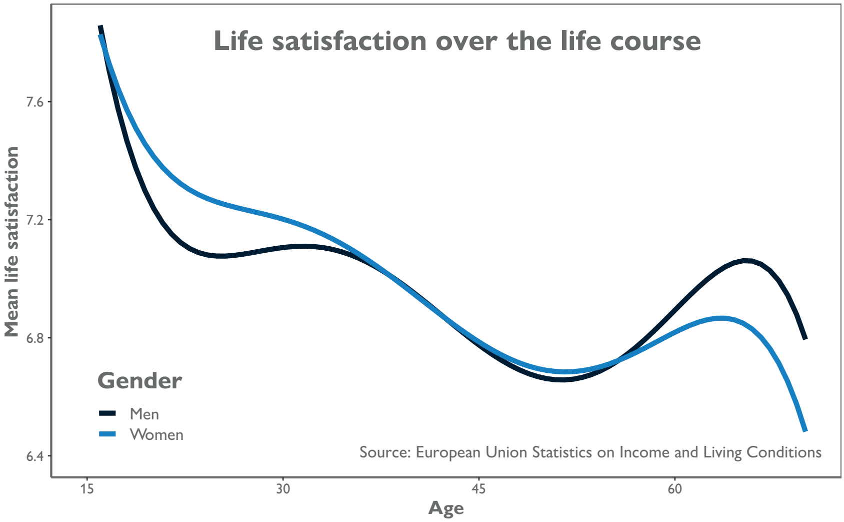
VIENNA YEARBOOK *of Population Research* 2021



"Demographic aspects of human wellbeing"

Guest Editors: Sonja Spitzer, Vanessa di Lego, Angela Greulich and Raya Muttarak

Vienna Yearbook of Population Research 2021
Austrian Academy of Sciences, Vienna

Special issue on
Demographic aspects of human wellbeing

Guest editors:
Sonja Spitzer, Vanessa di Lego, Angela Greulich
and Raya Muttarak

Vienna Yearbook of Population Research 2021 (Volume 19)

Guest editors for this issue:

Sonja Spitzer, Vanessa di Lego, Angela Greulich and Raya Muttarak

All Review and Research Articles in this publication were subject to international peer review.

Cover: Vanessa di Lego and Sonja Spitzer, stylised graph based on data from Clark et al., this volume

ISSN 1728-4414
ISBN 978-3-7001-8707-3

Copy editing: Miriam Hils, Bainbridge Island, USA
Layout: Sunrise Setting, Brixham, UK
Print: Prime Rate, Budapest, Hungary
Publisher: Austrian Academy of Sciences Press, Vienna, Austria (https://verlag.oeaw.ac.at)
https://epub.oeaw.ac.at/vypr_collection

Made in Europe

Authors for this volume

Hilal ARSLAN, Department of Social Research Methodology, Institute of Population Studies, Hacettepe University, Ankara, Turkey

Anna BARBUSCIA, INED, Paris, France and University of Lausanne, Switzerland

Felix BITTMANN, Leibniz Institute for Educational Trajectories (LIfBi), Department: Educational Decisions and Processes, Migration, Returns to Education, Bamberg, Germany

David E. BLOOM, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Thananon BUATHONG, Faculty of Sociology and Anthropology, Thammasat University, Thailand

Alanur ÇAVLIN, Department of Demography, Institute of Population Studies, Hacettepe University, Ankara, Turkey

Montakarn CHIMMAMEE, Social Research Institute, Chulalongkorn University, Thailand

Andrew E. CLARK, Paris School of Economics – CNRS, Paris, France

Chiara COMOLLI, University of Lausanne, Switzerland

Hippolyte D'ALBIS, Paris School of Economics – CNRS, Paris, France

Alessandra DE ROSE, Department of Methods and Models for Economics, Territory and Finance, Sapienza University of Rome, Italy

Vanessa DI LEGO, Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

Anna DIMITROVA, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

Victoria Y. FAN, University of Hawai'i at Mānoa, Office of Public Health Studies, Myron B. Thompson School of Social Work, Honolulu, HI, USA

Farid FLICI, Research Center in Applied Economics for Development (CREAD), Algiers, Algeria

Paul FRIJTERS, Department of Social Policy, London School of Economics, London, UK

Nina-Sophie FRITSCH, Institute for Sociology and Social Research, Vienna University of Economics and Business, Austria and Department of Sociology, University of Potsdam, Germany

Angela GREULICH, Sciences Po – OSC and Institut Universitaire de France, Paris, France

Nacer-Eddine HAMMOUDA, National Committee for Population, Health Ministry, Algiers, Algeria

Bernard HARRIS, School of Social Work and Social Policy, University of Strathclyde, Glasgow, UK

Carol JAGGER, Newcastle University, Population Health Sciences Institute, Newcastle upon Tyne, UK

Vadim KUFENKO, University of Hohenheim, Stuttgart, Germany

Richard E. LUCAS, Department of Psychology, Michigan State University, East Lansing, Michigan, USA

Wolfgang LUTZ, University of Vienna, International Institute for Applied Systems Analysis (IIASA), Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

Hideko MATSUO, Center for Sociological Research, Research Unit Family and Population Studies (https://soc.kuleuven.be/ceso/fapos), Leuven, Belgium

Koen MATTHIJS, Center for Sociological Research, Research Unit Family and Population Studies (https://soc.kuleuven.be/ceso/fapos), Leuven, Belgium

Raya MUTTARAK, Population and Just Societies Program, International Institute for Applied Systems Analysis (IIASA), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Laxenburg, Austria

Marília R. NEPOMUCENO, Max Planck Institute for Demographic Research, Rostock, Germany

Osondu OGBUOJI, Duke University, Duke Global Health Institute, Durham, NC, USA

Ronak PAUL, International Institute for Population Sciences, Mumbai, India

Klaus PRETTNER, Vienna University of Economics and Business, Vienna, Austria and Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

RASHMI, International Institute for Population Sciences, Mumbai, India

Claudia REITER, Department of Demography, University of Vienna, Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

Bernhard RIEDERER, Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Austria and Institute for Sociology, University of Vienna, Austria

Lena SEEWANN, Department of Sociology, University of Potsdam, Germany

Sonja SPITZER, Department of Demography, University of Vienna, Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

Erich STRIESSNIG, Department of Demography, University of Vienna, Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

Cássio M. TURRA, Demography Departament – Cedeplar, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Paolo Miguel M. VICERRA, College of Population Studies, Chulalongkorn University, Thailand

Gavin YAMEY, Duke University, Duke Global Health Institute, Durham, NC, USA

Dilek YILDIZ, International Institute for Applied Systems Analysis (IIASA), Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

Marina ZANNELLA, Department of Methods and Models for Economics, Territory and Finance, Sapienza University of Rome, Italy

Contents

Data & Trends

A demographic perspective on human wellbeing: Concepts, measurement and population heterogeneity

Sonja Spitzer[1,†,*]***, Vanessa di Lego***[2,†]***, Angela Greulich***[3] ***and Raya Muttarak***[4]

Abstract

This introduction to the 2021 special issue of the Vienna Yearbook of Population Research explores demographic perspectives on human wellbeing across time and space. While the idea of relating demographic parameters to wellbeing has been around for a while, a more concrete research agenda on this topic has only recently gained momentum. Reviewing the research presented in this volume, we show how existing theoretical concepts and methodological tools in demography can be used to make substantial advances in the study of wellbeing. We also touch upon the many challenges researchers face in defining and measuring wellbeing, with the most important debate being about whether the focus should be on objective or subjective measures. The studies discussed here define wellbeing as health and mortality; as income, education or other resources; as happiness or life satisfaction; or as a combination thereof. They cover wellbeing in historical and contemporary populations in high- and low-income countries, and also point out important barriers to research on wellbeing, including the lack of good quality data in many regions. Finally, we highlight the value of considering population heterogeneities when

[1]Department of Demography, University of Vienna, Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

[2]Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

[3]Sciences Po – OSC and Institut Universitaire de France, Paris, France

[4]Population and Just Societies Program, International Institute for Applied Systems Analysis (IIASA), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Laxenburg, Austria

[†]These authors contributed equally to this work.

[*]Correspondence to: Sonja Spitzer, sonja.spitzer@univie.ac.at

DOI: 10.1553/populationyearbook2021.int01

studying wellbeing in order to identify population subgroups who are likely to fall behind, which can have important policy implications.

Keywords: wellbeing; demography; subjective and objective measures; population heterogeneities

1 Background

Research on human wellbeing is often driven by the objective of "the greatest good for the greatest number" of people, a principle attributed to Jeremy Bentham, which is motivated by the desire to maximise wellbeing in a given population. However, since ancient Greece, there has been a longstanding debate about what constitutes wellbeing, and how it should be measured. While the hedonic Epicurean position was that wellbeing represents a life full of happiness, Aristotle proposed a definition of eudemonia, or of a life full of purpose, meaning and value. To date, it is still challenging to define and measure wellbeing, with the most important debates being about whether the focus should be on objective or subjective definitions of wellbeing, and about whether an individual or an aggregate perspective is the most useful for assessing wellbeing (Dodge et al. 2012; Voukelatou et al. 2020). Moreover, the approaches to conceptualising and measuring wellbeing differ substantially across disciplines. Economists, with their relatively long tradition of wellbeing research, have generally concentrated on objectively measurable dimensions, like GDP or household income (Reinhart et al. 2010). In contrast, sociologists and psychologists have tended to focus on a subjective concept of wellbeing that is linked to people's emotional and cognitive evaluations of their lives, psychosocial needs and an overall sense of purpose and emotional fulfilment (Diener et al. 2003; Tay and Diener 2011; Pleeging et al. 2021). These more subjective wellbeing indicators – like happiness and life satisfaction – have also recently been considered by other disciplines in the social sciences (Clark 2018).

As the scientific study of human population, demography seeks to measure various population characteristics, and to assess their drivers and implications. The ability to account for both macro- and micro-level changes across time and space allows for a broad, empirically based understanding of population trends and patterns (Billari 2015). This unique feature of demography makes the discipline ideal for the study of human wellbeing not only in the past and present, but also in the future, via forecasting (Muttarak et al. 2016). While the classic demographic events – birth, death and migration – have important links to wellbeing, mortality and health are used most frequently to conceptualise wellbeing.

In addition to existing theoretical and empirical tools, the ability to measure and incorporate population heterogeneity – i.e., variations among individuals in demographic behaviour, events and other outcomes – provides fruitful angles for the demographic analysis of wellbeing. Wellbeing differentials can be identified not only across the most commonly investigated population subgroups, like those based on age and gender, but also according to other key demographic characteristics, like level of education, relationship status and religion. Likewise, demography provides the resources for connecting these dimensions and analysing their interdependencies. By explicitly considering population heterogeneity, we increase our chances of preventing the infamous Benthamian pitfall of sacrificing the wellbeing of the small number for that of the greater number when seeking to maximise wellbeing.

This special issue of the Vienna Yearbook of Population Research serves as a starting point for exploring the contributions of demographic perspectives on the study of human wellbeing, while also bringing in different approaches and evidence from other scientific disciplines. It provides a deep dive into the concepts and measurements of human wellbeing in four contributions: two from the Demographic Debate; and two from Review Articles, in which the benefits and drawbacks of various wellbeing measures are discussed, along with their consequences for policy-making. Two additional Demographic Debates focus on specific policy targets, and on what is at stake when considering wellbeing. A further 13 Research Articles and a Data & Trends contribution present novel empirical results on different wellbeing dimensions and indicators for various regions of the world. Thus, these contributions confirm demography's role as a highly relevant discipline for advancing the field of wellbeing.

2 What is wellbeing in the demographic context, and how can we best measure it?

What constitutes wellbeing is both a philosophical and practical question. As demography is open to a range of fields and disciplines, it has the advantage of being able to accommodate a myriad of understandings, concepts and ways of measuring wellbeing that derive from other areas. This openness is reflected in the contributions in this special issue, which attempt to capture the multidimensionality of wellbeing using a wide range of data and methods. Here, wellbeing is variously defined based on health or mortality; income, education or other resources; happiness or life satisfaction; or a combination thereof, usually in the form of a composite wellbeing indicator representing both objective and subjective dimensions of wellbeing.

Even within these various interpretations of wellbeing, there is still considerable room for discussion about how best to measure it – or, even more importantly, about who should measure it. When analysing human wellbeing, who should judge wellbeing levels: the individuals themselves, or an objective outsider based on a common set of criteria? In his Demographic Debate, Clark reviews the current discussion surrounding the most appropriate measurement of human wellbeing. He shows how objective wellbeing can be complemented with subjective measures, which have the advantage of being democratic; i.e., individuals decide what is important when assessing wellbeing. Information on subjective wellbeing can also be easily collected based on only one survey question, and it typically suffers less from missing values than variables needed for other operationalisations of wellbeing.

Simply asking individuals about their wellbeing has empirical benefits, and many of the contributions in this special issue have applied this subjective approach. Yildiz, Arslan and Cavlin, for example, study the determinants of subjective wellbeing in Turkey based on the survey question: “All things considered, how satisfied are you with your life as a whole?” By conceptualising wellbeing as self-reported life satisfaction, the authors are able to provide a global evaluation of wellbeing that can account for all life domains over an extended period of time, as perceived by the individuals themselves. Similarly, Zannella and De Rose study Italian parents’ subjective wellbeing during the time they spend with their children. In particular, the impact of multitasking on subjective wellbeing related to time spent on childcare is evaluated based on the survey question: “How pleasant was the moment?” Again, the researchers rely on self-reports rather than objective sources when measuring wellbeing.

Lucas discusses the benefits and drawbacks of self-reported wellbeing in his Review. He focuses in particular on a major concern related to subjective wellbeing: namely, the challenge of providing global assessments of wellbeing. Global self-reports of wellbeing require individuals to consider and condense various life domains over a long period of time. Thus, such self-reports involve the tasks of remembering and aggregating – and the inability to perform these tasks effectively can distort the measurement of subjective wellbeing. In his contribution, he compares global self-reports of subjective wellbeing with experimental measures in which individuals evaluate their momentary experiences many times over a limited period of time. He concludes that while both types of measures have the potential to further our understanding of wellbeing, neither of them is without flaws.

These contributions thus pave the way for more “objective” measures of well-being to be considered. But which one of these measures is the most relevant? Historically, wellbeing has often been measured in monetary terms, a metric that – seemingly – allows wellbeing to be compared and aggregated. In particular, economists have generally considered GDP per capita relevant for measuring wellbeing. In the past decade, however, there has been a constant call to move away from using economic production as an indicator of social progress, and towards the adoption of a more holistic understanding of human wellbeing (Stiglitz et al. 2010; OECD 2021). It has become more common to complement purely economic measures with measures of social and environmental dimensions of quality of life. These dimensions can be measured at the aggregate and/or the individual level, and include, for example, employment, household debt, air and water quality, freedom of movement, freedom of family formation, freedom of political participation and access to institutions providing education and health services. The demographic measures that are used most often in this context are mortality and health, which are discussed in more detail in the next subsection.

3 Health and mortality as wellbeing concepts

While being alive is a precondition for enjoying any form of wellbeing, being in good health is an essential component of human welfare, and is an important factor in economic development that has both intrinsic and instrumental value (Bloom et al. 2019). The World Health Organization has defined health as a state of complete physical, mental and social wellbeing, and not just as the absence of illness (World Health Organization 2020). Thus, wellbeing is explicitly associated with being healthy. Based on this definition, promoting health and wellbeing at older ages was established as one of the three priority areas outlined in the 2002 Madrid International Plan of Action on Ageing, which aimed to ensure that population ageing is sustainable and equitable (World Health Organization 2020). Ensuring healthy lives and promoting wellbeing among people of all ages is also one of the 17 Sustainable Development Goals (United Nations Development Programme 2015). In addition, in 2008, the EU Innovation Partnership on Active and Healthy Ageing set a target of achieving an increase of two Healthy Life Years by 2020; and in 2018, the UK government set a target to "ensure people can enjoy at least 5 extra healthy, independent years of life by 2035, while narrowing the gap between the experience of the richest and poorest". Despite these efforts, Jagger argues in her Demographic Debate that the gap in Healthy Life Years between European Union member states has widened. Indeed, past trends in disability-free life expectancy in England and projections suggest that it will be difficult to achieve an increase of five healthy and independent years by 2035.

Several other contributions in this special issue also use health status to capture human wellbeing. Barbuscia and Comolli, for example, explore gender inequalities in Switzerland and France. The authors analyse the accumulation of gender inequalities with age using different measures of physical and mental health, as well as relational wellbeing. Their results show that while gender inequalities are greater for many dimensions of wellbeing in Switzerland than in France, socio-economic inequalities in health and wellbeing are more pronounced in France than in Switzerland. The researchers also report that while gender inequalities tend to accumulate with age, educational gaps do not seem to diverge over time. These findings suggest that gender inequalities tend to be based on structural disadvantages that grow and accumulate with life events and the ageing process.

Similarly, Nepomuceno, di Lego and Turra focus on gender inequalities by considering the intersection of wellbeing, health and gender-specific issues in Latin America and the Caribbean. Their work presents empirical evidence on the gender paradox in health and mortality, whereby women live longer than men, but with higher rates of disability. In particular, the authors address the challenges of measuring this paradox in a unique environment characterised by high levels of socio-economic inequality that interact with gender- and age-specific disparities in health and mortality. When coupled with high levels of widowhood and impoverishment among older women, improvements in life expectancy may increase the burden of health for women, which may, in turn, negatively affect women's wellbeing (Monteverde et al. 2009). While the authors note that a lack of data hampered their efforts to investigate as many countries in the region as possible, the evidence they present confirms the persistence of the gender paradox for countries in Latin America and the Caribbean.

Indeed, variations in the measurement of wellbeing are determined not only by differences in the theoretical concepts and methods used, but also by differences in the availability and quality of data. Gaining access to appropriate data is often a challenge, particularly when studying the wellbeing of historical populations, as proper data may not be available for a given time period; or when studying the wellbeing of low-income countries, as data quality is often poor in such contexts. Problems with data quality can make it difficult to monitor trends in wellbeing or to set targets for policy interventions. Flici and Hammouda explore this topic by extensively evaluating the quality of mortality data in Algeria, and pointing out important data issues that can lead to biased interpretations. They investigate questionable observations, such as sudden changes in the mortality age pattern, and excess female mortality at reproductive ages that unexpectedly appears and disappears in specific years. The authors find that while the population census of 1998, which coincided with the end of the civil war, had better data quality, reliable information on mortality and health in Algeria is still lacking. In particular, the authors show how controlling for data quality and understanding irregular patterns can enable researchers to identify and explain the data issues in a country context with limited data.

Likewise, Harris shows that in the absence of survey or census data, alternative data sources may be used to fill in the gaps in the data for certain historical populations. The author demonstrates how anthropometric data from sources such as military recruitment or prison records can be used to study wellbeing in historic contexts. In his Review Article, Harris explores and assesses population height as a historical measure of wellbeing, noting that population health can reflect the economic and environmental conditions at the time of measurement.

4 Education as another aspect of wellbeing

While health is widely used as a key dimension of wellbeing, in a similar function, education can also be considered as both a means to achieve and an ultimate end of human wellbeing. The articles by Paul and Rashmi as well as by Bittmann operationalise education in this manner. Using panel survey data for India and Germany, respectively, they are able to capture how the subjective and objective school performance of children affects their educational outcomes and other dimensions of wellbeing later in life.

Paul and Rashmi investigate for India how children's educational wellbeing influences their school dropout and grade repetition rates measured seven years later. Educational wellbeing is assessed using objective and subjective measures based on both theoretical concepts and empirical results from a principal component analysis.

They find that both objective and subjective measures of educational wellbeing are key determinants of educational outcomes, as measured by school dropout and grade repetition rates, even after controlling for relevant household and geographic characteristics.

Bittmann examines the effects of teachers' subjective assessments of which secondary school academic track is appropriate for each child, the level of compliance with the teachers' recommendations, and the socio-economic background of the parents on the subjective and objective educational outcomes of children in Germany. The author shows that pupils who chose to continue their secondary education on the academic track even though their teacher recommended a different track tend to perform worse on competency tests, to have lower grades, and to report lower levels of satisfaction and reading enjoyment than pupils who were advised to follow the academic track. Thus, Bittmann's results suggest that the teachers' assessments predict rather well how the students will perform academically in secondary school. Ignoring such a recommendation – which has been called academic track mismatch – can result in individuals having poor educational outcomes later in life, both objectively and subjectively. These two articles clearly highlight the importance of considering both the subjective and the objective aspects of wellbeing measurement, as well as the multidimensionality of wellbeing.

5 Population heterogeneity in wellbeing

Another important question for the study of wellbeing is how wellbeing differs across population subgroups, and whether certain demographic and socio-economic characteristics provide protective effects against the deterioration of wellbeing. Being able to identify which subgroups are likely to fall behind in wellbeing indicators could facilitate the design of policy interventions aimed at closing the gaps between groups.

Matsuo and Matthijs, for example, address heterogeneities by age and cohort by studying subjective wellbeing trajectories across generations. They find a U-shaped relationship between age and happiness that reflects lower levels of happiness in mid-life, followed by an increase in happiness at older ages. The authors also note, however, that the patterns are rather flat, and show considerable diversity after age 80 in many countries. In terms of cohort effects, the baby boomers are found to have lower happiness levels than pre-boomers and post-boomers, which may indicate that baby boomers have been negatively affected by their early and formative life conditions and experiences. For example, economic inequality rose from the late 1960s onwards, in part because a larger cohort size created greater competition for schools and jobs. The authors argue that the cohort-specific factors of subjective wellbeing is an interesting field of future research, particularly for studying the Millennials, who are facing new challenges in adulthood that were not experienced by the previous generations.

D’Albis, Clark and Greulich also investigate the U-shaped relationship between life satisfaction and age, with a focus on heterogeneities between partnered and unpartnered individuals. First, the authors demonstrate that the U-shape generally holds in Europe, but is flatter for partnered than for non-partnered individuals. They also show that gender differences in life satisfaction by age – namely, that women’s life satisfaction is higher at younger ages and is lower at older ages – disappear after controlling for partnership status. This pattern is, however, less clear in the panel data, which the authors analyse in a second step. Commenting on these mixed results, the authors emphasise that the results on life satisfaction by age strongly depend on the type of data analysed. They also point out that to study how key demographic, economic and psychological variables change over the lifecycle, reliable longitudinal datasets are needed for a larger set of countries.

Riederer, Fritsch and Seewann also explore the impact of relationship status on wellbeing by comparing the subjective wellbeing of single households in Vienna with that of other household types. Their contribution introduces another layer to this special issue: namely, that of various dimensions of subjective wellbeing. Their findings show that singles in Vienna are less satisfied with their family life than two-adult households, but also that singles are more satisfied with their financial situation, leisure time and housing than households with children.

6 Putting it all together: multidimensional approaches to sustainable wellbeing

Several contributions in this special issue are concerned with the multidimensional aspect of wellbeing, while emphasising that all of the metrics discussed above are potentially valid measures of wellbeing, and that various demographic aspects should be considered when using wellbeing as a basis for decision-making. Like Bentham’s principle of “the greatest good for the greatest number”, the measurement of wellbeing is usually not a means to an end, but is, rather, a tool for evaluating and judging decisions or actions at various levels. These wellbeing indicators often aim to “go beyond GDP” by replacing a simple monetary metric with a more holistic measure that encompasses different quality of life domains. Accordingly, several indicators designed to capture the multidimensional nature of wellbeing, including both its objective and the subjective components, have been developed.

The Research Article by Bloom, Fan, Kufenko, Ogbuoji, Prettner and Yamey emphasises that per capita GDP has significant limitations when used as a standalone tool for measuring wellbeing, as it does not capture many dimensions associated with a “good life,” such as health and equality of opportunity. Nonetheless, they acknowledge that GDP per capita has a number of advantages, including being easy to interpret and requiring data that are widely available for a large number of countries. With this motivation in mind, the authors propose a new parsimonious measure of wellbeing that preserves the advantages of per capita GDP, but that also includes health and equality. Thus, they offer an indicator that simultaneously accounts for income, health and inequality, and that can be readily interpreted as inequality-adjusted healthy lifetime income. This approach is useful for monitoring trends in wellbeing across countries and over time, and as an alternative strategy for performing cross-country comparative analysis.

Along these lines, Frijters critically discusses in his Demographic Debate article the practise of relying on cost-benefit analyses as the basis for making public spending decisions. As an alternative to assessing economic surplus, he suggests using the WELLBY indicator, which is one unit of self-reported life satisfaction on a scale of zero to 10 for one person for one year. To account for potential jealousy and envy effects – i.e., for the possibility that the higher consumption of others reduces an individual's wellbeing – the author proposes a reduction factor of 75% on all estimates of private consumption benefits for this new wellbeing framework. He suggests that decision-making should be based on the expected stream of WELLBYs, as estimated using the results of the several thousand studies on the determinants of life satisfaction.

The indicators mentioned above implicitly or explicitly take a demographic stance when measuring wellbeing. Frijters, for example, sums wellbeing over the population of interest. Similarly, Bloom, Fan, Kufenko, Ogbuoji, Prettner and Yamey include healthy life expectancy in their wellbeing indicator. In his Demographic Debate, Lutz emphasises that existing demographic tool kits, such as life tables and Sullivan methods, can be used to analyse human wellbeing, by, for example, calculating the person-years lived at any given age in the status of interest. This application of demographic methods allows for a straightforward and simple interpretation of wellbeing, as exemplified in Lutz, who introduces a new indicator of wellbeing called Years of Good Life (YoGL). Three of the empirical contributions in this special issue are dedicated to this newly introduced measure, which considers both subjective and objective measures of wellbeing, including self-reported life satisfaction, poverty and health. These research articles analyse YoGL across countries and population groups, as well as over time. Specifically, Striessnig, Reiter and Dimitrova reconstruct the indicator back to 1950 for 140 countries; Reiter and Spitzer use YoGL to decompose wellbeing by country and gender for the European population aged 50+; and Dimitrova and Buathong illustrate how the indicator can be applied using novel primary data collected in Thailand.

In his contribution, Lutz also outlines how demographic models, such as multidimensional demographic dynamics, can be used to project future population trends based on alternative scenarios. In particular, he notes that because demography is able to produce longer-term quantitative models, it can play a central role in efforts to account for the temporal dimension of changes in wellbeing. Such models can, in turn, be used to explore sustainable human wellbeing, which is the ultimate goal of sustainable development, while accounting for future challenges like climate change. Thus, holistic measures of human wellbeing are highly relevant for the Sustainable Development Agenda 2030, which seeks to achieve sustainable improvements in human wellbeing, while maintaining a balance between the environment and economic progress (Messerli et al. 2019). Monitoring the progress of sustainable development therefore requires overarching empirical measurements of human wellbeing.

Developing and constantly improving these wellbeing measures, as well as providing relevant data, particularly in low- and middle-income countries, are needed to formulate policies that promote wellbeing at the individual and the aggregate level. With contributions from authors from various countries, such as Algeria, India, Brazil, Turkey and Thailand, this special issue represents a first step towards producing evidence on wellbeing for countries outside of Europe and North America. This special issue also demonstrates that the study of wellbeing should consider population heterogeneity whenever possible, because failing to do so hampers efforts to identify population subgroups who are likely to fall behind in the development process. To promote policies designed to enhance wellbeing for the greatest number of people, a multidimensional approach that captures both subjective and objective dimensions, as well as different life domains, is needed, as the contributions in this volume vividly illustrate.

References

Billari, F. C. 2015. Integrating macro- and micro-level approaches in the explanation of population change. *Population Studies* 69: S11–S20. https://doi.org/10.1080/00324728.2015.1009712

Bloom, D. E., M. Kuhn and K. Prettner 2019. Health and economic growth. *Oxford Research Encyclopedia of Economics and Finance*. https://doi.org/10.1093/acrefore/9780190625979.013.36

Clark, A. E. 2018. Four decades of the economics of happiness: Where next? *Review of Income and Wealth* 64(2): 245–269. https://doi.org/10.1111/ROIW.12369

Diener, E., S. Oishi and R. E. Lucas 2003. Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual Review of Psychology* 54(1): 403–425. https://doi.org/10.1146/annurev.psych.54.101601.145056

Dodge, R., A. Daly, J. Huyton and L. Sanders 2012. The challenge of defining wellbeing. *International Journal of Wellbeing* 2(3): 222–235. https://doi.org/10.5502/ijw.v2i3.4

Messerli, P., E. M. Kim, W. Lutz, J.-P. Moatti, K. Richardson, M. Saidam, D. Smith, P. Eloundou-Enyegue, E. Foli, A. Glassman, G. H. Licona, E. Murniningtyas, J. K. Staniškis, J.-P. van Ypersele and E. Furman 2019. Expansion of sustainability science needed for the SDGs. *Nature Sustainability* 2(10): 892–894. https://doi.org/10.1038/s41893-019-0394-z

Monteverde, M., K. Noronha and A. Palloni 2009. Effect of early conditions on disability among the elderly in Latin America and the Caribbean. *Population Studies* 63(1): 21–35. https://doi.org/10.1080/00324720802621583

Muttarak, R., W. Lutz and L. Jiang 2016. What can demographers contribute to the study of vulnerability? *Vienna Yearbook of Population Research* 13: 1–13. https://doi.org/10.1553/populationyearbook2015s1

OECD 2021. Better life initiative: Measuring well-being and progress. OECD Better Policies for Better Lives. https://www.oecd.org/statistics/better-life-initiative.htm, accessed on 21 June 2021.

Pleeging, E., M. Burger and J. van Exel 2021. The relations between hope and subjective well-being: A literature overview and empirical analysis. *Applied Research in Quality of Life* 16(3): 1019–1041. https://doi.org/10.1007/s11482-019-09802-4

Reinhart, C. M., V. R. Reinhart, C. Reinhart and V. Reinhart 2010. After the Fall. NBER Working Paper 16334. https://doi.org/10.3386/w16334

Stiglitz, J. E., A. K. Sen and J.-P. Fitoussi 2010. Report by the commission on the measurement of economic performance and social progress. Paris: Commission on the Measurement of Economic Performance and Social Progress. https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf, last accessed 13 March 2015.

Tay, L. and E. Diener 2011. Needs and subjective well-being around the world. *Journal of Personality and Social Psychology* 101(2): 354–365. https://doi.org/10.1037/a0023779

United Nations Development Programme 2015. Human Development Report 2015. New York, NY. http://hdr.undp.org/en/content/human-development-report-2015

Voukelatou, V., L. Gabrielli, I. Miliou, S. Cresci, R. Sharma, M. Tesconi and L. Pappalardo 2020. Measuring objective and subjective well-being: dimensions and data sources. *International Journal of Data Science and Analytics* 11(4): 279–309. https://doi.org/10.1007/S41060-020-00224-2

World Health Organization 2020. Basic Documents: Constitution of the World Health Organisation. https://apps.who.int/gb/bd/pdf_files/BD_49th-en.pdf

DEBATE

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 15–21

EU and UK targets for healthy life expectancy – are they achievable?

Carol Jagger[1,*]

Abstract

In 2008, the EU Innovation Partnership on Active and Healthy Ageing (EIP-AHA) set a target of an increase of two healthy life years by 2020. More recently, in 2018, the UK Government set a target to "ensure people can enjoy at least 5 extra healthy, independent years of life by 2035, while narrowing the gap between the experience of the richest and poorest". This paper reviews the progress the EU member states have made towards meeting this target, and what the UK can learn from their experiences. I conclude that, although the EU target is likely to be reached, the gap in healthy life years between the member states has increased. Past trends in and projections of disability-free life expectancy in England suggest that it will be difficult to achieve an increase of five healthy and independent years of life by 2035.

Keywords: health expectancy; life expectancy; healthy life years; social inequality; European Union

1 Introduction

In recent decades, the idea of focusing on the quality of remaining life – "adding life to years" – has become popular with the public and the academic community as a means of offsetting some of the challenges, both personally and societally, of population ageing. Health expectancies are population health indicators that combine information on health and mortality, and measure the remaining years spent in good health at a particular age. Health expectancies have also gained credence among governments as indicators of the outcomes of policies. In the early 2000s, the European Union developed healthy life years (HLY), a disability-free life expectancy, as the first pan-European health indicator. The health measure underlying HLY is the global activity limitation indicator, which is available

[1]Newcastle University, Population Health Sciences Institute, Newcastle upon Tyne, UK

[*]Correspondence to: Carol Jagger, carol.jagger@ncl.ac.uk

DOI: 10.1553/populationyearbook2021.deb01

annually in the EU Statistics of Income and Living Conditions (EU-SILC) survey. HLY was later chosen as the outcome for the European Innovation Partnership on Active and Healthy Ageing (EIP-AHA) target to "increase HLY by 2 years by 2020". More recently, the UK has declared as part of the Ageing Grand Challenge the intention to "increase by 5 years healthy and independent life by 2035 whilst narrowing the gap between the richest and the poorest". In this paper, I reflect on the likelihood that the UK will achieve this goal by looking first at the experiences of the EU, and then at past and forecasted trends in England. I conclude by discussing some approaches for achieving this target.

2 EU trends

Between 2010 and 2017, men's HLY at birth for the 28 member states of the EU (EU28) rose by 1.7 years, from 61.8 years to 63.5 years. Women's HLY at birth rose less, by 1.4 years, from 62.6 years to 64.0 years. Over this period, the gender gap in HLY declined slightly from 0.8 years to 0.5 years. Thus, if these trends continue, by the target date of 2020, HLY at birth will have increased by 2.4 years for men and by 2.0 years for women, the goal of an increase of two HLY will have been achieved, and the gender gap in HLY will be 0.4 years. Moreover, the gain in life expectancy at birth between 2010 and 2017 was 1.4 years for men and 0.7 years for women, which is evidence of a compression of disability for men and women. Whether the narrowing of the gender gap will continue is unknown. However if it does, it will likely to be due to the greater increase in LE in men than in women, which has been attributed to declines in cardiovascular and stroke mortality, as well as to the higher rate of smoking among women in the EU28 compared to in other countries (Leon et al. 2019; Sundberg et al. 2018).

Nevertheless, the picture is far from perfect. Inequalities in HLY between the EU28 member states have increased between 2010 and 2017, by 4.9 years for men's HLY and by 2.1 years for women's HLY. One factor that contributed to these increased inequalities is that the economic crisis was experienced differently by the member states. For example, the largest decreases in gross domestic product and the largest increases in unemployment were experienced by Latvia and Lithuania, the member states with some of the lowest life and healthy life expectancies in the EU (Stuckler et al. 2017). Increasing inequalities within the member states might have also contributed to inequalities in HLY. Nevertheless, increased inequalities were anticipated in a simulation of how these trends would play out (Jagger et al. 2013). For all member states to achieve two additional HLY by 2020 along with a reduction in the inequalities between the member states, an increase of 6.4 HLY would be required, which would lead to a 50% reduction in the gap. However reducing the gap by 50% alone would result in all but two member states, Malta and Sweden (which currently have some of the highest HLY levels in the EU28), gaining two additional HLY.

3 UK trends

It is useful to look at the past trends in and current values of healthy life expectancy to assess whether the UK will be able to achieve its target. Through the Office for National Statistics (ONS), the UK has a long history of monitoring two health expectancies based on self-rated health (healthy life expectancy - HLE) and on limiting long-standing illness (disability-free life expectancy - DFLE). Between decennial censuses, these health expectancies do not include the population in residential care, but this lack of data has little effect on health expectancies at birth. Between 2009–11 and 2015–17, HLE at birth increased by 0.4 years for men and decreased by 0.2 years for women (Office for National Statistics 2018). However, over this period, gains in life expectancy at birth were greater, at 0.8 years for men and at 0.6 years for women. For both men and women, years lived in "not good" health increased in relative and in absolute terms, with a consequent expansion of morbidity. Over the 20-year period between 1991 and 2011, HLE at age 65 increased by 3.8 years for men and by 3.1 years for women (Jagger et al. 2016). As these increases were significantly smaller than the increases in life expectancy (4.5 years for men and 3.6 years for women), there was again an increase in years spent with ill health, and, therefore, an expansion of morbidity. DFLE at age 65 for men increased again, but by only 2.6 years, which was smaller than the increase in HLE. However, since DFLE at age 65 for women increased by only 0.5 years, the majority of the 3.6 years in life expectancy women gained were years spent with disability. The number of years spent living independently at age 65 followed a similar pattern, increasing by 1.7 years for men, but by only 0.2 years for women (Kingston et al. 2017).

Although there is a body of work on gender disparities in health expectancies, little is known about gender disparities in trends in health expectancies. While there is evidence that differential health reporting between men and women may contribute to differences in health expectancy (di Lego et al. 2020), the greater prevalence of chronic conditions among women (disabling conditions such as arthritis) than among men (more fatal conditions such as cardiovascular disease) appears to play a part as well. However, though differential reporting may affect the gender differences in health expectancies, it is unlikely to result in different trends. The decline in fatalities caused by chronic conditions among men may have led to the gap between DFLE and LE being narrowed more among men than among women.

Looking at the potential for narrowing the HLE gap between rich and poor, it is clear that inequalities in HLE between socio-economic groups are large in the UK, at around 19 years for men and women at birth, or double the inequalities in life expectancy at birth (Office for National Statistics 2019). There is evidence that inequalities increased slightly between 2012–14 and 2015–17. Large inequalities in multimorbidity have also been found (Chan et al. 2019). These findings suggest that when it comes to reducing inequalities in HLE in the UK, there is still a long way to go.

Simulation models that forecast the health of future populations have been developed, but few of these models incorporate forecasts of health expectancy (Jagger and Kingston 2020). The most comprehensive of these approaches is the Population Ageing and Care Simulation (PACSim) model, as it overcomes a number of the limitations of other models. Specifically, PACSim: (a) brings together the three national or nationally representative longitudinal studies of ageing covering England for the base population; (b) includes the real health and lifestyle profiles of people aged 35 years and over who will age into the older population aged 65 and over during the subsequent 20 years; and (c) includes a wide range of socio-demographic characteristics, health behaviours, and 12 diseases and geriatric conditions that allow for an estimate of multimorbidity. PACSim forecasts that between 2015 and 2035, years spent independent from age 65 will increase by 4.2 years for men and 0.9 years for women (Kingston et al. 2018). Since for men, life expectancy at age 65 will increase less than the years spent independent, there will be a compression of dependency for men. For women, on the other hand, there will be an increase in dependency – although the level of dependency among women will be relatively low, requiring care less frequently than daily. PACSim also forecasts considerable increases in the prevalence of and the years spent with multimorbidity (Kingston et al. 2018). The reasons for this trend are two-fold. First, the numbers of very old people, among whom multimorbidity is common, are rising. Second, people entering the older population in subsequent years have a greater likelihood of already having one disease.

4 Conclusions

The evidence from past trends and forecasts suggest that it will not be easy for the UK to achieve the target of increasing the number of healthy, independent years Britons can enjoy by five years, while narrowing the gap between the richest and poorest population groups. Lessons learned from the EU's experiences with HLY trends indicate that there is a need to focus on narrowing the gap, as national strategies designed to increase HLE are likely to result in the greatest increases occurring in the most advantaged areas. Nevertheless, narrowing the gap will be particularly challenging given the size of the current gap, and the direction of change (currently increasing). Gaining an understanding of the barriers to the uptake of healthy behaviours (healthy diet, physical activity, smoking cessation, reducing alcohol consumption) in the most disadvantaged communities, as well as making investments in communities, raising employment levels, and taking into account the wider social determinants of health, are all required (All Party Parliamentary Group for Longevity 2020).

In considering which strategies might improve HLE, there are three points worth noting. First, the increases in years spent with multimorbidity indicate that there is a need for a greater focus on prevention in early and mid-adulthood, as well as on delaying the disabling consequences of chronic conditions. Second, most of the remaining years of life after age 65 are spent living independently. Given that ageing is malleable, interventions and strategies aimed at delaying decline will be most beneficial in achieving gains in years of healthy, independent life (Gore et al. 2018). However, finding solutions will require gaining an understanding of the barriers that more disadvantaged adults face in adopting healthy lifestyles, as well as recognising that implementing these solutions may require greater investment in, for example, jobs, housing, leisure facilities, and neighbourhood improvements.

Third, and perhaps most important, it is crucial to monitor both HLE and LE together, since interventions may increase HLE, but if they increase LE by the same amount or more, there will be an overall increase in unhealthy years. A good illustration of this dynamic is the differing effects of smoking and obesity on DFLE. Evidence from cross-sectional studies suggests that, of the three lifestyle factors assessed – smoking, obesity, and alcohol consumption – obesity is most strongly associated with spending more years of life with disability, with obese individuals living on average more years with disability (5.9 years) than smokers (3.8 years) and drinkers (3.1 years) (Klijs et al. 2011). In contrast, there is evidence that smoking has the strongest effect on life expectancy, with life expectancy at age 55 differing by 4.0 years based on smoking, 3.0 years based on alcohol consumption, and 1.4 years based on BMI. Longitudinal observational studies have generated similar findings (Majer et al. 2011).

In conclusion, reaching the UK target of an increase of five healthy and independent years of life by 2035, while narrowing the socio-economic gaps in HLE and DFLE, may be attainable for men, but will not be easy for women. It appears likely that the overall European target will be achieved by 2020, but will be accompanied by increased inequalities between member states, many of which will not see the same gains, or any increase at all. Nevertheless, countries must aspire to reach such targets if we are to achieve the longer term goals of enabling healthy ageing for all, and of gaining some control over the rising demands for health and social care worldwide.

References

All Party Parliamentary Group for Longevity 2020. The Health of the Nation - a strategy for healthier longer lives. London: Longevity UK.

Chan, M. S., A. van den Hout, M. Pujades-Rodriguez, M. M. Jones, F. E. Matthews, C. Jagger, R. Raine and M. Bajekal 2019. Socio-economic inequalities in life expectancy of older adults with and without multimorbidity: a record linkage study of 1.1 million people in England. *International Journal of Epidemiology* 48(4): 1340–1351. https://doi.org/10.1093/ije/dyz052.

di Lego, V., P. Di Giulio and M. Luy 2020. Gender differences in healthy and unhealthy life expectancy. In *International Handbook of Health Expectancies*, eds C. Jagger, E. C. Crimmins, Y. Saito, R. Tiene de Carvalho Yokota, H. van Oyen and J.-M. Robine. Springer.

Gore, P. G., A. Kingston, G. R. Johnson, T. B. L. Kirkwood and C. Jagger 2018. New horizons in the compression of functional decline. *Age and Ageing* 47(6): 764–768. https://doi.org/10.1093/ageing/afy145.

Jagger, C. and A. Kingston 2020. Forecasting health expectancy – what the future might hold. In *International Handbook of Health Expectancies*, eds C. Jagger, E.C. Crimmins, Y. Saito, R. Tiene de Carvalho Yokota, H. van Oyen and J.-M. Robine. Springer.

Jagger, C., F. E. Matthews, P. Wohland, T. Fouweather, B. C. Stephan, L. Robinson, A. Arthur and C. Brayne 2016. A comparison of health expectancies over two decades in England: results of the Cognitive Function and Ageing Study I and II. *Lancet* 387(10020): 779–786. https://doi.org/10.1016/s0140-6736(15)00947-2.

Jagger, C., M. McKee, K. Christensen, K. Lagiewka, W. Nusselder, H. Van Oyen, E. Cambois, B. Jeune and J.-M. Robine 2013. Mind the gap—reaching the European target of a 2-year increase in healthy life years in the next decade. *European Journal of Public Health* 23(5): 829–833. https://doi.org/10.1093/eurpub/ckt030.

Kingston, A., A. Comas-Herrera and C. Jagger, for the Modem Project 2018. Forecasting the care needs of the older population in England over the next 20 years: estimates from the Population Ageing and Care Simulation (PACSim) modelling study. *Lancet Public Health* 3(9): E447–E455. https://doi.org/10.1016/s2468-2667(18)30118-x.

Kingston, A., P. Wohland, R. Wittenberg, L. Robinson, C. Brayne, F. E. Matthews and C. Jagger 2017. Is late-life dependency increasing or not? A comparison of the Cognitive Function and Ageing Studies (CFAS). *Lancet* 390(10103): 1676–1684. https://doi.org/10.1016/s0140-6736(17)31575-1.

Kingston, A., L. Robinson, H. Booth, M. Knapp and C. Jagger 2018. Projections of multi-morbidity in the older population in England to 2035: estimates from the Population Ageing and Care Simulation (PACSim) model. *Age and Ageing* 47(3): 374–380. https://doi.org/10.1093/ageing/afx201.

Klijs, B., J. P. Mackenbach and A. E. Kunst 2011. Obesity, smoking, alcohol consumption and years lived with disability: a Sullivan life table approach. *BMC Public Health* 11(1): 378. https://doi.org/10.1186/1471-2458-11-378.

Leon, D. A., D. A. Jdanov and V. M. Shkolnikov 2019. Trends in life expectancy and age-specific mortality in England and Wales, 1970–2016, in comparison with a set of 22 high-income countries: an analysis of vital statistics data. *The Lancet Public Health* 4(11): e575–e582. https://doi.org/10.1016/S2468-2667(19)30177-X.

Majer, I. M., W. J. Nusselder, J. P. Mackenbach and A. E. Kunst 2011. Life expectancy and life expectancy with disability of normal weight, overweight, and obese smokers and nonsmokers in Europe. *Obesity (Silver Spring)* 19(7): 1451–1459. https://doi.org/10.1038/oby.2011.46.

Office for National Statistics 2018. Health state life expectancies, UK: 2015 to 2017.

Office for National Statistics 2019. Health state life expectancies by national deprivation deciles, England and Wales: 2015 to 2017.

Stuckler, D., A. Reeves, R. Loopstra, M. Karanikolos and M. McKee 2017. Austerity and health: the impact in the UK and Europe. *European Journal of Public Health* 27 (suppl_4): 18–21. https://doi.org/10.1093/eurpub/ckx167.

Sundberg, L., N. Agahi, J. Fritzell and S. Fors 2018. Why is the gender gap in life expectancy decreasing? The impact of age- and cause-specific mortality in Sweden 1997–2014. *International Journal of Public Health* 63(6): 673–681. https://doi.org/10.1007/s00038-018-1097-3.

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 23–28

Demography and well-being

Andrew E. Clark[1,*]

Abstract

Demography studies the characteristics of populations. One such characteristic is well-being: this was the subject of the 2019 Wittgenstein Conference. Here, I discuss how objective well-being domains can be summarised to produce an overall well-being score, and how taking self-reported (subjective) well-being into account may help in this effort. But given that there is more than one type of subjective well-being score, we would want to know which one is "best". We would also need to decide whose well-being counts, or counts more than that of others. Finally, I briefly mention the potential role of adaptation and social comparisons in the calculation of societal well-being.

Keywords: subjective well-being; demography; measurement; policy

How are demography and well-being linked? The answer to this question of course depends on what is meant by both of these terms. Luckily, the definition of one of them appears to be unambiguous. Turning to an increasingly valuable research resource (Wikipedia), we see that demography is defined as "*the statistical study of populations, especially human beings*". That seems pretty straightforward. Demography and well-being can then interact by considering the well-being of individuals in different population subgroups.

Well-being, on the other hand, can be operationalised in many different ways. One popular way of summarising this concept has been as income, which represents an individual's command over resources. Another has been as access to basic needs (Townsend 1979 and Streeten 1981). A third has been as capabilities (what the individual has the opportunity to achieve: Sen 1985). These are objective definitions of well-being: we observe which individuals possess these resources or capabilities, and which do not.

[1]Paris School of Economics - CNRS, Paris, France

[*]Correspondence to: Andrew E. Clark, Andrew.Clark@ens.fr

DOI: 10.1553/populationyearbook2021.deb02

However, there are many different types of resources available to individuals apart from those that we can easily directly measure in surveys. Moreover, different individuals may value these resources differently. It would seem to be impossible to measure all of the possible resources and their importance at the same time for many thousands of individuals in different populations.

A simple solution to this data collection problem is to outsource it to the individuals themselves. We capture the importance of all of the many resources that individuals find valuable by asking them to provide an overall assessment of how their life is going. This subjective evaluation of one's own life is called subjective well-being. It has *a priori* much to recommend it: subjective well-being is democratic (individuals themselves, not the researcher, decide what is important), simple to collect (one question may suffice) and easy for respondents to understand (missing values for subjective well-being questions are an order of magnitude smaller than those for income, for example). One common measure of subjective well-being is life satisfaction (although, as we will outline below, there are many others).

Research that uses subjective well-being measures has flourished in recent years (see Clark 2018, for a recent review). Some of these contributions have looked explicitly at the correlations between subjective well-being and various objective well-being measures, which include, using our list above, income (Clark et al. 2008), access to basic needs (Clark and D'Ambrosio 2019) and capabilities (Bruni et al. 2008).

Going back to Wikipedia, the definition of well-being is a broad one: "*Well-being, wellbeing, or wellness is the condition of an individual or group. A higher level of well-being means that in some sense the individual's or group's condition is more positive*". But the spirit of the term seems to be subjective: the accompanying photo (Note: accessed 12 July 2020) is captioned "*Children appearing to experience well-being after an art class*". The "in some sense" could refer to the dimensions of well-being that the individual finds important. While we as outsiders can conclude that individuals appear to be experiencing well-being, we can also try to determine whether this is indeed the case by actually asking them.

Given the breadth of conditions that might be meant by the term well-being, I applaud the decision of the organisers to focus the 2019 Wittgenstein Conference on demographic aspects of human well-being. What did we learn there? I had three broad reactions to the 36 papers presented over the two days of this conference.

The first is that the gamble to mix demography with well-being paid off. The number of submissions, presentations and attendees surpassed my expectations. My second, and related, reaction is that the conference showed that it is possible to put researchers from different disciplines in the same room without (too much) blood being spilled. I have long believed that one of the positive aspects of the analysis of (subjective) well-being is that it appeals to so many different disciplines across the social sciences. Over time, the cross-disciplinary attention to the topic has produced some common ground in terms of what aspects of well-being we are most interested in, how we measure well-being, and what statistical tools we apply to test our hypotheses. My last broad reaction is more policy-oriented: almost every speaker I heard at the conference managed to get their point across in 12 minutes, which led me to wonder why faculty seminars have to be 75 or 90 minutes long.

To move into the details, I noted above that there was "some common ground". When it comes to measurement of well-being, the key word may really be "some". Regarding the objective measures of national well-being, there is, by now, probably a consensus that we should go beyond relying on GDP alone. There is, however, little agreement beyond that point. We have not yet reached a consensus about how many new variables need to be added to GDP in order to summarise societal well-being, or about how the different elements of this new index should be weighted. Is crime as important as inequality? How much health am I willing to give up to improve the environment? All political choices to improve one aspect of well-being involve the opportunity cost of not investing in another aspect. Thus, a well-being policy should make explicit the trade-offs between these components. "Traffic-light" indicators that reveal how well a country is doing in a variety of domains (but not how important these domains are compared to each other) do not resolve this problem. One example of a traffic-light indicator is the United Nations' Millennium Development Goals (and their successor, the Sustainable Development Goals). We also see these indicators used to provide health guidance via labelling on food items (for example, see https://www.food.gov.uk/safety-hygiene/check-the-label: But how important is fat relative to salt or sugars?).

The discussions surrounding the measurement of subjective well-being suffer from some of the same types of issues, although here I am rather more optimistic (as summary measures of subjective well-being, such as life satisfaction, should arguably already include all of the aspects of life that the individual finds important). The Wikipedia caption was "*appearing to experience well-being*"; as I noted above, we can ask people to find out for sure. This is indeed the Wiki definition of subjective well-being as a "*self-reported measure of well-being, typically obtained by questionnaire*". But there is an embarrassment of riches regarding the different questions that we can ask, a number of which were touched upon at the conference. We could, for example, ask respondents about (i) their life satisfaction (which appears in the main individual-level panel datasets used for subjective well-being, such as the British Household Panel Survey/Understanding Society, the German Socio-Economic Panel, SOEP, and the Household, Income and Labour Dynamics in Australia survey, HILDA); (ii) their evaluation of their life relative to the best and worst they could imagine (the Cantril ladder, as used in the Gallup World Poll); (iii) their psychological health; (iv) whether they experience autonomy, engagement, a sense of meaning, etc., in their life (eudaimonia); or (v) whether they feel/felt happy, bored, sad and so on right now or yesterday (i.e., their mood or affect). Is one of these measures our winner, or should we consider some combination of them?

I suspect that we are all ultimately interested in promoting good policy. The choice of the measure of subjective well-being would not matter if the policy variables on which we can act were correlated in the same way with each of the subjective well-being measures (this is the spirit of the empirical analysis in

Clark 2016). But the idea of achieving such a correlation across even a small number of key policy variables is probably fanciful. So what is the “right” measure? We are unlikely to be able to make this determination based on polemics alone; we need to put some objective data into the matrix. There may be mileage in looking at how these different subjective well-being measures are correlated with objective neurological and physiological measures (Although this approach then leads us to the question of which of these latter measures best reflects well-being: Is brain activity more or less important than cortisol levels?).

Which brings me to a second option: namely, using beauty contests or cage fights between the different survey measures to see which one best predicts an objective outcome. Individuals wish to maximise their well-being, and will act in order to do so. Using that assumption as a starting point, let us take some panel data and consider some common decisions that individuals make in order to evaluate which subjective well-being measure best predicts these decisions. Among the behaviours we might examine are geographical mobility, marital separation or quitting one’s job (for example, Clark 2001 and Senik et al. 2012). At the more aggregate level, it has been suggested that life satisfaction is a good predictor of election results (Ward 2020). But is it possible that other subjective well-being measures perform even better?

Let us now imagine that we have reached an agreement about the best measure of well-being to use. But we are still not out of the woods, as for policy purposes, we also need to decide whose well-being matters more. With respect to objective well-being, societies are often described in terms of both their wealth and how unequally this wealth is distributed. Analogously, we should probably go beyond simply looking at the aggregate subjective well-being in a country, and worry about the level of inequality in the country as well. Some policies may be more (or less) effective in addressing well-being or misery, or the well-being of different groups in a society.

A more subtle point is that in a society, the average well-being will increase, and well-being inequality will decline, if that society experiences greater mortality among those who are the worst-off (something of an analogous argument can be made regarding restrictions on immigration). We need to be explicit about who will be present in the society whose well-being we are evaluating, and who is not there, but could have been. This is a general point that applies equally to the average levels of objective individual outcomes, such as income, education and health.

Formulating good policy requires a lot of knowledge. In our research, many of us have asked about the size of the effect of some variable Z on well-being. When making policy choices, we would also need to know whether we can change Z, and how much it costs to change it relative to some other policy variable (some of these considerations appear in Frijters et al. 2020). Moreover, if there is adaptation, the immediate effect of Z on well-being may not continue down the road. And given that there are spillovers between individuals, a policy that makes one individual better off may reduce the well-being of another. Here, the economist’s mantra of revealed preference (if I buy something, or, in general, carry out some action, it must be in my interest to do so) may not be a good barometer of societal well-being. Under adaptation, the fact of wanting something now does not mean that we will still want or enjoy it once we have it. Since we may regret in the future what we are buying/doing now, our current actions are not a good guide to our future well-being (a recent contribution on this point is Odermatt and Stutzer 2019). My current behaviour may be a good measure of my own well-being, but because there are spillovers, it affects the well-being of others as well. When we try to add up individual well-being to produce a national figure, we may find that the purchases made by each individual do not increase aggregate well-being if the positive effect on the individuals making the purchases is potentially cancelled out by the lower well-being of others. Both adaptation and spillovers are taken into account in simple subjective well-being scores: your adaptation to a new job or a new house will be reflected in a gradual drop in well-being back towards its initial level, while any spillovers will appear directly in the well-being scores of those who are harmed by your actions.

Once the most appropriate definition of well-being is sorted out, we can move on to some fundamental questions, such as are some populations or areas doing better than others? If well-being differentials are indeed found, do they reflect the (age/sex/ethnicity/education) composition of the population, or some underlying (un)attractive feature? To what extent do well-being differentials cause individuals to move between areas? And are there spillovers between areas, such that factors like criminality or a good environment in an area affect not only those who live in that area, but those who live nearby as well? This is a fascinating time in which to be carrying out this kind of research, and conferences such as the conference on the link between demography and well-being organised by the Wittgenstein Centre help to pave the way forward.

Acknowledgments

I am grateful to the editors for their feedback, and to Conchita D'Ambrosio for useful discussions.

References

Bruni, L., F. Comim and M. Pugno (eds) 2008. *Capabilities and happiness*. Oxford: Oxford University Press.

Clark, A. E. 2001. What really matters in a job? Hedonic measurement using quit data. *Labour Economics* 8(2): 223–242. https://doi.org/10.1016/S0927-5371(01)00031-8

Clark, A. E. 2016. SWB as a measure of individual well-being. In: *Oxford handbook of well-being and public policy*, eds M. Adler and M. Fleurbaey. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199325818.013.17

Clark, A. E. 2018. Four decades of the economics of happiness: Where next? *Review of Income and Wealth* 64(2): 245–269. https://doi.org/10.1111/roiw.12369

Clark, A. E. and C. D'Ambrosio 2019. Living conditions and basic needs: Evidence from African Countries. *South African Journal of Economics* 87(2): 91–109. https://doi.org/10.1111/saje.12219

Clark, A. E., P. Frijters and M. Shields 2008. Relative income, happiness and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature* 46(1): 95–144. https://doi.org/10.1257/jel.46.1.95

Frijters, P., A. E. Clark, C. Krekel and R. Layard 2020. A happy choice: Wellbeing as the goal of government. *Behavioural Public Policy* 4(2): 126–165. https://doi.org/10.1017/bpp.2019.39

Odermatt, R. and A. Stutzer 2019. (Mis-)predicted subjective well-being following life events. *Journal of the European Economic Association* 17(1): 245–283. https://doi.org/10.1093/jeea/jvy005

Sen, A. 1985. *Commodities and capabilities*. Amsterdam: North-Holland.

Senik, C., H. Stichnoth and C. Guven 2012. You can't be happier than your wife. Happiness Gaps and Divorce. *Journal of Economic Behavior and Organization* 82(1): 110–130. https://doi.org/10.1016/j.jebo.2012.01.003

Streeten, P. 1981. *First things first: Meeting basic human needs in developing countries*. New York: Oxford University Press.

Townsend, P. 1979. *Poverty in the United Kingdom*. Harmondsworth: Penguin Books.

Ward, G. 2020. Happiness and voting: Evidence from four decades of elections in Europe. *American Journal of Political Science* 64(3): 504–518. https://doi.org/10.1111/ajps.12492

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 29–37

Sustainable human wellbeing: What can demography contribute?

Wolfgang Lutz[1,*]

Abstract

This note considers the role that demography as a discipline can play in addressing some of the key questions in the context of human wellbeing and sustainable development. Starting with the wellbeing function of sustainability science that tries to explain an indicator of human wellbeing as being determined by a set of capitals and explanatory factors, it gives an example of how the constituents of such a wellbeing indicator can be combined based on a demographic approach. It also highlights how a broadened view of demographic methodology that goes beyond the conventional focus on age and sex alone can help to make demography more relevant for studying the key challenges of humanity.

Keywords: sustainability science; multi-dimensional demography; years of good life; wellbeing indicator

1 Introduction

The improvement of human wellbeing has been an essential human aspiration for as long as written records of humanity exist. Aside from some ancient philosophers, the focus of most scholars until the time of the Enlightenment was on the collective wellbeing of a kingdom or a certain community. Since then, the idea of human rights as individual rights has given rise to a more individualistic understanding of wellbeing in modern Western cultures (Russell 1945). Today, from an international perspective, more collectivistic and more individualistic understandings of wellbeing tend to co-exist. Demography, as a discipline that focuses on population-level

[1]University of Vienna, International Institute for Applied Systems Analysis (IIASA), Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

[*]Correspondence to: Wolfgang Lutz, lutz@iiasa.ac.at

DOI: 10.1553/populationyearbook2021.deb03

changes as well as their micro-level foundations in terms of individual characteristics, is well-positioned to bridge these two understandings.

Demography is also well-equipped to cover the temporal dimension of changes in wellbeing over time. The naïve expectation that wellbeing has been increasing and will continue to increase more or less automatically for all segments of the global human population for the foreseeable future has become untenable, particularly considering what humanity has done to the environment and our natural life support systems. The very successes societies can claim in improving human living conditions through the extensive use of fossil energy, material consumption and land-use changes with associated losses of biodiversity have resulted in planetary-scale environmental changes that threaten our life support systems, and, thus, the wellbeing of future generations. In response to these insights, the focus of attention in demography has recently been shifting away from assessing only past and current conditions, and toward the notion of sustainable human wellbeing, which takes future challenges into account (Lutz et al. 2021b).

Today, the adjective "sustainable" is used excessively, and has many different connotations. But above all, it explicitly introduces a temporal dimension, and thus frames the phenomenon it defines as something that has the potential to go on into the future, for as long as we can imagine. In the context of human wellbeing, the term shifts the focus away from assessments of past improvements and current levels of wellbeing, and toward likely future changes, with the goal of maintaining or further improving the level of wellbeing, even when potential threats – including threats related to political, economic, pandemic or environmental problems – are also factored in. Since future trends cannot yet be observed, researchers must rely on scientific models and scenarios to assess the likelihood that alternative future trends will occur. Since demography has more quantitative models for analyzing the longer-term future to offer than most other scientific disciplines, it is highly relevant to the study of sustainable human wellbeing. In particular, models of multi-dimensional demographic dynamics have been producing alternative scenarios of future population trends that go beyond the conventional demographic focus on changing age structures by also considering changing structures in other relevant demographic characteristics, such as educational attainment, labor force participation and health status (Lutz and KC 2011). The Shared Socioeconomic Pathways (SSP) scenarios, which are already widely used in the global environmental research community (Riahi et al. 2017), can be translated into demographic scenarios, and applied to project alternative multidimensional demographic trajectories.

The combination of the words "sustainable" and "development" in the currently ubiquitous notion of sustainable development has evolved into an effort to combine the more traditional focus on socioeconomic development with more recent concerns about environmental degradation and climate change. The classic definition by the Brundtland Commission in 1987 (World Commission on Environment and Development 1987) refers to a pattern of development that meets the needs of present generations, but without compromising the rights of future generations to fulfil their needs. Three decades later, in a preface to the Global Sustainable

Development Report 2019, which was written by a group of 15 independent scientists (including the author) who were charged by the UN Secretary General to assess the progress made toward meeting the Sustainable Development Goals (SDGs), Gro Harlem Brundtland revisited this definition, and highlighted some key aspects from today's perspective. The first of these aspects listed as priorities for sustainable development is "Strengthening human wellbeing and capabilities" (Brundtland 2019).

This focus on human capabilities and intergenerational aspects of human well-being again places demography on center stage as the discipline that, through its analytical focus on cohort changes over time, is very well-positioned to conduct such intergenerational analyses using a quantitative approach. As a key driver of sustainable development, human capabilities (Bengtsson et al. 2018) can be quantitatively captured, and the changes in these capabilities over time can be modelled through the multi-dimensional demographic approach of modelling human capital trends, while explicitly considering education, health and labor force participation, in addition to age and sex. These characteristics cover all of the essential aspects of human capabilities – with labor force participation as a proxy for not being in absolute poverty – following the approach of Sen (1997). A demography that goes beyond the rather narrow conventional focus on population size and age structure, and that explicitly includes education, skills and health status as the key aspects of capabilities, can assume a key role in this field of studies (Lutz and KC 2011; Lutz 2017; Lutz et al. 2021a). Finally, as has been shown at this conference, demographic methods, data and knowledge can contribute a lot to the definition and the measurement of human wellbeing indicators.

Given that demography has so much to contribute to all of these key dimensions of sustainable development and the study of sustainable human wellbeing, it is rather surprising how few demographers have taken the opportunity to enrich this field through their expertise.

In the following, I will give an example of how demographers can make a contribution to the very core of sustainability science by also applying core demographic methods, such as the life table and associated tools. This example illustrates that to become more relevant for the rest of the world, demography does not have to evolve into some form of hyphenated demography, such as social-demography or bio-demography. Instead, by broadening its unique methodological core to encompass multi-dimensional analysis, demography can be applied to some of the most pressing questions of humanity, which will, in turn, boost the relevance of the discipline.

2 Demography and sustainability science

In the remainder of this brief commentary, I will highlight one possible entry point of demography into what has come to be called "sustainability science," and I will introduce the approach on which I currently work with a team at the Wittgenstein

Centre as part of an ERC Project entitled "The demography of sustainable human wellbeing." At the same time, in this short text, I will try to provide a theoretical framing for the newly developed Years of Good Life (YoGL) indicator, which has been used and applied in three of the papers in this volume.

Broadly defined, sustainability science focuses on the interactions between natural and social systems in the most comprehensive way possible. Nonetheless, in the words of one of its founders, it still suffers from the "plurality of purpose in characterizing and measuring sustainable development, and the confusion of terminology, data, methods and measurement" (Parris and Kates 2003). Because of this extremely broad and often diffuse focus, it is also sometimes called an "inter-discipline" (Stycos 1989). An explicit aim of sustainability science is to cross all disciplinary boundaries if doing so helps to address one overriding and all-important question: What are the conditions that assure long-term human thriving and wellbeing in harmony with nature? In this context, the distinction between strong and weak sustainability is sometimes made, depending on the views on the substitutability of natural capital. According to the most extreme view of strong sustainability, nature needs to be protected for its own sake, preferably without any disturbance by humans. If such a position is taken, then a world completely without human beings would, ultimately, be seen as most desirable. In this view, demography would have little to offer (except for providing demographic scenarios of a one-time mortality rate of 1.0, or a somewhat softer alternative scenario of zero fertility over the next half century).

When advocating for the – in my view – more meaningful concept of weak sustainability, which is more anthropocentric in the sense that it argues that nature needs to be preserved in order to serve as a human life support system, then sustainable human wellbeing becomes the ultimate sustainability criterion. In this context, the specification of an analytical function has been broadly agreed upon in which human wellbeing W is being generated through a combination of human, manufactured and natural capital, as well as institutions and knowledge (Clark 2012).

$$W = f(C_i, I, K)$$

W is "human well-being" (intra- and intergenerational); and
C_i are "capital assets" (from which services flow).

This can be further decomposed into:

- C_m is "manufactured capital" (factories, homes, roads);
- C_h is "human capital" (population, health, education);
- C_n is "natural capital" (ecosystem and their services);

I is "institutions" (laws, rules, norms, expectations); and
K is "knowledge" (scientific, practical, innovation).

While this specification offers a promising way forward, there are still huge challenges in empirically assessing the different elements of this function and their interactions. Most fundamentally, one has to decide whether to focus the measurement on the left- or the right-hand side of this function; or, in other words, on the constituents or the determinants of human wellbeing. Much of the recent literature in this field has used economic approaches to quantify the determinants on the right-hand side. Here, the notion of inclusive wealth has become prominent as a broadening of conventional concepts of economic wealth and material capital to include other capitals, particularly the value of natural capital (Dasgupta 2001). While there is a rich body of economic theory and a well-equipped economic toolbox for such approaches, a fundamental problem that arises in their application is that it is hard to estimate the relative sizes of the effects of these different determinants on the right-hand side if there is no quantification of W on the left-hand side.

My ERC project focuses on the left-hand side of this function. The aim of the project is to come up with a tailor-made indicator of human wellbeing that is fit for use in estimations of this analytical function. We have specified important criteria that any such indicator should meet; and, as we will see, demography plays an important role in almost all of these criteria.

3 An indicator for sustainable human wellbeing

In recent years, there has been a mushrooming of wellbeing and quality of life indicators, with 31 of them described in the appendix of Lutz et al. (2021b). A common feature of most of these indicators is that they are designed to go beyond the outdated focus on GDP per person as an indicator of general wellbeing and social progress. Many of them are composite indicators that try to combine different aspects that are assumed to be important for wellbeing. Here, the question of how those aspects should be combined and weighted is difficult to answer. The solutions offered range from applying fixed equal weights, as is done in the well-known UNDP Human Development Index (HDI); to leaving the weights entirely up to the user, as is done for the 11 aspects covered by the OECD Better Life Index, in which users can freely choose which aspects they value more. Thus, in both approaches, the weighting is more or less arbitrary. In the new tailor-made indicator, this problem should be avoided.

There are other important criteria that such a new tailor-made indicator for W should meet. Most importantly, it should be based on **characteristics of people** that can be flexibly aggregated to apply to **subpopulations**. This is a deeply demographic approach, because any such indicator that refers to the wellbeing of people should be built from the bottom up based on individual characteristics that are observable and measurable for individual people. The use of such an approach also makes it possible to calculate the indicator for groups of people other than just national populations. A focus on subpopulations is essential for answering many of the important questions that are asked in sustainability science, and that also tend to be important for ethical and political considerations: e.g., how does wellbeing differ by gender; how does it differ among various ethnic or socioeconomic groups in a population; and, how does it differ between residents of urban and rural populations or different administrative districts or other geographic units? The criteria used to assess the wellbeing of subpopulations cannot include any indicators that only exist at the national level.

In addition, such a tailor-made indicator needs to have a meaning in its absolute value that makes the indicator **comparable over time and across subpopulations**. For the indicator to be useful for comparing the wellbeing of certain populations at two different points in time, and for determining whether the wellbeing of these populations has improved or deteriorated, it must have a meaning in its absolute value. Thus, the indicator cannot simply be defined on a relative scale that depends on other time-varying indicators. As an example, the life expectancy component of the HDI is defined as a fraction of the maximum national life expectancy observed in any given year. Hence, when this mortality index (given in the form of a fraction) is compared for a given population between, say, 1990 and 2015, it is impossible to determine whether survival conditions in this population actually improved over time, and, if so, by how much. Thus, in its relative form, the mortality index can only show whether the given population improved its standing relative to that of the country with the highest life expectancy.

Furthermore, such an indicator needs to describe some highly desirable state as an ultimate end goal that has the potential to **be universally shared across all cultures** (Lutz et al. 2021b). The use of any wellbeing indicator only makes sense if there is near consensus that it describes a highly desirable target. The aspiration is to capture the single most important ultimate end goal that people with very different orientations, values and cultural backgrounds already to subscribe to. It should describe some aim that a Wall Street broker, a Buddhist monk and a member of a traditional nomadic tribe would all, at least in principle, find most desirable. While the actual acceptance of the indicator as an ultimate end goal will be empirically tested in different settings later in this project, at this point, we can only *ex ante* exclude some candidates for such an indicator that would clearly be rejected by certain groups. Avoiding the premature death of oneself or a person one cares for seems to be a good candidate for such a universally accepted goal. Here, demography can again contribute a lot to the measurement of this end goal, including the use of subpopulation-specific mortality rates and life table-based methods, such as the Sullivan method, to link the person-years lived at any age to other relevant characteristics.

Finally, ideally, it should be possible to interpret such an indicator through some **real-life analogy,** rather than simply as an abstract index (Lutz et al. 2021b). Thus, the hope is that the indicator will provide not only abstract index values, but outcomes that have a real-life analogy, just as life expectancy gives the number of years that a person can expect to still be alive. For an indicator of wellbeing, it is more attractive and intuitive to say, for instance, that in population *x*, the average person has 66 years of good life, while in population *y*, s/he only has 54 years of good life; rather than to say that a wellbeing index has values of 0.753 and 0.683, which are numbers that people cannot relate to. Here again we can draw on the strength of demographic life table methods, which can be used to calculate the number of years a person is expected to live in good health.

The application of these criteria to the definition of the tailor-made indicator Years of Good Life (YoGL) and its specific quantification will not be discussed here, but will be shown in three of the following papers (Reiter and Spitzer 2021; Striessnig et al. 2021; Buathong et al. 2021). A comprehensive description of this concept has been published recently in PNAS (Lutz et al. 2021b). Here, it is sufficient to say that in order to meet these criteria, a clear hierarchy of the dimensions that will be covered in the YoGL indicator has been established on theoretical grounds. First and foremost, the indicator assumes that survival is the most essential prerequisite for enjoying any quality of life. Indeed, it seems odd that most existing wellbeing indicators do not explicitly include the length of human life. If avoidance of premature death is not explicitly factored into the criterion, then the average value of any indicator can always be improved by eliminating (killing) the members of the population who have lower values for this indicator, as doing so would increase the average value for the survivors. For this reason, it seems imperative to explicitly include longevity in the indicator.

However, since mere survival is not considered good enough to ensure wellbeing by many people, in a next step, the YoGL indicator goes on to define good years of life as the years in which people live above a minimum level in terms of both objectively observable conditions and subjectively assessed life satisfaction. At a third level, the objective conditions that are used to define capable longevity are further broken down into three dimensions: being out of poverty, not being cognitively disabled, and not being physically disabled. All of these aspects are integrated and brought together using a uniquely demographic method called the Sullivan method, in which the person-years in a life table are multiplied by the proportions of men and women at each age and for each subpopulation who have values that are above these specified minimum thresholds. The indicator can then give the average number of years a person can expect to be alive and above minimum levels in terms of subjective life satisfaction, and to be out of poverty and in good cognitive and physical health. We call these years the Years of Good Life (YoGL), and their calculation is based on a genuinely and uniquely demographic approach.

In the conclusion of their PNAS paper, Lutz et al. (2021b) also speculated about the possibility that YoGL could be used in the future as a kind of demography-based currency for assessing costs and benefits. They wrote: “An indicator like YoGL also has the potential to become a broadly used currency in which the costs and benefits of certain developments and actions can be expressed, complementing assessments based on purely monetary units. For example, the social costs of carbon could potentially be assessed in terms of years of good life lost among future generations, rather than only in some dollar terms” (Lutz et al. 2021b). Thus, demographic units of measurement could even be used to replace units of money in assessments of sustainable development and human wellbeing.

References

Bengtsson, S. E. L., B. Barakat, R. Muttarak, E. B. Kebede and W. Lutz 2018. *The role of education in enabling the sustainable development agenda*. Abingdon, UK and New York, USA: Routledge. https://doi.org/10.4324/9781315142708

Brundtland, G. H. 2019. Prologue. In *Global sustainable development report 2019: The future is now - science for achieving sustainable development*, ed. Independent Group of Scientists appointed by the Secretary-General, xv–xvii. New York: United Nations.

Buathong, T., A. Dimitrova, P. M. M. Vicerra and M. Chimmamee 2021. Years of good life: An illustration of a new well-being indicator using data for Thailand. *Vienna Yearbook of Population Research* 19. https://doi.org/10.1553/populationyearbook2021.dat.1

Clark, W. C. 2012. Sustainability science: Its origins, evolution and challenges. Paper presented at IIASA 40th Anniversary Conference: Worlds Within Reach: From Science to Policy, 24–26 October 2012. http://conference2012.iiasa.ac.at/person.html?code=clark&scode=s2s7

Dasgupta, P. 2001. *Human well-being and the natural environment*. Oxford, New York: OUP Oxford.

Lutz, W. 2017. Global sustainable development priorities 500 y after Luther: Sola schola et sanitate. *Proceedings of the National Academy of Sciences* 114(27): 6904–6913. https://doi.org/10.1073/pnas.1702609114

Lutz, W. and S. KC 2011. Global human capital: Integrating education and population. *Science* 333(6042): 587–592. https://doi.org/10.1126/science.1206964

Lutz, W., C. Reiter, C. Özdemir, D. Yildiz, R. Guimaraes and A. Goujon 2021a. Skills-adjusted human capital shows rising global gap. *Proceedings of the National Academy of Sciences* 118(7): e2015826118. https://doi.org/10.1073/pnas.2015826118

Lutz, W., E. Striessnig, A. Dimitrov, S. Ghisland, A. Lijad, C. Reiter, S. Spitzer and D. Yildiz 2021b. Years of good life is a wellbeing indicator designed to serve research on sustainability. *Proceedings of the National Academy of Sciences* 118(12): e1907351118. https://doi.org/10.1073/pnas.1907351118

Parris, T. M. and R. W. Kates 2003. Characterizing and measuring sustainable sevelopment. *Annual Review of Environment and Resources* 28(1): 559–586. https://doi.org/10.1146/annurev.energy.28.050302.105551

Reiter, C. and S. Spitzer 2021. Well-being in Europe: Decompositions by country and gender for the population aged 50+. *Vienna Yearbook of Population Research* 19. https://doi.org/10.1553/populationyearbook2021.res4.1

Riahi, K., D. P. van Vuuren, E. Kriegler, J. Edmonds, B. C. O'Neill, S. Fujimori, N. Bauer, K. Calvin, R. Dellink, O. Fricko, W. Lutz, ... and M. Tavoni 2017. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change* 42: 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Russell, B. 1945. *History of Western Philosophy*. New York: Simon & Schuster.

Sen, A. 1997. Editorial: Human Capital and Human Capability. *World Development* 25(12): 1959–1961.

Striessnig, E., C. Reiter and A. Dimitrova 2021. Global improvements in years of good life since 1950. *Vienna Yearbook of Population Research* 19. https://doi.org/10.1553/populationyearbook2021.res1.2

Stycos, J. M. (ed) 1989. *Demography as an interdiscipline*. New Brunswick, NJ: Transaction Publishers.

World Commission on Environment and Development 1987. *Our common future*. Oxford: Oxford University Press

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 39–64

WELLBYs, cost-benefit analyses and the Easterlin Discount

Paul Frijters[1,*]

Abstract

The current practise of cost-benefit analysis in Western countries consists of a collection of various incompatible ideas and methodologies to obtain replicable numbers for the costs and benefits of major public spending plans. This paper describes the main elements of the dominant methodology, which combines consumer and producer surplus, price-taking, government-inputs-as-outputs, hedonic pricing of externalities, and the issue-specific use of partial or general equilibrium thinking. The paper then discusses how that methodology can be augmented and partially replaced by looking at how prospective policies would change the total number of WELLBYs (life satisfaction-adjusted years of life) of the population. The ability of the WELLBY methodology to address complex externalities is illustrated by the Easterlin Discount, which is a proposed reduction factor of 75% on all estimates of private consumption benefits to offset the envy caused in others.

Keywords: cost-effectiveness; wellbeing; consumer surplus; Easterlin Discount; public advocacy

1 Introduction

Cost-benefit analysis (CBA) has been used for over a century in Western countries to assess whether large projects are worth doing. Dupuit (1848) is said to have initiated this approach in the 19th century when he proposed a specific methodology for looking at the social spending case for building a new bridge in France.

In the UK, the ministry of finance, known as the Treasury, has stated that other ministries and agencies should use some form of CBA when seeking public funds. The Treasury Green Book (2018), which has been maintained and updated by departments for decades, lays out the basic methodology for doing so. Moreover,

[1]Department of Social Policy, London School of Economics, London, UK

[*]Correspondence to: Paul Frijters, p.frijters@lse.ac.uk

DOI: 10.1553/populationyearbook2021.deb04

dozens of additional publications by individual departments specify the CBA methodologies that should be applied for particular purposes. For a general text on cost-benefit analysis as it is applied in broad terms throughout the world, see Boardman et al. (2017, 4th Edition).

The reality of how funding is approved is, of course, not that simple, as many large-scale policies are driven by politics. Nonetheless, it has been a long-standing aim of the UK Treasury to move increasingly towards a world in which large public expenses have to be justified on the basis of some measure of what the gains and the costs of such policies are. The same pressures can be found in most other Western countries, as the approval of infrastructure projects in particular is almost always preceded by a cost-benefit analysis.

When performing a cost-benefit analysis, it would, ideally, be possible to understand everything about the world, and to thus be able to say with perfect foresight what the investment would change. Again, this is too much to ask. The practise of cost-benefit analysis resembles a leaking bucket cobbled together from different intellectual traditions and pragmatic considerations.

Nevertheless, this leaking bucket is the main tool our societies have to assess what matters to the government; i.e., its goal function. A point that is seldom noticed is that the rules of thumb and habits of accounting represented in government cost-benefit analysis serve as *de facto* measures of forms of welfare that our societies strive to maximise. Loosely speaking, it is the notion of economic surplus that Jules Dupuit envisaged.

As a measure of what matters, economic surplus is wholly inadequate. Many things that people care about and spend resources on, such as their children and the quality of the environment, are not naturally part of economic surplus. Thus, CBA analyses have adopted numerous ad hoc assumptions in an effort to capture some of its value. Many things that governments spend resources on are not based on any CBA, nor could they be determined by a CBA. For example, questions like how large the army should be or how the tax authorities should be organised cannot be determined by a CBA. While models that calculate the costs and benefits of these issues can be envisaged, they do not exist in reality. In weighing these questions, political judgment and organisational learning are relied upon instead.

In this piece, I first provide a simple account of what cost-benefit analysis entails in Western countries, and then propose an alternative measure of welfare: the stream of WELLBYs of the entire population of the country. A WELLBY is one unit of life satisfaction on a 0–10 scale for one person for one year (introduced in Frijters et al. 2020). While I advocate using the WELLBY as a replacement for economic surplus, I show how it can also be used as a patch for some of the problems with existing CBA analyses. In particular, I introduce the idea of the Easterlin Discount, which can be applied to all private economic surplus in order to account for consumption externalities. The idea is that the private consumption of someone else creates jealousy and envy in others, which leads to the need to only count a proportion of the private consumption benefit as social benefits. The proportion that needs to be discounted is called the Easterlin Discount, introduced in Frijters and Krekel (2021).

To explain the practise of CBA in the UK, I first set up a standard CBA calculation loosely based upon the extension of Heathrow in the UK (which was actually proposed as a project by the Airports Commission 2015). I use this example to explain the leaky bucket of the current approach to CBA. Then, I introduce the alternative, the WELLBY, and show how it can be used to either replace the bucket entirely, or as a patch for particular problems.

2 The practice of CBA in the UK

Here, we refer to the practise of CBA in the UK. Each Western country has subtly different habits in the use of CBA. Hence, the observations we make about the UK will not perfectly hold for other countries.

Consider Table 1 below, which is partly populated by the actual numbers from a government-commissioned report by the Airports Commission (2015), and partly consists of made-up entries (regarding health quality-adjusted life years, QALYs) that are included to illustrate further aspects of the practise of CBA.

This table illustrates four elements involved in standard cost-benefit analyses that are, on closer inspection, unexpected and incompatible. Let us examine these four elements in turn.

2.1 Consumer surplus

The big positive element in the first column, worth £54.8 billion, is the supposed consumer surplus of the additional runway at Heathrow. While this number comes from a model, at its core, the calculation comes from the estimation of demand curves. To remind the reader: consumer surplus is the difference between the willingness to pay and the price of something, and thus indicates how much additional money a person would have been willing to pay to buy something.

What is crucial to realise is that consumer surplus is not actually observed. It is not the production or the price of a traded good; instead, it is some notion in the head of the consumer of the utility he or she would enjoy by consuming a good, versus the price of that good. Hence, it is a psychic good that is not part of GDP or some other notion of national income, but is nonetheless valued in terms of money.

Thus, the inclusion of consumption surplus in cost-benefit analyses indicates that the government economists who run CBA believe in unseen pleasure and the value of traded goods, which they measure by figuring out what individuals were willing to pay for a good versus what they actually paid for it.

This is also the approach that "CBA people" prefer to use to value everything, whether it is the environment, sex, noise or democracy (Boardman et al. 2017 or

Table 1:
Appraisal results for the Heathrow Airport Northwest Runway scheme, present value (£billion, 2014 prices). 50% Easterlin Discount and rearranged

	With or without Easterlin Discount (ED)	
Appraisal results	**Without ED**	**With ED**
Monetised		
Consumer surplus	54.8	27.4
Producer surplus	−38.4	−19.2
Scheme capex and private paid surface access cost	−13.6	−6.8
Primary surplus change	2.8	1.4
Delays	1	1
Wider economic impacts post-tax	6.9	3.45
Noise	−1	−1
Air quality	−0.8	−0.8
Carbon emissions	−0.9	−0.9
Biodiversity	0	0
Health costs (100,000 QALY)	**−6,0**	**−6,0**
Government paid surface access costs	−2.5	−2.5
Taxes (40%) from wider economic impact	4.6	4.6
Government revenue	1.8	1.8
NPV (net social benefits and PVC)	**5.9**	**1.05**
Non-monetised		
Surface access	Light green	
Quality of life	Neutral	
Community	Light red	
Place	Light red	
Local economy	Dark green	
Water and flood risk	Light red	

Note: NPV = net present value, PVC = Present Value Cost.
For the moment, we ignore the distinction between the first and the second column, because the Easterlin Discount will not be explained until much later. Instead, we look purely at the first column (the "without ED" column).

Campbell and Brown 2016): i.e., they look for some way to measure what people's willingness to pay for something is (often by asking them about their willingness to pay in a survey), and then they look for an appropriate price so that they can calculate the consumer surplus involved.

There are many hidden value judgments in this practise. Let me list four of the main ones:

1. The economic surplus of every person is added up as if who does the consuming was unimportant. This simple process of adding up is used partly by necessity, since it is often not known who buys a good. Hence, it is not possible to "correct" for how rich or poor the person is. This means, however, that the rich person generates far more consumer surplus than the poor person, and thus counts far more than the poor person in cost-benefit analyses. It is the equality of every pound, not of every person. Thus, economic surplus is deeply anti-democratic. Moreover, it does not conform to any normal economic notion of "utility", i.e., the amount of psychic benefit, because that utility would then, of course, have to be measured. We can therefore conclude that the concept is based largely on convenience, and the need to get a number.
2. It is assumed that consumers know the full costs and benefits of their consumption, such that everyone who consumes is enjoying a surplus, and there are no systematic mistakes. Hence, there is an untested presumption of consumer rationality and responsibility in CBA: i.e., that there are no regrets of any kind.
3. The consumption of a good is seen as a stand-alone purchase that is not connected to other purchases, such as when an airline ticket is only part of a whole consumption package ("the trip"). This stand-alone assumption means that the surplus of the whole trip is assigned to an individual element of it, while the same surplus can also show up in different calculations involving elements of the package (like the hotels and the attractions consumed on a trip). Thus, it is assumed that consumers are not able to obtain the same economic surplus with some other purchase; i.e., that there are no other goods in the economy on which a surplus is based. The partial equilibrium assumption embedded in models of a single sector implies, for example, that there is no opportunity surplus that is foregone when a consumer buys an airline ticket. There is a deliberate one-sided blindness to this assumption that inflates the supposed social value of consumption increases, and that is again basically made for the sake of convenience.
4. It is assumed that there are no consumption externalities between individuals: e.g., that no one else loses or gains if air trips are made. This is unwarranted, because individuals will usually travel somewhere to meet other people, thereby creating a positive externality for those other people. A husband who is travelling to meet his wife will, presumably, create surplus through his wife as much as through himself. By ignoring such externalities, social life is implicitly assumed to be of no importance. This observation also applies to negative consumption externalities, like jealousy. It is, for example, assumed that no one is made to feel inadequate if other, richer people fly around all the time. This, too, is unlikely.

Thus, I argue that the practise of estimating and counting consumer surplus has no obvious "economic foundations", and that it does not take into account social life, issues of inequality, general equilibrium or limited rationality.

Nonetheless, CBA is a very powerful approach that can assign numerical values in many cases. For anything with observed market prices and some history of demand curves, CBA can be used to estimate changes in consumer surplus if capacity is increased. That is why consumer surplus is a cornerstone of CBA around the world: it is a concept that can be applied to generate numbers. Whether the numbers are reasonable or nonsensical cannot be known unless we are able to measure the underlying pleasure that consumers derive from the consumption of the goods in question. These numbers meet the needs of a CBA industry that has to come up with non-negative numbers that demonstrate how much an expansion of the consumption possibility space would benefit consumers.

Another important advantage of the notion of consumer surplus is that the procedure used to obtain a given number is replicable, and would yield the same number if someone else followed the same procedure. This makes it possible to defend the number on the basis of a list of assumptions included in the model that estimates the demand curves and other aspects of the problem. Thus, the number is defensible and replicable, which are very important qualities from the perspective of bureaucracies.

We will encounter this dynamic frequently when considering how CBA is applied. The driving logic of many aspects of CBA is the need to come up with a number that others would also come up with if they used the same "accepted procedure".

Let us now turn to the next number in Table 1.

2.2 Producer surplus

The next number in Table 1 is the anticipated change in producer surplus, estimated as a negative value of £38.4 billion. This number seems more straightforward, as it represents the anticipated reduction in the profits of airlines due to the anticipated lower prices following an increase in the possible number of flights to Heathrow. This figure is a combination of lower profits on existing flights and low profits on additional flights.

On closer inspection, however, this number is indefensible, primarily because the airline industry is a highly competitive sector in which almost no profits are made (Morrison and Winston 2010). If there are no profits, the claim that £38.4 billion can be subtracted from profits is false. It is worth noting that if airlines were truly forced to reduce their earnings by that much, they would either go bankrupt or stop flying to Heathrow. Thus, this number does not make much economic sense.

Though indefensible, there is an important effect of the claim that consumer surplus is highly positive whilst producer surplus is highly negative: it makes the proposed runway look like a good deal for consumers. Thus, consider the conclusion of the Airports Commission (2015) report, which makes the following observation on the basis of these numbers:

> "Against the objective of maximising **economic benefits** and supporting the competitiveness of the UK economy the Heathrow Airport Northwest Runway option performs most strongly, generating £69.1 billion of benefits, compared to £58.7 billion from the Extended Northern Runway scheme and £60.1 billion from the Gatwick Second Runway".

Hence, the report notes that this figure is based almost entirely on the supposed consumer surplus, which it calls an "economic benefit", even though it is not a number that refers to GDP. Thus, it puts weight on what is a metaphysical number (an unseen psychic benefit presumed to be revealed via willingness to pay) that is more closely aligned with psychology than with economics.

Deep within the report we find the information on what is actually likely to happen following the airport expansion: namely, that prices will not drop at all for consumers because it is anticipated that the regulator will allow price increases. The report thus states on page 89 that "the airport scheme would be financed privately and offset via rising aero charges levied on the passengers and users of the airport (**not accounted for in this calculation**)".

Thus, the profits of the airlines will not suffer after all. Instead, the airport will simply attract more business, which is why the airport operator commissioned the modelling, and is pushing for this additional runway. Heathrow commissioned the modelling that led to the high estimated consumer surplus. And yet it appears that the airport operator wants to convince the reader that the producers are going to lose nearly £40 billion.

So the headline numbers have been twisted to appear as if the consumers, rather than the producers, will win out. This is done by not using the actual anticipated prices in the calculations. It is a straightforward case of cheating: one cannot say in the main conclusions that the airport expansion is good for consumers, and then say in a passage hidden deep in the report that one actually expects a totally different consumer surplus (via a higher price) than the surplus that is used to come up with the main conclusions.

Thus, on closer inspection, we can see that both the consumer surplus and the producer surplus are essentially hoaxes; that is, they are the products of political posturing meant to convince politicians and consumers of the advantages of building a new runway. While the Airport Commission is supposedly independent and on the side of the British public, it is clear that it has been captured by vested interests. This is hardly surprising, given that cost-benefit analyses are typically paid for by stakeholders with the expectation of private gain. Bureaucracies want CBA to be performed because they want it to appear as though the decisions are made based on scientifically credible numbers. Meanwhile, interest groups want projects from which they expect to earn a profit to go ahead. Those interest groups then simply make up the numbers the bureaucrats need, while pursuing their own objectives.

Similar objectives can be discerned when we look at the "scheme CAPEX..." number, which supposedly denotes the direct costs producers would have to pay to build the additional runway. In addition to being inflated, this number hides the fact that the government usually pays a substantial share of these costs. The government may, for example, help to cover the costs of building the infrastructure and of buying the land. This point was also conveniently hidden in the report, because in the appendices of the report it is revealed that the government usually pays half the access costs.

The strategic forms of miscommunication detected in this Airport Commission report are not, of course, replicated in all CBA analyses. But the way in which changes in producer surplus are calculated in the report is indicative of how this is done elsewhere. The important elements of this calculation are as follows:

1. The prices and costs for producing something are taken as given, and are thus not subject to negotiation. This often means that the government is expected to pay high prices for something it could get much cheaper. The producer surplus calculations are based not on the lowest price the government could achieve, but on the price demanded by private companies. This might have been reasonable when a town was building one bridge in France in 1848, but it is not reasonable today. The government is not a price taker, and should not make the mistake of acting like one.
2. In most cases, there are many elements of the costs of producing something that involve the government, either as the source of the money or as the regulator. For instance, the runway cannot be built unless the government uses its power to force everyone living on that land to sell their property for a particular price. Moreover, a wide range of utility and infrastructure services are involved in an airport expansion, such as water and electricity supply and transport to the extended airport. Providing these services also involves a great deal of government investment and regulation. Judgment calls and agreements on all these services are also part of the actual costs and benefits, but unless the entity drawing up the CBA is very honest and independent, it is very easy to understate the true costs of the project to the government: only real experts would notice the subterfuge.

Thus, the producer surplus line in this table, which is really the main "cost" line, is, in fact, a reflection of the outcomes of lengthy negotiations and choices by governments in the coming decades. One of the main drawbacks of CBA is that it does not automatically make negotiations and regulation part of the process by, for example, informing the government of the cost savings that can be achieved via optimal negotiations and regulations.

The very notion of calculating a cost and a benefit has within it the presumption that the costs can be known. However, these costs cannot be known if they are the outcomes of strategic negotiations with large private parties. Here again, we see that CBA is not aligned with many aspects of economics. We also find that simply taking the prices named by producers as a given makes for verifiable numbers. Although this is not standard economics and makes for irrational government, this approach is a more easily replicable bureaucratic procedure.

Let us now go on to a third important aspect of CBA, which is how it treats government revenues versus private surplus.

2.3 Government revenue and private surplus

There are four lines in this table that involve government in relation to the private sector. Let us take them in turn.

First, we look at taxes (£4.6 billion) and government revenue (£1.8 billion). The crucial point to note is that these figures are simply added up together with the private economic surplus, regardless of whether they represent the consumer surplus or profits.

While this may seem like an innocuous thing to do, especially as it is the standard used in nearly all government CBA, it is actually quite strange. The presumption is that the value of what the government does is exactly equal to how much revenue the government collects via taxation and other sources (such as from the sale of land).

Hence, it is assumed that the government generates no surplus of any kind. The implicit model of the world is thus one in which the government is practically useless. Indeed, since consumers would enjoy a greater consumer surplus if they had additional resources, the government is assumed to be worse than useless, as the taxes paid to the government represent net losses to society – at least according to CBA.

Obviously, this assumption is not based on economics. So where does it come from?

It seems to come from GDP accounting, which also adds up government spending and private consumption as part of national consumption. Within GDP, which does not measure the pleasure from consumption, but simply adds up everything with a market value, government activity is valued as equal to its costs (inputs are outputs). Thus, the CBA practise seems to consist of just adding private economic surplus together with government revenue.

Yet, as we saw in the case of consumer surplus, CBA does not measure the change in GDP, because it would otherwise not include the difference between the willingness to pay and the price as part of its calculation.

Note also that it makes life easy for the CBA calculator to simply assume that the value of any change in government revenue is exactly equivalent to that change: i.e., there is no need to understand what government does and might or might not add. Once again, we see that CBA is a highly verifiable way to come up with a number for the value of government activity.

But is it remotely reasonable to assume that government spending would generate no surplus? Basically, the answer is: “Not at all”. Governments run many monopolies that make it a relatively cheap provider of services. The education and health services provided by the government are far less expensive than those provided by private firms, and they frequently deliver better outcomes. There is, for example, evidence that private health provision in the US is more expensive and has worse outcomes than European systems (Papanicolas et al. 2018).

We also have ample evidence that government services are highly effective when they are first set up. When Turkey rolled out its General Practitioners (GP) services, each doctor was estimated to save about half a life for every year of practise (Cesur et al. 2017), which is a hugely positive surplus. Economic historians similarly credit the large increase in longevity to government investments, including investments in sewage, inoculations, food safety and cheap transport (Smith 1999).

Is there any empirical evidence that government spending is useless? Most estimates show that government spending is, on the contrary, highly effective. A key number that can be cited in making this case is that according to its internal calculations, the UK government estimates that it produces a year of high quality life for every £15,000 spent on health services. It has also estimated that the equivalent costs of consumers buying a QALY on the private market is £60,000 (see Department of Health Report 2019). Thus, the UK government assumes that it is four times as effective at producing QALYs as the private market is when it comes to health.

There is also an inherent contradiction in assuming that the government is useless in a CBA, as the whole point of a CBA is to convince the government that a project is worth undertaking, based on the premise that goods that produce a positive surplus should be funded and organised. How is it possible to simultaneously assume that the government has the capacity to judge the CBA under consideration, but that all other activities funded by the government are useless?

These assumptions seem even stranger if we reflect on another number in this table: namely, the number denoting the supposed value of 100,000 QALYs generated by the project. This number is not in the Airport Commission report, and I added it purely to illustrate a point. It is intended to reflect how within the UK government's machinery changes in health are now valued: any health benefits (or disbenefits) now have to be valued at £60,000 per QALY generated or lost. Thus, the assumption is that individuals are willing to pay £60,000 for an additional high quality year of life, whereby high quality is thought of purely as a high level of health.

Therefore, QALYs are valued at £60,000, if generated by a project. However, at the same time, the government itself calculates that it can produce a QALY at a cost of £15,000. This is obviously incompatible with the idea that the net added value of the whole of government is zero.

These three incompatible prices lead to a number of incompatible implications when taken to their logical conclusion.

One implication is that the government is assumed to be able to "print surplus" by usurping all of the resources involved in a project, and can simply spend these resources producing QALYs. Basically, it is assumed that the government makes a surplus of £4 for each pound of investment in health.

Thus, if CBA were taken seriously and to its logical conclusion, the benefit-to-cost ratio of any project should be at least four for the government to prefer it to spending more tax revenues on health production. The fact that this is not actually the case means one of two things: either the government does not actually take CBA seriously at all, or all government spending has an equal multiplier in its benefits that implies that it should not spend all tax revenues on health services.

There are more impossible implications of these price inconsistencies: if we assume that health spending has a four-to-one cost-benefit ratio, but total government spending is worth exactly the same as the inputs, the implicit assumption is that non-health spending has less value than its costs. Since health spending in the UK is about 25% of all government spending, the health spending alone is worth the full value of all government revenues. Thus, logically speaking, all other government activities are assumed to have a value of zero: i.e., they destroy the surplus, as if all other government activities benefit no one, not even a fraction of the inputs.

We can therefore state that the trifecta of assuming that individuals are willing to pay £60,000 for a QALY, that governments produce health at a cost of £15,000 per QALY, and that government revenue has the same value as consumer surplus, are inconsistent. These assumptions imply that government is both useless and four times as useful as consumer surplus.

This point might be dismissed by noting that health has a special status within the UK government, and that the convenience of making mutually incompatible assumptions – which has been common practice for the last 10 years – protects the health budget from cuts, and makes every other ministry and large organisation keen to count health benefits and avoid adverse health consequences. Thus, it is possible to argue that even though it is clear that cost-benefit analyses involving health are nonsensical, their use leads to better incentives within different departments and organisations relative to some prior status quo. This may be true, but the use of CBA also leads to perverse incentives, as any ministry can make any project look good by also producing some health benefits at a cost of £15,000 per QALY, while claiming that they are worth £60,000 per QALY. It therefore appears that any bad project can be made to have a high net present value by including some health expenses. The health ministry in the UK now uses exactly that trick in many cases (see, for example, Department of Health Report 2019).

Moreover, I should briefly note that the labour market benefits of the airport expansion plan – i.e., the additional jobs – are valued using general equilibrium arguments, which is to say they are not valued at all. In the UK, additional jobs only count in CBA if the case can be made that these individuals would otherwise remain unemployed. To do so, it would be necessary to argue that jobs are created faster in the industry related to this project than in the rest of the economic sectors. Hence, when it comes to the labour market, general equilibrium thinking is enforced, while for most other purposes, general equilibrium thinking is avoided. One for instance cannot interpret the consumer surplus estimates in a general equilibrium way: they would then be meaningless.

Finally, we should briefly mention that in Table 1, various externalities like noise and biodiversity are listed. These have been valued by generating an as-if willingness to pay a certain number, either by presenting people in the neighbourhoods with a scenario in which they have to choose between having less money and greater biodiversity; or by making hedonic calculations, such as by inferring a low willingness to pay for noise by the low degree to which noise is reflected in house prices. Thus, again, a rather extreme version of rationality is used to generate rather low numbers for the costs of noise and biodiversity. If the report had instead used the actual estimates of the health costs of noise, or of the production costs of more biodiversity (e.g., Stansfeld and Matheson 2003), the numbers would have been far higher. But then, the airport expansion would have been rendered much less attractive.

The final point is, therefore, that the absence of a standard and accepted measure for valuing non-market goods like noise and biodiversity allows CBA calculators to choose from a wide range of methods. Obviously, they then choose the method that makes those who pay them look good.

2.4 Summing up the practise and problems of standard cost-benefit accounting

As practiced in the UK, CBA is based on a very odd collection of mutually inconsistent assumptions. To sum them up:

1. It adds up private consumption and government revenue, thereby copying the GDP accounting practises of the 1940s.
2. Since about 2010, health benefits have been counted as four times as much as the government production costs of health. Together with (1), this basically means government is assumed to be both useless at anything but health, and to have a benefit-to-cost ratio of four relative to private consumption.
3. Consumer surplus, an unseen form of pleasure derived from estimated demand curves, is added. This means that CBA is not the same as GDP maximisation, nor is it aligned with classic economics in that it values some notion of internal pleasure. No distributional effects or general equilibrium aspects are involved in this calculation. Thus, this 19th-century practise is deeply inconsistent with standard macro-economics and GDP accounting; and with any empirical evidence of how people truly feel, including of how poor people value goods relative to rich people.
4. CBA applies partial equilibrium for any particular purchase or outcome, and general equilibrium for the labour market and some other aspects of the economy.
5. No consumption externalities of any kind are included. People are assumed to be an island unto themselves, with no role for jealousy or social production.

6. When something cannot easily be valued, either it is implicitly seen as “not economic” (and hence of no value), or an attempt is made to find a willingness to pay.

Thus, CBA is a leaky bucket with elements of the 19th, 20th and 21st centuries that combines economics, 19th-century mysticism (consumer surplus), GDP accounting practises and politically convenient habits. Its main merit is that numbers come out of its procedures, and that others who follow the same procedure will obtain similar numbers. The high degree of malleability in terms of what is counted and what negotiated prices are assumed allow the CBA calculator to reach almost any conclusion.

These problems underscore why CBA has never truly become the standard for government decision-making, and political judgments are still prevalent in many areas. As we saw in the case of the Heathrow expansion, a CBA is more likely to be done to overcome a necessary hurdle within a bureaucracy than to reach an informed decision.

3 An alternative: the WELLBY

A very different way of looking at what government produces is to base government decision-making explicitly on the idea that the government should look after the interests of the population, measured as the wellbeing of the population. It is an old Enlightenment idea that government should care about the happiness, or wellbeing, of the population. The US Constitution, for example, characterises the pursuit of happiness as a fundamental right.

While governments in places like Bhutan have long advocated taking this ideal seriously (see Ura et al. 2012), an explicit measurement of wellbeing is needed to turn this stated ideal into a reality. As we saw above, current reality is the leaking bucket of cost-benefit analysis in which there is a strong bias towards private consumption of the rich, and against government spending and taking consumption externalities into account. To move away from that reality and towards the long-standing Enlightenment ideal, we need an actual statistical measure to base policy on. In the absence of such a measure, the ideal is just window-dressing (see Frijters et al. 2020).

The wellbeing measure proposed here is the WELLBY, which is based on the life satisfaction question. In the case of the UK, the following question is used in all the surveys conducted by the Office of National Statistics since 2011:

> “Overall, how satisfied are you with your life nowadays?” 0 not at all, 10 completely.

There are many variants of this question used in other surveys, but the basic idea is to ask respondents for an evaluation of the whole of the person’s life on an explicit numerical scale. A WELLBY is then one unit of life satisfaction on a 0–10 scale for one person, for one year. The purpose of government is then to maximise the expected discounted stream of the WELLBYs of the country's population from now until eternity.

Readers interested in the philosophy and history of this approach, its many advocates, and various measurement issues, are referred to Frijters and Krekel (2021) or Clark et al. (2018) for lengthy discussions. Here, I only wish to briefly summarise the case.

In terms of a single equation, the aim of public policy at any moment in time is then the discounted stream of WELLBYs of the population:

$$\sum_t (1-\rho^W)^t \sum_i SW_i * (LS_{it} - LS_0)$$

Here, time t is in whatever unit is most convenient (years, months, minutes, decades); ρ^W is the pure discount rate on the future, the individual counter i sums over the relevant population (which will change over time); SW_i is a non-negative weight that denotes the possibility that some people are given a higher social weight (for instance, because they are full citizens rather than joint citizens with another country); LS_{it} is the life satisfaction level expected for person i at time t; and LS_0 is the zero point of life satisfaction, denoting the level of life satisfaction equivalent to death.

For each aspect of this basic maximand, technical standards are needed. The general proposal is that decision-making would be based on maximising the expected stream of WELLBYs. Please note that this usually will not require us to measure the stream of wellbeing. What we want instead is good evidence on the causal effects of proposed policies on wellbeing, such as evidence from experimental studies or high-quality inferential studies. We can then draw on the more than 170,000 studies that have already been done on life satisfaction (cf. Diener et al. 2018). Basically, the intention is to use this huge literature on the causes of life satisfaction to inform policy-making.

In the following, I briefly discuss the basic reasoning of this methodology; address some questions raised by public sector and academic organisations; and offer some brief remarks on the technical standards.

3.1 Rationale

The basic rationale is that individual responses to life satisfaction questions should be seen as a secondary form of voting. When individuals answer these questions, they are providing information about how they evaluate their life – and, thus, about what they think about their current circumstances. By analysing the circumstances associated with higher or lower evaluations, we can find out what issues matter in people's lives that can be improved by individual and government choices. More than 80 years of experience with this question has produced a social science literature of well over 100,000 articles on its determinants, and a "grey literature" of government reports and other non-academic publications that is easily twice as large. While the quality of this research varies widely, it provides an enormous amount of information on what matters for individuals, and thus represents an alternative to traditional cost-benefit analysis.

A good example of a real strength of this methodology is provided by the Covid-19 crisis in 2020. That crisis has led to massive changes in many different dimensions, and thus calls for the development of a methodology for making decisions about potentially painful trade-offs. How many additional lost years of life via Covid-19 deaths are worth a million more people being unemployed for another year? How many permanently damaged lungs are worth the loneliness and increased suicide rate among a billion elderly people locked up for three months? How much anxiety is alleviated due to social distancing rules, and many jobs that involve touching each other are destroyed by social distancing rules?

Standard cost-benefit analyses do not have ready answers to most of these questions, but the WELLBY methodology does have ready answers for many of the important elements (Clark et al. 2018). In brief, a year of normal life is valued at six WELLBY; a year spent unemployed costs 0.7 WELLBY a year; forced social distancing and lockdowns of the UK variety cost 0.5 WELLBY a year; and the government produces one WELLBY for about £2,500 of government revenue (Frijters and Krekel 2021). With these basic numbers alone, coupled with estimates of the economic collapse during the Covid-19 crisis and the lethality of the virus, I was able to already state that as of April 2020, that the costs of the various suppression strategies imposed by the UK government were easily 50 times higher than the possible benefits (Frijters 2020).[1]

Thus, one of the big selling points of the WELLBY approach is that it makes assessing difficult trade-offs relatively simple, because it provides a rough estimate of how important various things are to people that can be applied quickly when evaluating potential trade-offs.

We next discuss how this approach could be applied more generally, and which individuals within a government bureaucracy would do the calculations.

3.2 Practices and roles

The general idea is that all civil servants would be given some basic training in wellbeing, such that they are all aware of the basic insights on what matters a lot, and what matters only a little, for individual wellbeing and societal wellbeing. Government departments would generate their own narrative as to how their activities support wellbeing in the country as a whole, with a continuous re-setting of that narrative as departments slowly discover where they can make additional gains,

[1] See also https://clubtroppo.com.au/2020/03/21/the-corona-dilemma/ for initial estimates; and here for a whole workshop on the issue: https://clubtroppo.com.au/2020/06/08/a-seminar-workshop-on-wellbeing-cost-benefit-analysis-applied-to-covid/.

and where they can save money on activities that, in hindsight, do not generate much wellbeing at all. There would be a gradual process of learning by individual departments and other public organisations about how they could generate more wellbeing for less money.

There would be obvious roles for universities and public intellectuals if improving the WELLBY level were taken as the goal of government. Universities would need to provide training in wellbeing-related methodology, including in the widespread use of randomised control trials to ascertain the effects of this or that factor on individual and aggregate wellbeing, as well as the many intricacies of measurement and interpretation. There would also be a role for more theoretical and "macro-wellbeing" academics to discern more complex patterns, inter-dependencies and general equilibrium effects. Frijters and Krekel (2021) sketched what that could look like, but the current reality is that there are very few wellbeing theorists. Hence, there are many opportunities for ambitious young scholars in that field.

There would also be a natural role for civic society and public intellectuals in this sphere. For example, government budgets can be presented and discussed in terms of wellbeing changes in the population, using regular measurements. Journalists may be interested in the question of whether the expected increase in wellbeing from this or that new proposal is really likely to emerge. "Experts" could discuss on the news what the latest wellbeing numbers for certain groups say about the previous successes and failures of policies. In the UK, some of this activity is starting to happen via the What Works Centre for Wellbeing, which is a central hub for scientific knowledge and public discussion around wellbeing.

A focus on wellbeing has long been advocated in the UK (see Frijters et al. 2020). For example, Prime Minister David Cameron declared wellbeing to be the goal of the UK government in 2011. Moreover, in 2004, the local health care act made it mandatory for local councils to care for the wellbeing of their populations. Nonetheless, it is expected to take a long time for institutions to truly adopt these new objectives.

3.3 Technical standards

There are many practical and political questions that need to be addressed when actually using the WELLBY to decide whether to undertake a given policy. First, there is the question of who the affected population is. Are people who have yet to be born or to migrate to the relevant country also counted? Second, there is the question about the appropriate discount rate that governs how much future generations count relative to current generations. Third, there is the question of what stand-in measure should be used as a proxy for the WELLBY for sub-populations who cannot themselves answer questions about life satisfaction, such as babies and those who have lost their cognitive faculties. Fourth, there is the question of what level of life satisfaction is taken to be the same as death – and, hence, at what point living counts as worse than death.

Although I have suggestions for all these technical standards (Frijters and Krekel 2021), much more important than those suggestions is the process via which technical standards would be established and could be challenged. What process would a society trust to come up with reasonable standards that produce outcomes that the society would, on balance, agree with?

The process would have to be transparent and subject to challenge; not hidden in bureaucratic processes, as is currently the case for cost-benefit standards. If public policy is to become based on the subjective opinions of the population, the key elements that are taken into account must also be visible and debatable.

One way to do this is to have an initial list of technical standards, as well as an initial list of current central estimates of the effects of various outcomes on wellbeing. The technical standards largely regulate the question of who matters. Specifically, who matters in the current and future population, and how much does death matter? That is clearly a core question of democracy and social power that can be addressed using a debatable set of standards that are occasionally adjusted by a democratic process.

It is also desirable to have an open price list with elements such as the estimated effects of a year of unemployment on wellbeing, the effects of higher nitrogen oxide in the air on wellbeing and the effects of national pride on wellbeing. Clearly, the aim is develop a price list that will reflect the actual outcomes of appeals to the public purse, and that is based on the best science available. However, as with everything scientific issue, new information is generated continuously, Thus, a transparent process is needed to determine which scientific findings are the best at some point in time.

My suggestion is to have a large, open set of prices that is, perhaps, initially compiled by a team of central civil servants, key academics and politicians who represent all of the main political parties. There would then have to be a process for reviewing and challenging these numbers.

It is crucial that the review process only engages with proposed alternatives, and does not set itself up to answer critiques. The potential for blind critique is endless. However, to allow for scientific progress to be incorporated into the process, serious alternatives to the numbers on an initial list should be considered. Hence, if there are new insights about some area, or problems with a current number, there should be a process for judging whether a specific alternative number is better than the one currently used.

The Intergovernmental Panel for Climate Change (IPCC) is a good example of this process. Its modus operandi is well-explained and critiqued by Porter et al. (2018): in the IPCC reports, there is a headline number, plus a process for challenging that number in order to get better numbers over time. It channels different scientific teams towards offering numbers. Blind critique is simply ignored.

There are, of course, questions of causality and negotiations. What effect would implementing a given policy have compared to an appropriate counterfactual? How can the costs of implementing a certain policy be minimised, and, thus, the expected stream of WELLBYs be maximised? Such questions will appear in thousands of contexts, and will require a slow working through of the many issues as they arise.

There are also questions of bureaucracy. How should a system in which decisions are made based on the WELLBY calculations be organised? Who should be taught what about form of wellbeing so that the available knowledge is there where it is needed? Who does the experimenting, and how do we learn from these experiments? These are difficult questions of public administration that do not have crystallised answers.

I discuss all of these questions and more in a forthcoming handbook (Frijters and Krekel 2021).

Next, I wish to outline how the WELLBY changes the nature of CBA analyses. I do so by using the same list of problems associated with CBA, and then briefly discuss how a WELLBY CBA would address these problems differently.

3.4 Government revenue and private consumption

As I showed earlier, traditional CBA adds up private consumption and government revenue, thus copying the approach to GDP accounting of the 1940s.

If the WELLBY is adopted as the maximand of government, then the expected life satisfaction increase due to the higher consumption can be added on top of the expected increase in population WELLBYs due to government revenue. Hence, the same comparable metric of a final outcome is used to judge both private consumption and government spending. This makes the question of whether more consumption of this or that good or service truly buys WELLBYs an empirical one. The same then holds for government revenue. While using this approach is more work than the previous practise of simply assuming that government outputs are the same inputs, it goes to the very core of the benefits and the purpose of government, and thus provides true accountability.

3.5 The production cost of health and its value

Since about 2010, health benefits have been counted as four times as much as the government production costs of health. Together with the previously mentioned practise of inputs-as-outputs, this basically means that government is assumed to be useless at anything but health, and to have a benefit-to-cost ratio of four relative to private consumption.

When the WELLBY is adopted, an attempt needs to be made to calculate the average number of WELLBYs bought with more general government revenue, and with particular types of spending, with the aim in principle of obtaining the same marginal benefit across all government spending. Hence, how many WELLBYs are produced with the expansion of government health services or by private spending on health becomes an empirical question. Thus, health changes need to be translated into WELLBY changes, which is frequently done in the wellbeing literature; i.e., Cookson et al. (2020).

3.6 The treatment of consumer surplus

Consumer surplus, an unseen form of pleasure derived from estimated demand curves, is added in traditional CBA. This sets CBA apart from GDP maximisation, and from classical economics, in that it values some notion of internal pleasure. Distributional effects or general equilibrium aspects are ignored in CBA. Thus, this 19th-century approach is deeply inconsistent with standard economics and GDP accounting; and with any empirical evidence of how people truly feel, or of how poor people value things relative to rich people. How would the treatment of consumer surplus change when the WELLBY is applied?

With a WELLBY, consumer surplus would have to be empirically shown via effects of consumption on life satisfaction. The effect of private income, which can be seen as a proxy for consumption, on life satisfaction is well-studied, and can provide initial answers (basically, a one log point increase in income increases yearly life satisfaction by about 0.4, and thus buys about 0.4 WELLBY, Lindqvist et al. 2018).

3.7 Partial equilibrium and general equilibrium

As we saw above, traditional CBA has a pick-and-mix approach to assuming partial equilibrium or general equilibrium for different problems. Partial equilibrium is presumed for any particular purchase or outcome, and general equilibrium is assumed for the labour market and some other aspects of the economy. This is mainly done because there are simply no general equilibrium models that incorporate everything.

This is an aspect that would not change much in a WELLBY approach. While the WELLBY method does provide access to a lot of numbers that reflect the value of various forms of consumption and government production, those numbers are all estimates that come from slightly different models using different assumptions, where it is far more normal to see estimates from partial equilibrium models than more general equilibrium models.

The estimation of wellbeing effect then usually happens in a partial equilibrium manner, such as from randomised control trials that compare one group of people who get a public service (like health insurance or better police services) with a group who get standard services. Such an approach yields a partial equilibrium estimate, but then specific elements of general equilibrium, like consumer externalities, also get estimated, usually by different researchers. So the field might discover slowly over time when the general equilibrium effects are strong or weak.

Having an explicit measure of wellbeing does allow for a more natural discovery of general equilibrium effects, since it can be used to examine wellbeing changes between countries and over time in order to identify the effects of major changes on the macro level. However, that kind of procedure is fraught with identification issues, including measurement error, difficulties in comparing wellbeing measures over countries, and difficulties in identifying the macro effects of policies in a fast-changing environment. Thus, we should not expect progress to be particularly fast in this realm. Realistically speaking, pragmatic pick-and-mix in terms of the sophistication of the underlying theoretical structure of the socio-economic system will be a general feature of WELLBY CBA as much as it is of existing CBA.

3.8 Consumption externalities

In traditional CBA, there are few consumption externalities of any kind: i.e., each person is an island unto her or himself, with no role for jealousy or social production.

This radically changes with the adoption of the WELLBY as the government maximand. These consumption externalities can be estimated, which is the basis of the Easterlin Discount (see later).

3.9 The space of goods and commodities considered

In traditional CBA, when something cannot easily be valued, it is implicitly often seen as not economic (and hence of no value), or an attempt is made to find a willingness to pay.

Examples of elements that are difficult to value with a CBA are national pride, individual self-esteem, jealousy, anger, hurt, joy and a sense of identification with a group. While such elements are certainly studied by behavioural economists and other social scientists, they are not commonly part of CBA, because it is not easy to ask people how much they are willing to pay to, for instance, feel less jealous. Similarly, deducing a market value for the willingness to pay for being part of a group of people is far from trivial, as it is, for instance, hard in a market to deduce the degree to which anyone is part of a group.

The adoption of the WELLBY approach makes measuring these elements much easier, as it is not necessary to assume that people are aware of how much they value self-esteem, national pride, etc. The importance of those elements can be deduced by measuring over time how the wellbeing of individuals and of whole countries changes with shocks to their self-esteem, national pride, etc. This sort of measurement has been done extensively in the wellbeing literature. For instance, the hosting of the Olympics is used as identifying information for temporary local changes in national pride, as well as for its effects on wellbeing (Dolan et al. 2019). I will spell out in a later section how the found wellbeing effect of jealousy and social status can be used to change CBA methodology.

However, it is not the case that the use of the WELLBY immediately solves all issues regarding what the appropriate space of goods and commodities is. Moreover, the degree of attachment with groups remains a difficult thing to measure by assessing wellbeing (although there have been many attempts to do so; e.g., Pearce 2017 studied arts education classes as investments in social bonding).

However, using an empirical measure of wellbeing does have a huge advantage over using traditional CBA methods in terms of finding “missing commodities”: i.e., changes in wellbeing that cannot be explained by any menu of currently observed commodities point to the existence of something unseen that is driving wellbeing changes. Thus, differences over time and between groups can be used as indicators of what the missing commodity group might be that is important for how people evaluate life.

An example of how wellbeing data indicate a missing commodity is the finding that Latino cultures, such as the cultures of Colombia and Costa Rica, are unusually happy, holding economic development constant (World Happiness Report 2019). This has led many wellbeing scholars to try to ascertain what it is about those cultures that makes the population relatively happy – which then leads to the question of whether the identified commodities can be produced elsewhere. Rojas (2020), for instance, has argued that families and friendship groups in Latin America are particularly good at creating warm social relations in which people feel free to dance, display emotions and support each other. Whether these features can be transferred to Anglo-Saxon cultures or to other cultures is very much an open question.

3.10 Conclusion on CBA and the WELLBY

By applying an explicit measure of what people themselves value, the most glaringly odd and inconsistent elements of standard CBA can be eliminated: i.e., we can do away with consumer surplus and inconsistent assumptions about the value created by government expenditures versus private consumption. However, not all problems disappear or are made simpler by the use of this measure. The question of the macroeconomics of wellbeing remains a very difficult one.

Next we show how a particularly important type of social externality – i.e., status and jealousy effects, also known as consumption externalities – has been valued in the wellbeing literature, and how those findings can be used to change CBA methodology.

4 The Easterlin Discount: jealousy and envy effects of the consumption of others

It has long been known that humans care about their outcomes relative to what others have. In the “The Theory of Moral Sentiments” of 1759, Adam Smith already rhetorically asked why humans were so busy working hard and out-doing each other. He answered himself that it was not because they cared for comforts, but, rather, that “it is our vanity that urges us on”. Nowadays, we would say people care about their position in society; or, equivalently, about ensuring that others are jealous and envious of them, rather than the other way round.

Famously, Richard Easterlin (1974) argued that average happiness was no longer going up in the US after 1950 because although incomes were still rising, this increasing trend did not buy extra happiness, since people were not getting richer relative to their neighbours. Currently, the claims being made in this field are more sophisticated: 29 leading wellbeing researchers around the world were asked in the April 2018 World Wellbeing Panel poll whether they agreed with the statement: "Is the main effect of economic growth on national wellbeing via employment and public welfare programs?" Twenty-four out of 29 panellists agreed. The four who disagreed added only small caveats, such as that it depended on whether economic growth was inequality-neutral. Thus, the current version of Easterlin's original hypothesis is that the main WELLBY benefit of economic growth lies in increased government services, with greater private consumption providing very little additional benefit, as the individual benefit of more income is offset by more jealousy among one's neighbours and others in the country.

Just how much of the value to an individual of additional consumption is offset by jealousy among everyone else in the country is an empirical question. Richard Easterlin originally argued it was 100%. The field has not yielded a consensus number, but 50% would be at the lower edge of the estimates (see Frijters and Krekel 2021). For example, the Leiden school, which has estimated this figure since the 1970s, put it at 60% (see the survey by Van Praag and Frijters 1999).

Therefore, one way to use the wellbeing finding that a higher consumption of others reduces the wellbeing of an individual others is to apply a discount to the benefits calculated in current CBAs to the presumed benefits of more material consumption. I call that discount the Easterlin Discount. The next question is how an appropriate discount rate can be applied; i.e., to what it should be applied.

We should be very careful in selecting the items to which we apply the Easterlin Discount. The logic is that it applies to items that convey social status, and are thus things that are exclusive; i.e., that one person can have while others cannot.

Thus, the discount can be applied to anything that individuals can brag about, including services, holidays, and all manner of social experiences. It should not, however, be applied to most government services, since they usually are equally available to all people in the same situation: i.e., they are not exclusive. An individual cannot brag about having access to that which everyone else also has access to, such as education, police protection and nature parks.

As a rule of thumb, the discount can be applied to anything in the private sphere and to nothing in the public sphere, although there are some things that governments buy that are actually only enjoyed by a privileged few (like the Royal Opera or palaces for the royals).

Table 1 shows how the Easterlin Discount could be applied to the various numbers in column 2. The table indicates that it is applied to things people could brag about; namely, consumer and producer surplus. Importantly, the discount is not applied to pain or to government revenue because the latter buys public goods and the first is not visible to others.

It is debatable whether the Easterlin Discount should or should not have applied to particular elements in Table 1, as one might, for example, argue for or against applying an Easterlin Discount to noise. This hinges on the question of whether noise or the absence of it is an observable exclusive status-conveying good. In principle, the goal is to have empirical evidence that addresses this question, such as surveys in which people are asked if they are jealous of the tranquillity some others buy.

However, to be operational, a default position is needed that can be applied to cases for which there is not yet good specific evidence. The default advocated is not to think of the (lack of) pain of unpleasant noise as something people brag about. Rather, the absence of irritating noise is seen as a basic right because "disturbing the peace" is not allowed by law, and because discomfort from noise is not easy to measure. However, while this might be reasonable to say of noise in the UK, this position will not hold for all countries and all periods.

Let us then give some basic numbers to show how a WELLBY and the Easterlin Discount changes the priorities in our society.

The best number we have so far on how much additional income wellbeing buys comes from an analysis of Swedish lotteries (Lindqvist et al. 2018), which states that a one-point log increase in yearly personal income buys 0.4 WELLBY. For a country with a median income of £30,000, like the UK, this essentially means that with a spread-out increase in income (everyone experiences the same absolute increase), 0.4 WELLBY can be purchased for every £30,000. In other words, £75,000 buys one WELLBY via personal consumption.

The Easterlin Discount would then have to be subtracted from this figure. If the Easterlin Discount is 50%, £150,000 is needed to buy a WELLBY via personal consumption. If the discount is 75%, £300,000 is needed to buy a WELLBY via personal consumption.

This outcome can be compared to government spending: one QALY that is bought via increases in life expectancy would basically buy six WELLBY (because the average life satisfaction of the healthy is about an eight, and people are probably indifferent to death at a level of two[2]). The UK health services can then produce a WELLBY at a cost of £2,500 per unit.

Thus, with an Easterlin Discount of 50%, marginal government expenditure can buy 60 WELLBYs for the same amount of money as individuals could buy a WELLBY with more personal spending.

[2] This number comes from a study by Peasgood et al. (2018). They asked respondents about hypothetical trade-offs between differing lengths of life and the life satisfaction levels associated with that additional length of life. They found that at a life satisfaction level of two, about an equal number of respondents choose death over continued life, which denotes the point at which the marginal respondent is indifferent.

Hence, a WELLBY orientation places far less value on personal consumption, and much more value on social relations and the things governments can provide.

5 Conclusions

In this short paper, I first dissected the mainstream practise of cost-benefit analysis as used by governments to decide on policy priorities, and showed that this approach is based on several inconsistencies and articles of faith, such as the assumption that social and consumption externalities are unimportant. I then introduced the WELLBY as an alternative empirical basis for cost-benefit analyses in which many of the articles of faith within the current methodology become relatively simple questions of empirical science. These questions include just how much wellbeing is created in various government programmes; how individuals are affected by the consumption of others; and what the value is of many non-market goods that are important to people, such as their children and national pride. I then illustrated the power of that methodology by discussing how knowledge of the vanity of individuals gives rise to an appropriate discount rate (the Easterlin Discount) that can be used to reduce the benefit of individual consumption relative to social benefits.

The methodology advocated in this paper is both very old and very new. It is a very old ideal to base government and public sector decision-making on that which is good for the wellbeing of the population. To truly do this using an actual measure of wellbeing based on the scientific literature is new, and will, of course, be resisted by all those who gain from the current situation. However, the explicit adoption of a wellbeing criterion at the heart of public decision-making is necessary to overcome the pro-consumption bias in current standard methods. That pro-consumption bias leads to unsustainable pressure on finite resources, and neglects many of the social and emotional aspects of humanity. We cannot go on pretending that humans are like solitary individuals quietly consuming airplane rides, not affecting anybody else nor being affected by anyone else. To take account of our interdependencies, we must move towards adopting explicit methods through which they can be measured and accounted for.

A good start towards a happier and more sustainable earth would be the adoption of the WELLBY as the object of public policy. Certainly, we can always decide to adopt something even better down the line.

References

Airports Commission 2015. Business case and sustainability assessment - Heathrow north west runway. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/440315/business-case-and-sustainability-assessment.pdf

Boardman, A. E., D. H. Greenberg, A. R. Vining and D. L. Weimer 2017. *Cost-benefit analysis: concepts and practice*. Cambridge University Press.

Campbell, H. F. and R. P. C. Brown 2016. *Cost-benefit analysis: financial and economic appraisal using spreadsheets*. 2nd edition, Routledge.

Cesur, R., P. M. Güneş, E. Tekin and A. Ulker 2017. The value of socialized medicine: The impact of universal primary healthcare provision on mortality rates in Turkey. *Journal of Public Economics* 150: 75–93. https://doi.org/10.1016/j.jpubeco.2017.03.007

Clark, A. E., S. Flèche, R. Layard, N. Powdthavee and G. Ward 2018. *The origins of happiness: the science of well-being over the life course*. Princeton, NJ: Princeton University Press.

Cookson, R., O. Cotton-Barratt, M. D. Adler, M. Asaria and T. Ord 2020. Years of good life based on consumption and health: A practical well-being metric for economic evaluation. In *Measuring the global burden of disease: philosophical dimensions*, eds N. Eyal, S. A. Hurst, C. J. L. Murray, S. A. Schroeder, and D. Wikler, 126–150. Oxford University Press. https://doi.org/10.1093/med/9780190082543.003.0009

Department of Health Report. 2019. Transforming children and young people's mental health provision: a green paper. IA No: 14001, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664442/MHGP_IA.pdf

Diener, E., R. E. Lucas and S. Oishi 2018. Advances and open questions in the science of subjective well-being. *Collabra: Psychology* 4(1): 15. https://doi.org/10.1525/collabra.115

Dolan, P., G. Kavetsos, C. Krekel, D. Mavridis, R. Metcalfe, C. Senik, S. Szymanski and N. R. Ziebarth 2019. Quantifying the intangible impact of the Olympics using subjective well-being data. *Journal of Public Economics* 177: 104043. https://doi.org/10.1016/j.jpubeco.2019.07.002

Dupuit J. 1848 translated and reprinted 1969. On the measurement of the utility of public works. In *Readings in welfare economics*, eds K. J. Arrow and T. Scitovsky. London: Allen and Unwin.

Easterlin, R. A. 1974. Does economic growth improve the human lot? Some empirical evidence. In *Nations and households in economic growth: Essays in honor of Moses Abramovitz*, eds R. David and R. Reder, 89–125. New York: Academic Press.

Easterlin, R. A. 1995. Will raising the incomes of all increase the happiness of all? *Journal of Economic Behavior & Organization* 27(1): 35–47.

Frijters, P. 2020. How many WELLBYs is the corona panic costing. *Blog Post*, https://clubtroppo.com.au/2020/04/08/how-many-wellbys-is-the-corona-panic-costing/

Frijters, P., A. E. Clark, C. Krekel and R. Layard 2020. A happy choice: Wellbeing as the goal of government. *Behavioural Public Policy* 4(2): 126–165. https://doi.org/10.1017/bpp.2019.39

Frijters, P. and C. Krekel 2021. *Handbook of wellbeing policy for the UK*. Forthcoming with Oxford University Press. LSE Working Paper. 338 pages.

Lindqvist, E., R. Östling and D. Cesarini 2018. Long-run effects of lottery wealth on psychological well-being. National Bureau of Economic Research No. w24667. https://doi.org/10.3386/w24667

Morrison, S. and C. Winston 2010. *The evolution of the airline industry*. Brookings Institution Press.

Papanicolas, I., L. R. Woskie and A. K. Jha 2018. Health care spending in the United States and other high-income countries. *JAMA* 319(10): 1024–1039. https://doi.org/10.1001/jama.2018.1150

Pearce, E. 2017. Participants' perspectives on the social bonding and well-being effects of creative arts adult education classes. *Arts & Health* 9(1): 42–59. https://doi.org/10.1080/17533015.2016.1193550

Peasgood, T., C. Mukuria, M. Karimi and J. Brazier 2018. Eliciting preference weights for life satisfaction: A feasibility study. Mimeo.

Porter, A. J., T. R. Kuhn and B. Nerlich 2018. Organizing authority in the climate change debate: IPCC controversies and the management of dialectical tensions. *Organization Studies* 39(7): 873–898. https://doi.org/10.1177%2F0170840617707999

Rojas, M. 2020. Latin America and well-being. In *Well-being in Latin America: Drivers and policies*, 1–9. Springer, Cham. https://doi.org/10.1007/978-3-030-33498-7_1

Smith, A. 1759. *The theory of moral sentiments* (2nd edition, 1761). Strand & Edinburgh: A. Millar; A. Kincaid & J. Bell.

Smith, J. P. 1999. Healthy bodies and thick wallets: The dual relation between health and economic status. *Journal of Economic Perspectives* 13: 145–166.

Stansfeld, S. A. and M. P. Matheson 2003. Noise pollution: non-auditory effects on health. *British Medical Bulletin* 68(1): 243–257.

Treasury, H. M. S. 2018. *The green book: Central government guidance on appraisal and evaluation*. London: HM Treasury.

Ura, K., S. Alkire, T. Zangmo and K. Wangdi 2012. A short guide to gross national happiness index. The Centre for Bhutan Studies.

Van Praag, B. M. S. and P. Frijters 1999. The measurement of welfare and well-being; the Leyden approach. In *Well-being: the foundations of hedonic psychology*, eds D. Kahneman, E. Diener, and N. Schwarz, 413–432. New York: Russel Sage Foundation.

World Wellbeing Panel 2018. April 2018: Economic Growth, Welfare Programs & Wellbeing. https://www.barcelonagse.eu/research/world-wellbeing-panel/survey-archive

REVIEW ARTICLES

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 67–89

Comparing global reports of subjective well-being to experiential measures

Richard E. Lucas[1,*]

Abstract

Subjective well-being (SWB) is an overall evaluation of the quality of a person's life from his or her own perspective. One common method of assessing this construct requires respondents to think about their life as a whole and to provide a "global" evaluation that summarizes across life domains or affective experiences over extended periods of time. The validity of these global measures has been challenged, however; and experiential measures, which ask respondents to report on their momentary evaluative experiences many times over a constrained time period, have been suggested as a more valid alternative. This paper addresses the empirical evidence for one important challenge to global measures: the possibility that temporarily salient information overwhelmingly influences global judgments, reducing their reliability and validity. This paper critiques prior evidence for this challenge and presents new concerns about the assumed validity of the proposed alternative: experiential measures.

Keywords: subjective well-being; life satisfaction; measurement; experience sampling method; day reconstruction method

1 Introduction

Subjective well-being (SWB) is an overall evaluation of the quality of a person's life from his or her own perspective. According to Diener et al. (1999), the construct and its measures are characterized by three features. First, SWB, as its name suggests, is *subjective*; it captures the respondent's own perspective about how life is going rather than an objective evaluation or an outside observer's perspective. Second, it is *global* in scope. In other words, measures of SWB are intended to capture overall evaluations that incorporate multiple domains of a person's life. Finally, SWB is

[1]Department of Psychology, Michigan State University, East Lansing, Michigan, USA

[*]Correspondence to: Richard E. Lucas, lucasri@msu.edu

DOI: 10.1553/populationyearbook2021.rev01

balanced in its focus, capturing both positive and negative aspects of life. These distinguishing features of SWB make it especially well-suited for use in a variety of applied and theoretical investigations. For instance, by focusing on subjective evaluations of life as a whole, SWB researchers allow respondents to consider, evaluate and weight the importance of a broad range of factors that may contribute to their perceived quality of life, rather than having those factors selected for them. The subjective nature of the construct contrasts with alternative approaches that rely on expert judgments or theoretical arguments to determine which features of the good life are most important (e.g., Hurka 2014; Ryan and Deci 2000; Ryff 1989).

The subjectivity of SWB and its measures are especially important when investigating decisions, events or interventions that may affect some life domains in a positive manner, while simultaneously affecting other life domains negatively. For instance, specific medical treatments might successfully reduce pain or other symptoms, while also causing negative side effects for cognitive or social functioning. Thus, SWB may be useful if the aim is to find out whether, on balance, the treatment has improved a respondent's life. The global, subjective nature of SWB judgments allows respondents to evaluate and weight the impact of the treatment on these and other life domains. By assessing SWB, researchers can more fully evaluate the total effect of treatments or other interventions. These features have allowed SWB measures to be incorporated into a wide range of applied and theoretical research programs, and they have led to an increased focus on SWB judgments in policy decisions (Diener et al. 2009; Dolan and White 2007; Frijters and Krekel 2021; Kahneman et al. 2004b; Krueger et al. 2009).

Of course, the usefulness of SWB as a construct depends critically on the quality of the measures that are available to assess it, and debates about measurement issues have existed for as long as SWB has been studied. The greatest strength of SWB—its subjective nature—is also responsible for its greatest measurement challenge: SWB cannot be directly observed, and no "gold standard" measure of the construct exists.[1] Thus, substantive investigations of SWB must simultaneously grapple with standard issues of internal and external validity, and with ever-present concerns about the quality of the measures used to assess the central construct.

Throughout much of the early history of research on SWB, social scientists relied on *global evaluative* measures. This class of measures requires respondents to reflect on their lives and to provide an overall evaluation. Examples include the Satisfaction With Life Scale (Diener et al. 1985), along with a range of single-item life satisfaction scales used in many population-based surveys. Although global evaluative measures have a number of advantages, including their simplicity and high face validity, they are, of course, not perfect. One important challenge to these

1 One might argue that subjective evaluations should correspond to internal, physiological states, and thus, that physiological measures could ultimately serve as a gold standard. Although this is a possible outcome as psychologists' understanding of physiological indicators develops, there are currently no candidate measures that have been found to have substantial levels of validity as measures of the underlying construct (Ito and Cacioppo 1999).

measures—a challenge that is one primary focus of this paper—has been proposed by researchers from the judgment model perspective (Schwarz and Strack 1999). Proponents of the judgment model suggest that constructing a global evaluation should be time-consuming and difficult; i.e., it should require respondents to remember, evaluate and aggregate evaluations of a potentially infinite set of life domains and events. As a result, respondents are thought to rely on a variety of heuristics and biases that simplify the evaluation process, but also affect the validity of the measures that are found.

If these judgment processes affect the reliability and the validity of global measures, then it would be fruitful to investigate alternative measurement strategies that address these specific concerns. Indeed, global measures can be contrasted with *experiential* measures, which focus on narrower evaluations of specific moments in time (Kahneman and Riis 2005). For instance, with experience sampling methods, respondents are asked repeatedly over an extended period to provide ratings of how they feel at each moment. These momentary ratings can capture emotional reactions to specific events, or they can capture longer-lasting moods that may be influenced by an individual's temperament or lingering effects of events or circumstances that are not currently the focus of attention. Presumably, people whose lives are going well will evaluate more of the moments of their lives positively (Kahneman 1999). Importantly, from a measurement perspective, experiential measures should be easier for respondents to complete because respondents are not required to remember and aggregate across multiple moments or life domains. In other words, experiential measures might solve the problems of global reports because they simplify the judgment that is required: i.e., respondents only need to consider the moment that they are in.

The goals of this paper are twofold. First, I critically review the evidence that supports the judgment model critique of global evaluative measures. Although a number of highly cited studies purport to provide evidence for the idea that temporarily salient information can dramatically impact global judgments of SWB, these studies have many characteristics that limit the strength of the evidence they provide. The second goal of the paper is to evaluate the psychometric properties of experiential measures, which have been proposed as a useful alternative to global measures. I will argue that although these measures directly address the concerns about global measures, they may pose their own threats to validity that have previously not been considered.

1.1 Measurement challenges: The judgment model of subjective well-being

One of the clearest challenges to the validity of self-reports of SWB comes from Schwarz and Strack's *judgment model* of SWB (for a review, see Schwarz and Strack 1999). The judgment model focuses on global evaluative measures of subjective well-being: i.e., measures like life satisfaction judgments, which ask respondents to consider their life as a whole and to provide a summary evaluation. According to the judgment model, responses to these global measures result from a judgment process that is susceptible to systematic sources of bias and invalidity. Multiple studies from the judgment model perspective have investigated specific processes that are purported to occur when these judgments are made.

For instance, one concern is that SWB judgments are made too quickly to reflect an overall evaluation of all relevant information about a person's life. Thus, people may quickly scan their memories (or even their current thoughts and feelings) for relevant information, relying too heavily on the information—including potentially unimportant information—that happens to be accessible. Indeed, Schwarz and Strack (1999) suggested that "information that has just been used—for example, to answer a preceding question in a questionnaire—is particularly likely to come to mind later on" (p. 63), and that this temporarily accessible information biases the resulting responses.

In a famous demonstration of this effect, Strack et al. (1988) randomly assigned respondents to answer two questions, one about their global life satisfaction and one about their satisfaction with dating, in one of two orders: dating satisfaction first or global satisfaction first. According to the judgment model, when dating satisfaction is presented first, this life domain should be made especially salient, which should result in strong correlations between the two questions. In contrast, when the global evaluation is presented first, the respondent's romantic life may not be on his or her mind, and may thus have less of an impact. Results from two separate studies supported this prediction: Correlations between the two questions were only .16 and −.12 when the global satisfaction question was asked first, but they were .55 and .66 when the satisfaction with dating question was asked first. Thus, these studies suggest that the information that happens to be accessible at the time of a global judgment affects the judgment itself.[2]

Importantly, it is not just information about a person's life that can affect these judgments; according to judgment model researchers, even feelings that people are experiencing at the time of judgment can affect their judgments in undesirable ways. In one of the most cited examples of research from this tradition, Schwarz and Clore (1983) conducted two studies to assess whether respondents tended to rely on potentially irrelevant current moods when making judgments about their life as a whole. In each of these studies, the authors manipulated the respondents' current mood (using recall of positive and negative life events in one study and changes in daily weather in the second) to assess whether these manipulations also affected their life satisfaction judgments. In both studies, the respondents in the positive mood conditions reported substantially higher life satisfaction than those

[2] Deaton and Stone (2016) describe an unexpected item-order effect in a large sample of Americans that could be interpreted as further support for this type of context effect. However, because the study was not developed specifically to elicit this effect, the interpretation of this effect is somewhat ambiguous, and more research into the processes that underlie it is needed (Lucas et al. 2016).

in the negative mood conditions (with effect sizes close to a full standard deviation difference). These compelling results were also replicated in other studies that used different mood induction procedures, including placing respondents in a pleasant or an unpleasant environment (Schwarz et al. 1987), surveying people before and after important soccer games with varying outcomes (Schwarz et al. 1987), and even simply arranging for respondents to find a dime before completing the relevant survey (Schwarz 1987).

The findings of other studies have even suggested that feelings that arise simply as a result of the process of searching one's memory for relevant information can affect judgments. For instance, although this study was not conducted with well-being judgments as an outcome, Schwarz et al. (1991) found that the ease with which respondents could recall examples relevant to a characteristic could affect their ratings of that characteristic. Specifically, participants who were asked to list six examples of times when they acted assertively rated themselves as more assertive than those who were asked to list 12 examples of assertive behaviors. Presumably, the feelings respondents had about the difficulty associated with producing 12 examples provided information that was used in the judgment itself. To my knowledge, studies of this accessibility effect have not been conducted in the well-being domain. However, evidence of this general phenomenon adds support to the idea that a variety of feelings that people experience (including those that are arguably irrelevant for the judgment at hand) can influence their broad judgments.

Taken together, these studies suggest that irrelevant information that happens to be on a person's mind at the time of judgment can have strong effects on the resulting judgment. These findings represent an important challenge to the validity of global reports of SWB because they suggest that people do not always make SWB judgments in a manner that is consistent with theory or intuition, and that irrelevant information may play a large role in their responses. Indeed, after reviewing research from the judgment model tradition, Schwarz and Strack suggested that one plausible conclusion could be that "There is little to be learned from global self-reports of well-being . . . [What] is being assessed, and how, seems too context dependent to provide reliable information about a population's well-being, let alone information that can guide public policy" (1999, p. 80). These authors argued that the instability inherent in the judgment process limited the extent to which SWB judgments reflected the actual quality of the respondents' lives. Research from the judgment model tradition is often cited in critiques of well-being research, and the presumed problems that these judgment processes create for the validity of these measures have led some scholars to suggest that alternatives to global measures of well-being are needed (e.g., Kahneman 1999; National Academy of Sciences 2013; OECD 2013; Testoni et al. 2018).

2 The promise of experiential measures

The concerns that have been raised about global evaluative measures have led some researchers to propose and investigate alternative approaches for assessing SWB. Many of these approaches have focused on addressing the possibility that respondents may have difficulties answering questions that call for global judgments because answering such questions requires respondents to call to mind information from a broad range of life domains, and to aggregate that information into a single judgment. Both tasks—remembering and aggregating—may be challenging and subject to biases.

Robinson and Clore (2002) proposed an accessibility model focused on self-reports of emotion that is relevant (and that has been applied) to judgments of SWB. In their model, they proposed that different time frames for emotion reports result in the use of different processes. For instance, they argued, when evaluating current emotional experiences, people may rely on experiential knowledge, which is likely to capture actual variance in emotional experiences. At the very least, they pointed out, reports of current emotions cannot be affected by flawed memory, which suggests that these reports should be relatively accurate.

As the time frame of the emotional report increases in duration (the amount of time it covers) or temporal distance (the amount of time that has elapsed since the emotion occurred), respondents must rely on episodic memory—i.e., memory of a specific episode and the details of the experience—to provide a report. This process requires the respondent to reconstruct the episode, and memory limitations may prevent respondents from accurately recalling how they felt in the moment. Furthermore, biases related to salient features of specific experiences may systematically distort their recall of emotional experiences. Thus, it might be expected that reports of recent emotional experiences are somewhat accurate, but less so than emotions reported in the moment.

As the length of time between the experience and the report grows, the details of the experience become more difficult to remember, and, according to Robinson and Clore (2002), individuals may rely even more heavily on situation-specific beliefs to reconstruct their experienced emotions. Indeed, respondents may rely more on semantic knowledge about the nature of events than on episodic memory for a specific event when reconstructing the emotional episode. If general expectations about situations contradict the unique features of a specific experience, the reports of that experience may be inaccurate. For instance, if a respondent believes that she typically experiences joy at parties and forgets that during a particular party one month previously she was concerned about a work-related conflict, she may overestimate the level of joy she experienced. Finally, Robinson and Clore (2002) suggested, if respondents forget what they were doing during the time period in question, then they may rely on general beliefs about how they typically feel, even if their experience had actually deviated from these typical levels.

Robinson and Clore's (2002) model raises concerns about people's ability to remember relevant information when providing responses to global well-being questions. However, there are also concerns about people's ability to aggregate across experiences even if their memory was flawless. For instance, in a series of studies, Kahneman and colleagues provided evidence that when asked to evaluate an extended experience that varied in hedonic valence, respondents tended to place less emphasis on the duration of the experience, and to weigh the peak experience and the end experience heavily when providing an overall evaluation (Fredrickson and Kahneman 1993; Kahneman et al. 1993; Redelmeier and Kahneman 1996). In a particularly striking demonstration of this effect, Redelmeier et al. (2003) randomly assigned patients undergoing colonoscopies to either have the procedure end normally after an especially painful part of the procedure, or to have the procedure extended with a short interval of much milder pain. The authors predicted that even though the patients in the extended condition experienced more pain overall, the tendency to neglect the duration of an experience, combined with the greater salience of the much milder pain at the end of an episode, would lead to more positive evaluations of the extended procedure than of the normal procedure. In support of this prediction, the participants in the extended procedure condition evaluated it more positively, and were even more likely to return for a follow-up than those in the normal procedure condition (although the latter effect was small and just barely significant). Because the two procedures were, on average, identical other than the added interval of mild pain, the tendency to evaluate the longer experience as more desirable than the shorter one violated reasonable expectations about which experience was objectively better. Although debates continue about the processes that underlie effects like these (e.g., Tully and Meyvis 2016), there are plausible concerns about the extent to which people can accurately aggregate across experiences.

Due in part to these concerns, researchers turned to procedures for assessing well-being that reduce reliance on memory and the need for aggregation by respondents. Specifically, *experiential* measures of SWB shift the emphasis from people's evaluations of their life as a whole to their evaluations of the individual experiences that they have (Kahneman and Riis 2005). The potential value of these approaches is highlighted most clearly by Kahneman (1999) in a chapter titled "Objective Happiness." In this chapter, Kahneman argued that something close to an objective measure of SWB can be obtained by repeatedly asking respondents a very simple question that should be unaffected by the concerns raised above: namely, whether, at the current moment, their experience is positive or negative. These momentary evaluations can then be aggregated to get an assessment of the respondent's overall experience.

Although Kahneman (1999) focused on a very specific type of question for his proposed measure of objective happiness (the simple dichotomous evaluation of whether the person was feeling positive or negative), a broad range of experiential measures have been examined as alternatives to global evaluative measures. There is a long history of research that samples people's experiences repeatedly over time, while focusing on the emotions and feelings that people report in those moments (for an overview, see Mehl and Conner 2013). These momentary reports can be aggregated to get an overall sense of the positivity of people's experiences, which could be considered to be a reasonable proxy for an evaluation of their life as a whole (for a discussion of whether such measures should be seen as equivalent, see Diener et al. 2018).

Of course, methods that repeatedly sample experiences over time are quite burdensome both for participants and researchers. To tackle this problem, researchers have developed alternative procedures that streamline assessments while attempting to generate a rich set of data over time. Specifically, Kahneman et al. (2004a) proposed the *day reconstruction method*, in which respondents reconstruct an entire day at a time, listing distinct episodes that they experienced during the day, along with the moods and feelings that they experienced during those episodes. The goal of the day reconstruction method is to obtain rich data about momentary experiences using techniques that are less burdensome than experience sampling, but that do not have the same problems as global measures of SWB. Because the day reconstruction method can be administered in a single assessment, it has been incorporated into several large-scale studies, including the German Socio-Economic Panel Study (Goebel et al. 2019), the Panel Study of Income Dynamics,[3] and the American Time Use Survey.[4] This increasingly widespread adoption of experiential measures shows that they are a promising complement or alternative to traditional global measures.

3 Comparing global reports to experiential measures

If the validity of global reports of SWB is severely affected by the processes described by judgment model researchers, then experiential measures provide a sensible alternative that solves the worst of these problems. Specifically, if the problem with global reports is that people are unable or unwilling to consider all relevant information when making global judgments, and instead rely on whatever information happens to be accessible, experiential measures solve the problem by only asking respondents to consider a narrow moment of time, as that moment occurs. Alternatively, if the primary problem is that people focus on especially salient moments or features when aggregating across a range of experiences, experiential measures can solve this problem by repeatedly assessing moments, while taking the task of aggregation out of the hands of the respondents.

In short, researchers have argued for the benefits (if not the superiority; cf. Kahneman 1999) of using experiential measures instead of global measures precisely because the former directly address concerns that have been raised about the latter. However, it is possible to question whether the evidence that the judgment model researchers have presented really does robustly challenge the validity of global report measures. In the next section, I argue that the evidence for the existence of

[3] https://psidonline.isr.umich.edu/data/Documentation/UserGuide2017.pdf

[4] https://www.bls.gov/tus/atususersguide.pdf

such problems has been overstated. In addition, because experiential measures so clearly solve two problems associated with global measures (problems of memory and aggregation), there has been a sense of complacency about the validity of these experiential measures. Thus, few explicit tests of their validity have been performed. Indeed, I argue that this complacency has prevented researchers from considering whether experiential measures may pose their own unique challenges to validity that do not affect global measures. If this is the case, then the important questions about the relative strengths of the two types of measures of SWB would remain.

3.1 Concerns about the judgment model

Over the past decade, scientists have been grappling with concerns about the replicability of previously published work (e.g., Open Science Collaboration 2015). Although concerns about replicability have touched many areas of science, the field of social psychology—the field that encompasses judgment model research—has been at the heart of this debate. Social psychological studies fared poorly in the first major attempt to replicate large sets of high-profile studies (Open Science Collaboration 2015), and more targeted investigations of specific high-profile findings have also repeatedly failed to replicate initial claims (e.g., O'Donnell et al. 2018; Cheung et al. 2016; Eerland et al. 2016; McCarthy et al. 2018). To be sure, researchers do not know the true rate of replicability of social psychological studies, but the substantial number of failed replications that has emerged has led to a renewed skepticism about even the most foundational studies in the field.

Specific concerns about problematic research practices have strengthened this skepticism. For instance, researchers (not just from social psychology) have admitted to a variety of such practices, including optional stopping (i.e., checking the results before the data collection is complete and stopping once significance is obtained); the strategic omission of outliers, measures, or even experimental conditions that do not support hypotheses; and post hoc theorizing that makes it appear as though unpredicted effects were hypothesized (John et al. 2012; Kerr 1998). When combined with a tendency to use underpowered designs (Button et al. 2013) and a broad bias against publishing null results, these practices can lead to a high rate of false positives in the literature (Ioannidis 2005). Indeed, in a famous investigation of the consequences of these practices, Simmons et al. (2011) showed that support for clearly wrong (and nonsensical) hypotheses can easily be found when a combination of these practices is used.

These issues are important for discussions about the measurement of SWB. When highly cited studies from the judgment model are evaluated from this modern, skeptical perspective, there are reasons for concern. Specifically, almost all original effects in support of the judgment model of SWB came from a single research laboratory; they typically rely on extremely small sample sizes (with per-cell samples often around 10 to 15 participants; see Yap et al. 2017); they often have effect sizes that could be considered implausibly large; and there have been increasing numbers of failures to replicate the basic findings from the original studies. For instance, in the original small sample study showing that asking about one life domain (in this case, dating) before asking about general life satisfaction affected the size of the correlation between the two items, the correlations in the condition where general life satisfaction was asked first were much smaller than those that are typically found. These correlations ranged from −.12 to .16 (Strack et al. 1988), whereas previous research on domain satisfaction consistently shows that satisfaction with just about any domain of life tends to correlate moderately (i.e., in the range of .2 to .5) with global life satisfaction judgments, regardless of the order in which they are assessed (Schimmack and Oishi 2005).

Indeed, years after Strack et al.'s (1988) initial item-order study, Schimmack and Oishi (2005) reviewed all of the studies that they could find that used a similar methodology, and they concluded that these initial results from Strack et al. (1988) were outliers. The average correlation between global satisfaction and narrower domain satisfactions was $r = .32$ when the global measure was asked first, compared to a just slightly larger $r = .40$ when the domain satisfaction was asked first. Importantly, the difference was even smaller when the Strack et al. (1988) studies were excluded. Moreover, Schimmack and Oishi (2005) conducted five new studies, all of which found extremely small differences across conditions. Using a different design, Lucas et al. (2018) also showed that questions asked before a life satisfaction measure do not have a strong influence on responses to that global question. Thus, although making information salient before asking questions about life satisfaction may have small effects on those judgments, the striking effects presented by Strack et al. (1988) appear to be a dramatic overestimate of the typical effect of temporarily accessible information on global well-being judgments.

The concerns about replicability extend to other evidence from the judgment model. For instance, to test whether ease of retrieval affects judgments of personality characteristics, Yeager et al. (2019) used a large, nationally representative sample to test whether being asked to list six versus 12 assertive behaviors affected people's ratings of their own assertiveness. In contrast to the original small sample studies (with *N*s with approximately eight and 20 participants per cell), the effects found in this large study ($N = 1{,}338$ for a two-cell design) were not significant. Again, independent replication that used a high-quality, large sample failed to find results that were consistent with the original findings.

The most well-supported part of the judgment model concerns the mood effects on life satisfaction judgments, which have been found in multiple studies conducted by Schwarz, Strack and their colleagues. Again, however, the original studies often have characteristics associated with low rates of replicability (e.g., small sample sizes, inconsistent measures and analytic approaches, and p-values that are close to .05), and few if any successful replications have been published. Indeed, a growing number of large-scale replications have failed to find effects consistent with the original results.

For instance, Schwarz's (1987) study, which purported to show that finding a dime can affect life satisfaction ratings, had only eight participants in each of the two conditions, and the differences between the two conditions were not significant. In another study that focused on people's reactions to soccer games, Schwarz et al. (1987) called separate groups of respondents before and after two important soccer games, one resulting in a win for the local team, and one resulting in a tie. The authors found a significant interaction between the time (before the game versus after) and the outcome (win versus tie) when predicting life satisfaction and concluded that life satisfaction judgments were influenced by the outcome of the game. However, a closer examination of the results from this study shows that there was a cross-over interaction that was driven as much by theoretically irrelevant and unexplainable differences in life satisfaction before the game as by the differences after the game, and no follow-up comparisons were conducted. Especially considering the small sample size used in the study, the robustness of these results can be questioned.

The study that received the most attention was Study 2 from Schwarz and Clore (1983). In this study, the researchers called respondents on sunny days or rainy days and found that those contacted on rainy days reported lower life satisfaction than those contacted on sunny days. The explanation was that weather affects mood, and respondents rely on their mood when making life satisfaction judgments. However, this study had six separate conditions, each of which included approximately 14 participants. Importantly, five of six conditions were not significantly different from one another (and had means that were quite similar), and just one very large discrepancy drove the entire effect. The difference in this one 14-person cell was quite large, exceeding the decline in SWB that is typically found after the onset of a serious disability or in the first year of widowhood (e.g., Lucas 2007; Lucas et al. 2003). This effect is, arguably, implausibly large given the design of the study.

In recent years, a number of researchers have attempted to replicate this and other mood studies from the judgment model tradition with little success. For instance, Lucas and Lawless (2013) used data from over one million Americans from all 50 states assessed over a five-year period, and found almost no evidence that any form of weather affected life satisfaction judgments, despite having extremely high power to detect even exceedingly small effects. Using panel data, Feddersen et al. (2016) did find significant associations between weather and life satisfaction, but they also had extremely high power, and the effect of weather found in their study was approximately one one-hundredth the size of the effect reported in the original study; i.e., a size that is arguably inconsequential and unimportant for validity.[5] Lucas and Lawless reviewed additional evidence that calls into question the robustness of the original finding of a weather effect.

Finally, because studies like those conducted by Lucas and Lawless (2013) and Feddersen et al. (2016) were conceptual replications that relied on existing data with correlational designs, Yap et al. (2017) conducted nine new experimental studies

---

[5] Simonsohn (2015) calculated that a one-standard-deviation increase in sunshine was associated with .01 additional points of life satisfaction on a 0-10 scale.

using procedures very similar to the classic judgment model mood studies, but with much larger samples. Although significant effects of mood induction procedures on life satisfaction were found in some of their studies (although not for weather), these effects were substantially smaller than those reported in the original study (e.g., *d*s of .09 and .11 for the two focal outcome measures). Thus, it is not clear that these effect sizes were large enough to affect the validity of traditional global well-being measures. These findings imply that concerns raised by judgment model studies about the validity of SWB measures may not be well-founded.

3.2 Testing the validity of experiential measures

The previous sections raised concerns about the robustness of existing evidence that challenges the validity of global measures of SWB. An additional issue that should be considered when comparing different measures is whether the proposed alternatives can solve the problems that global measures are thought to have. Experiential measures, such as those that can be obtained from experience sampling methodologies or the day reconstruction method, address concerns about global measures by reducing reliance on the respondents' ability to remember and aggregate. Thus, these methods address two of the most salient concerns about global measures. However, these potential strengths do not ensure that scores from experiential measures are in fact more valid than global measures of SWB. Indeed, as was noted above, it is not even clear that the concerns that have been raised about global measures translate into demonstrable decrements in validity. Thus, issues of relative validity are empirical questions that must be assessed directly.

Unfortunately, few direct comparisons of the validity of the two types of measures exist. It seems that researchers have relied on the face validity of experiential measures as the justification for their use, without subjecting these measures to more rigorous tests of validity for the purposes of assessing individual- and population-level SWB. For instance, in introducing the day reconstruction method, Kahneman et al. (2004a) simply compared absolute sample-level frequencies and diurnal patterns of affect found using this new method to those found with the existing experience sampling method. Since the publication of this initial report, few explicit tests of validity have been reported.

Admittedly, the lack of testing is attributable in part to some uncertainty about precisely how researchers should validate experiential measures. There is no gold standard measure that can be used as a criterion, and—as I have noted—widely used global measures have been criticized. Thus, the use of experiential measures has been justified primarily based on their assumed beneficial attributes. In recent years, however, my colleagues and I have attempted to address this oversight by considering additional approaches to validating these methods.

For instance, if experiential measures like those obtained from the experience sampling method and the day reconstruction method are valid, they should provide similar results when used to assess well-being in the same sample over the same time period. Initial investigations that compared the two methods (e.g., Kahneman et al. 2004a) used different samples and only compared sample-level statistics. Studies like this, however, provide an incomplete picture of the convergence across the two methods.

Recently, we asked two samples of students to complete a day of experience sampling (in which they reported on what they were doing and how they were feeling up to nine times per day), followed by a reconstruction of the exact same day using the day reconstruction method (Lucas et al. 2021). This allowed us to compare responses at multiple levels: the sample level (e.g., were the overall sample-level estimates of time spent in specific situations and the average affect similar across methods?), the person level (e.g., did people who reported being happy with others using one method also report being happy with others using the other method?) and the moment level (e.g., when people reported being happy at a particular moment using one method, did they report being similarly happy using the other method?). Results varied depending on the level examined.

When looking at sample-level statistics, the level of agreement was quite high. For instance, when comparing estimates of time spent in various situations, the correlations across the two methods of assessment (using situations as the unit of analysis) exceeded .95. Similarly, when looking at the average positive and negative affect reported in each situation, the correlations across methods were again quite high, ranging from .76 to .85. Thus, if the goal was to estimate sample-level statistics about these experiences, both methods provided similar information.

However, when moving to the person level of analysis, the results were more varied and problematic. First, when assessing individual differences in affect experienced over the course of the day, the two methods converged reasonably well: the between-person correlations (i.e., the correlation between individual differences in average affect estimated with the ESM and average affect estimated with the DRM) for positive and negative affect were .83 and .88, respectively. Yet reports of time spent in specific situations varied across types of situations: person-level agreement was high (ranging from .74 to .83) for reports of how much time a person spent in specific locations, like home, school or work; but it was relatively low for reports of specific activities, like commuting ($r = .25$), eating ($r = .21$), or doing housework ($r = .28$).

Finally, when looking within individuals over time, reports of both affect and activities diverged across methods. These analyses assessed the changes in a person's affect and activities over time and examined whether the changes measured using one method corresponded to the changes measured using the other. For within-person affect, correlations ranged from $r = .34$ for positive affect to $r = .36$ for negative affect (in Study 1; results were similar for Study 2). We calculated within-person *kappa* coefficients to assess the level of agreement for dichotomous situation variables, and the average *kappa* across situations was .34. Thus, it appears that two forms of experiential measures—experience sampling and day reconstruction—did not provide the same information when used to assess a day of experience. These results raise some concerns about the validity of these measures.

Moreover, although there is no gold standard against which to compare either global or experiential measures, it is possible to compare both sets of reports to alternative measures that have different strengths and weaknesses. Lucas et al. (2021) obtained informant reports of life satisfaction from friends and family members who knew the respondent well and used these as a criterion that could be predicted from global self-reports, experience-sampling-based reports, and day-reconstruction-based reports. Although the differences in the correlations were not significant, these associations were somewhat higher for the global measures than for the experiential measures, which suggests that at the very least, the validity of the experiential measures did not exceed the validity of the global reports when informant reports were used as a criterion (also see Hudson et al. 2020, 2017; and Hudson et al. 2019a, for similar results). An additional study that looked at the associations over time between well-being and health also found that the associations were slightly stronger for the global measures than for the experiential measures (Hudson et al. 2019b). Overall, these results suggest that there is at least some cause for concern about the validity of experiential measures; and that at the very least, it should not be assumed that such measures provide scores that are more valid than those provided by traditional global measures.

3.3 Concerns about experiential measures

The goal of this paper is not simply to defend global measures and critique experiential alternatives. Instead, the goal is to raise awareness about asymmetries in the ways these classes of measures have been evaluated. Critiques of global measures often proceed by focusing on the processes that underlie the judgments that respondents are asked to make. Experiential measures are often assumed to have advantages over global measures precisely because they do not rely on the same underlying processes. However, as the evidence reviewed in the previous section shows, these fundamental differences in these measures do not always translate into improved reliability and validity for experiential alternatives to global reports. Moreover, few researchers have considered whether there are additional psychological processes underlying these experiential measures that have the potential to systematically (and negatively) affect their psychometric properties. In this final section, I discuss some potential concerns about the processes that underlie experiential measures.

Experiential measures clearly solve two challenges of global measures of SWB: they reduce reliance on memory, and they avoid the need for aggregation on the part of the respondent. However, SWB researchers have, for the most part, failed to consider whether these measures have unique problems that global measures avoid. For instance, it is possible that the increased respondent burden associated with experiential measures affects the quality of the responses that respondents provide (Eisele et al. 2020). In addition, because of logistical challenges associated with experiential measures, these measures are often more difficult to implement than global measures in large-scale survey work. This means that selection bias may affect studies that use experiential measures more than studies that rely on global measures, as studies that use experiential measures must often rely on weaker sampling plans.

One salient difference between experiential measures and global measures is that the former measures require respondents to answer the same question multiple times so that multiple experiences can be assessed. Moods and other evaluative experiences fluctuate from moment to moment (e.g., Epstein 1979); therefore, to assess something that is representative of a person's life, multiple moments must be aggregated. Thus, respondents must report on many such occasions, either as the occasions occur (in experience sampling), or soon after they have happened (as in the day reconstruction method). However, simply asking the same question repeatedly can change the way that respondents interpret the question, which can, in turn, affect the validity (e.g., Meade and Craig 2012).

Schwarz (1999) described a model of survey response that detailed how the specific questions that researchers ask can shape the answers that respondents provide. Although researchers may assume that respondents interpret survey questions literally, subtle features of the testing environment can influence respondents' interpretations, which can, in turn, affect their responses. As just one relevant example, Strack et al. (1991) presented two similar questions to respondents: one question about happiness and one about life satisfaction. When these two questions were presented as if they were part of two separate questionnaires, respondents provided very similar responses to the two questions. However, when the questions were presented as part of the same questionnaire, respondents appeared to interpret the questions differently, and provided more discrepant responses.[6] Specifically, Strack et al. (1991) argued that in the condition in which the two questions were presented as a part of the same questionnaire, respondents assumed that the researcher would not ask the same question twice, and therefore interpreted the two questions in subtly different ways. Schwarz (1999) noted that findings like this suggest that respondents rely on *conversational norms* to interpret questions; and one important norm is that people do not ask for redundant information.

It is not difficult to see how such a conversational norm could affect experiential measures; i.e., measures that ask respondents to answer the same question over and over again across multiple experiences. It is possible that by presenting questions in this way, respondents interpret the researchers' focus as being on the *change* that occurs from moment to moment, rather than on conditions that remain stable. Thus, respondents may be more likely to report on this change, resulting in exaggerated reports of variability.

[6] Note, however, that this study also has similar characteristics to those from the judgment model, including a very small sample size and somewhat implausible effect sizes. Therefore, even this result should be interpreted with caution until it is replicated in a large-scale, pre-registered study.

At least some evidence exists to support this possibility. In a slightly different context—i.e., an investigation into personality variability—Baird and Lucas (2011) tested whether asking the same personality questions multiple times led respondents to report more variability than they otherwise would have. All participants first provided a global evaluation of their personality in an online questionnaire. Approximately one week later, they appeared in person to complete an additional set of questionnaire measures. Participants were randomly assigned to take part in one of two conditions. In the single-role condition, participants were asked to answer a set of 50 different questions about their personality in a single role (e.g., what their personality is like as a friend, as a worker or as a student). In the multiple-role condition, participants were asked a set of 10 different questions five separate times, once for each role. The hypothesis was that the variability in the responses that participants provided (as assessed by comparing each role-specific score to the general score provided a week earlier) would be greater in the multiple-role condition than in the single-role condition.

The results supported this hypothesis. Compared to the personality scores in the multiple-role condition, the personality scores in the single-role condition were more strongly correlated with and less discrepant in an absolute sense from the global scores. Thus, it appeared that compared to the reports they gave when only asked about a single role, simply asking the same question multiple times led respondents to exaggerate the variability of their personality. Moreover, participants provided responses in the multiple-role condition that were more in line with stereotypes of how people typically behave in specific situations. For instance, participants reported greater differences between the ratings of what their extraversion was like in general and what their extraversion was like as a friend in the multiple-role condition than in the single-role condition. Although it is not clear whether such effects occur in standard experiential measures of well-being, Lucas et al. (2021) reported suggestive evidence that this could be the case, especially for the day reconstruction method. Thus, the unique features of experiential measures—especially the fact that these measures rely on repeated questions—may affect the validity of the scores that result when these measures are used.

These results do not directly address the validity of experiential measures of SWB. However, like the judgment model studies of global reports, they provide insight into the processes by which participants create and report responses to the questions that are posed. Baird and Lucas's (2011) study suggests that simply asking the same question over and over may communicate to respondents that the researchers are more interested in change than in stability. This may, in turn, lead to responses that exaggerate the impact of changing situational factors on the resulting aggregate ratings. Indeed, experiential measures like those obtained from the day reconstruction method are often more strongly correlated with situational variables assessed at the same time and using the same method than they are with other major life circumstances that are often correlated with more global reports (Lucas et al. 2021). Although it is possible that this is because these situational factors truly do affect momentary experiences, it is also quite possible that these associations appear because the use of experiential measures makes these situational factors especially salient to respondents. Again, more work that directly evaluates the validity (and the relative validity compared to global reports) of experiential measures is needed.

4 Summary

I want to be clear that in reviewing this research, I am not making the claim that the validity of global reports of well-being clearly exceeds that of experiential measures.[7] Indeed, there are many remaining questions about the validity of both types of measures, and much additional research should be done to clarify these issues. However, many prior reviews of the literature on the measurement of well-being have made two claims that I believe should be reconsidered. First, these reviews have suggested that the attention-grabbing findings from the judgment model tradition show that global measures of SWB are substantially affected by irrelevant features that reduce the reliability, stability and validity of the measures of the responses that lead to the results. The first goal of this paper was to review this evidence with a critical eye. A careful look at this literature suggests that the initial studies and results have characteristics that should raise questions about their replicability, and an increasing number of replications have failed to obtain results that were anywhere near the size of the original studies. Although the processes identified by judgment model researchers may influence well-being judgments, the bulk of the evidence suggests that the effects are small, and are unlikely to substantially affect the validity of the resulting measures.

The second claim that prior critiques of global SWB measures often made was that by solving the problems of memory and aggregation, experiential measures may represent a more valid choice for measuring SWB. Although experiential measures clearly do have these benefits, these features do not guarantee that the measures are more valid than alternatives. For one thing, the memory and aggregation problems may not have been so bad in the first place. In addition, experiential measures may have their own problems that global measures do not have. In this paper, I argued that researchers should not assume the validity of experiential measures, and that their validity must be examined empirically. I also highlighted some initial evidence that should raise some concerns about the validity and the utility of these measures, especially given their costs and burdens.

To be sure, more work on both global and experiential measures is needed. Just because the validity of global measures is not strongly affected by the processes identified by judgment model researchers does not mean that other threats to validity

[7] Though a case has been made that even if global and experiential measures are valid indicators of somewhat different constructs, global measures like life satisfaction may capture something closer to what is typically meant by SWB or may be more useful for policy purposes (Diener et al. 2009; Frijters and Krekel 2021).

do not exist. Questions about inter- and intra-personal comparability have been raised, and the answers to these questions will affect how researchers interpret responses to SWB measures. Similarly, just because challenges to the validity of experiential measures can be proposed, this does not mean that these measures do not have substantial levels of validity. It may even be the case that for some purposes, experiential measures are more valid than global measures. Indeed, studies that include both types of measures have the potential to further our understanding of the strengths and weaknesses of each. Given the wide range of applications of SWB research, an increased focus on these issues will help the field of SWB and all other fields that rely on SWB measures.

References

Baird, B. M. and R. E. Lucas 2011. "…And how about now?": Effects of item redundancy on contextualized self-reports of personality. *Journal of Personality* 79(5): 1081–1112. https://doi.org/10/fg647f

Button, K. S., J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson and M. R. Munafò 2013. Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* 14(5): 365–376. https://doi.org/10.1038/nrn3475

Cheung, I., L. Campbell, E. P. LeBel, R. A. Ackerman, B. Aykutoğlu, Š. Bahník, J. D. Bowen, C. A. Bredow, C. Bromberg, P. A. Caprariello, R. J. Carcedo, … and J. C. Yong 2016. Registered replication report: Study 1 From Finkel, Rusbult, Kumashiro, and Hannon (2002). *Perspectives on Psychological Science* 11(5): 750–764. https://doi.org/10.1177/1745691616664694

Deaton, A. and A. A. Stone 2016. Understanding context effects for a measure of life evaluation: How responses matter. *Oxford Economic Papers* 68(4): 861–870. https://doi.org/10.1093/oep/gpw022

Diener, E., R. A. Emmons, R. J. Larsen and S. Griffin 1985. The satisfaction with life scale. *Journal of Personality Assessment* 49: 71–75. https://doi.org/10/fqqbmr

Diener, E., R. E. Lucas and S. Oishi 2018. Advances and open questions in the science of subjective well-being. *Collabra: Psychology* 4(1):15. https://doi.org/10.1525/collabra.115

Diener, E., R. E. Lucas, U. Schimmack and J. Helliwell 2009. *Well-being for public policy*. Oxford University Press, USA. https://doi.org/10.1093/acprof:oso/9780195334074.001.0001

Diener, E., E. M. Suh, R. E. Lucas and H. L. Smith 1999. Subjective well-being: Three decades of progress. *Psychological Bulletin* 125: 276–302. https://doi.org/10.1037/0033-2909.125.2.276

Dolan, P. and M. P. White 2007. How can measures of subjective well-being be used to inform public policy? *Perspectives on Psychological Science* 2(1): 71–85. https://doi.org/10.1111/j.1745-6916.2007.00030.x

Eerland, A., A. M. Sherrill, J. P. Magliano, R. A. Zwaan, J. D. Arnal, P. Aucoin, … J. M. Prenoveau 2016. Registered replication report: Hart & Albarracín (2011). *Perspectives on Psychological Science* 11(1): 158–171. https://doi.org/10.1177/1745691615605826

Eisele, G., H. Vachon, G. Lafit, P. Kuppens, M. Houben, I. Myin-Germeys and W. Viechtbauer 2020. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. Preprint. PsyArXiv. https://doi.org/10.31234/osf.io/zf4nm

Epstein, S. 1979. The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology* 37(7): 1097–1126. https://doi.org/10/c6w73c

Feddersen, J., R. Metcalfe and M. Wooden 2016. Subjective well-being: Why weather matters. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 179(1): 203–228. https://doi.org/10.1111/rssa.12118

Fredrickson, B. L. and D. Kahneman 1993. Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology* 65(1): 45–55. https://doi.org/10/dc7rrm

Frijters, P. and C. Krekel 2021. *A handbook for wellbeing policy-making: History, measurement, theory, implementation, and examples*. Oxford: Oxford University Press.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder and J. Schupp 2019. The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik* 239(2): 345–360. https://doi.org/10.1515/jbnst-2018-0022

Hudson, N. W., I. Anusic, R. E. Lucas and M. B. Donnellan 2020. Comparing the reliability and validity of global self-report measures of subjective well-being with experiential day reconstruction measures. *Assessment* 27(1): 102–116. https://doi.org/10.1177/1073191117744660

Hudson, N. W., R. E. Lucas and M. B. Donnellan 2017. Day-to-day affect is surprisingly stable: A 2-year longitudinal study of well-being. *Social Psychological and Personality Science* 8(1): 45–54. https://doi.org/10.1177/1948550616662129

Hudson, N. W., R. E. Lucas and M. B. Donnellan 2019a. A direct comparison of the temporal stability and criterion validities of experiential and retrospective global measures of subjective well-being. Manuscript in Preparation, Dallas, TX: Southern Methodist University.

Hudson, N. W., R. E. Lucas and M. B. Donnellan 2019b. Healthier and happier? A 3-year longitudinal investigation of the prospective associations and concurrent changes in health and experiential well-being. *Personality and Social Psychology Bulletin* 45(12): 1635–1650. https://doi.org/10.1177/0146167219838547

Hurka, T. 2014. Objective goods. In *The Oxford Handbook of Well-Being and Public Policy*, eds M. D. Adler and M. Fleurbaey, 379–402. Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199325818.013.12

Ioannidis, J. P. 2005. Why most published research findings are false. *PLoS Medicine* 2(8): e124. https://doi.org/10/chhf6b

Ito, T. A. and J. T. Cacioppo 1999. The psychophysiology of utility appraisals. In *Well-being: The Foundations of Hedonic Psychology*, eds D. Kahneman, E. Diener and N. Schwarz, 470–488. New York, NY: Russell Sage Foundation.

John, L. K., G. Loewenstein and D. Prelec 2012. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science* 23(5): 524–532. https://doi.org/10/f33h6z

Kahneman, D. 1999. Objective happiness. In *Well-being: The foundations of hedonic psychology*, eds D. Kahneman, E. Diener and N. Schwarz, 3–25. New York, NY: Russell Sage Foundation.

Kahneman, D., B. L. Fredrickson, C. A. Schreiber and D. A. Redelmeier 1993. When more pain is preferred to less: Adding a better end. *Psychological Science* 4(6): 401–405. https://doi.org/10/djnzpd

Kahneman, D., A. B. Krueger, D. A. Schkade, N. Schwarz and A. A. Stone 2004a. A survey method for characterizing daily life experience: The day reconstruction method. *Science* 306(5702): 1776–1780. https://doi.org/10.1126/science.1103572

Kahneman, D., A. B. Krueger, D. A. Schkade, N. Schwarz and A. A. Stone 2004b. Toward national well-being accounts. *American Economic Review* 94(2): 429–434. https://doi.org/10.1257/0002828041301713

Kahneman, D. and J. Riis 2005. Living, and thinking about it: Two perspectives on life. In *The science of well-being*, eds F. A. Huppert, N. Baylis and B. Keverne, 285–304. New York: Oxford University Press, USA. https://doi.org/10.1093/acprof:oso/9780198567523.003.0011

Kerr, N. L. 1998. HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review* 2(3): 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Krueger, A. B., D. Kahneman, D. A. Schkade, N. Schwarz and A. A. Stone 2009. National time accounting: The currency of life. In *Measuring the subjective well-being of nations: National accounts of time use and well-being*, ed A. B. Krueger, 9–86. Chicago: University of Chicago Press.

Lucas, R. E. 2007. Long-term disability is associated with lasting changes in subjective well-being: Evidence from two nationally representative longitudinal studies. *Journal of Personality and Social Psychology* 92(4): 717–730. https://doi.org/10.1037/0022-3514.92.4.717

Lucas, R. E. 2018. Reevaluating the strengths and weaknesses of self-report measures of subjective well-being. In *Handbook of Well-Being*, eds E. Diener, S. Oishi, and L. Tay. Salt Lake City: DEF Publishers.

Lucas, R. E., A. E. Clark, Y. Georgellis and E. Diener 2003. Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status. *Journal of Personality and Social Psychology* 84(3): 527–539. https://doi.org/10/d3pbg4

Lucas, R. E., V. A. Freedman and J. C. Cornman 2018. The short-term stability of life satisfaction judgments. *Emotion* 18(7): 1024–1031. https://doi.org/10.1037/emo0000357

Lucas, R. E. and N. M. Lawless 2013. Does life seem better on a sunny day? Examining the association between daily weather conditions and life satisfaction judgments. *Journal of Personality and Social Psychology* 104(5): 872–884. https://doi.org/10.1037/a0032124

Lucas, R. E., S. Oishi and E. Diener 2016. What we know about context effects in self-report surveys of well-being: Comment on Deaton and Stone. *Oxford Economic Papers* 68: 871–876. https://doi.org/10/gh5sxc

Lucas, R. E., C. Wallsworth, I. Anusic and M. B. Donnellan 2021. A direct comparison of the day reconstruction method (DRM) and the experience sampling method (ESM). *Journal of Personality and Social Psychology* 120(3): 816–835. https://doi.org/10.1037/pspp0000289

McCarthy, R. J., J. J. Skowronski, B. Verschuere, E. H. Meijer, A. Jim, K. Hoogesteyn, R. Orthey, O. A. Acar, B. Aczel, B. E. Bakos, F. Barbosa … and E. Yıldız 2018. Registered Replication Report on Srull and Wyer (1979). *Advances in Methods and Practices in Psychological Science* 1(3): 321–336. https://doi.org/10.1177/2515245918777487

Meade, A. W. and S. B. Craig 2012. Identifying careless responses in survey data. *Psychological Methods* 17(3): 437–455. https://doi.org/10.1037/a0028085

Mehl, M. R. and T. S. Conner, eds. 2013. *Handbook of Research Methods for Studying Daily Life*. New York, NY: Guilford Press.

National Academy of Sciences 2013. *Subjective Well-Being: Measuring Happiness, Suffering, and Other Dimensions of Experience*. National Academies Press. https://doi.org/10.17226/18548

O'Donnell, M., L. D. Nelson, E. Ackermann, B. Aczel, A. Akhtar, S. Aldrovandi, N. Alshaif, R. Andringa, M. Aveyard, P. Babincak, N. Balatekin … and M. Zrubka 2018. Registered Replication Report: Dijksterhuis and van Knippenberg (1998). *Perspectives on Psychological Science* 13(2): 268–294. https://doi.org/10.1177/1745691618755704

OECD 2013. *OECD Guidelines on Measuring Subjective Well-being.* OECD Publishing, Paris. https://doi.org/10.1787/9789264191655-en

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349(6251): aac4716. https://doi.org/10.1126/science.aac4716

Redelmeier, D. A. and D. Kahneman 1996. Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain* 66(1): 3–8. https://doi.org/10/czjgb4

Redelmeier, D., J. Katz and D. Kahneman 2003. Memories of colonoscopy: A randomized trial. *Pain* 104(1-2): 187–194. https://doi.org/10.1016/S0304-3959(03)00003-4

Robinson, M. D. and G. L. Clore 2002. Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin* 128(6): 934–960. https://doi.org/10.1037/0033-2909.128.6.934

Ryan, R. M. and E. L. Deci 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55(1): 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryff, C. D. 1989. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology* 57(6): 1069–1081. https://doi.org/10.1037/0022-3514.57.6.1069

Schimmack, U. and S. Oishi 2005. The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of Personality and Social Psychology* 89(3): 395–406. https://doi.org/10.1037/0022-3514.89.3.395

Schwarz, N. 1987. *Stimmung als Information: Untersuchungen zum Einfluss von Stimmungen auf die Bewertung des eigenen Lebens*. Berlin: Springer.

Schwarz, N. 1999. Self-reports: How the questions shape the answers. *American Psychologist* 54(2): 93–105. https://doi.org/10/fqrx56

Schwarz, N., H. Bless, F. Strack, G. Klumpp, H. Rittenauer-Schatka and A. Simons 1991. Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology* 61(2): 195–202. https://doi.org/10.1037/0022-3514.61.2.195

Schwarz, N. and G. L. Clore 1983. Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology* 45(3): 513–523. https://doi.org/10.1037/0022-3514.45.3.513

Schwarz, N. and F. Strack 1999. Reports of subjective well-being: Judgmental processes and their methodological implications. In *Well-being: The foundations of hedonic psychology*, eds D. Kahneman, E. Diener, and N. Schwarz, 61–84. New York, NY: Russell Sage Foundation.

Schwarz, N., F. Strack, D. Kommer and D. Wagner 1987. Soccer, rooms, and the quality of your life: Mood effects on judgments of satisfaction with life in general and with specific domains. *European Journal of Social Psychology* 17(1): 69–79. https://doi.org/10.1002/ejsp.2420170107

Simmons, J. P., L. D. Nelson and U. Simonsohn 2011. False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22(11): 1359–1366. https://doi.org/10/bxbw3c

Simonsohn, U. 2015. Small telescopes: Detectability and the evaluation of replication results. *Psychological Science* 26(5): 559–569. https://doi.org/10.1177/0956797614567341

Strack, F., L. L. Martin and N. Schwarz 1988. Priming and communication: Social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology* 18: 429–442. https://doi.org/10.1002/ejsp.2420180505

Strack, F., N. Schwarz and M. Wänke 1991. Semantic and pragmatic aspects of context effects in social and psychological research. *Social Cognition* 9(1): 111–125. https://doi.org/10.1521/soco.1991.9.1.111

Testoni, S., L. Mansfield and P. Dolan 2018. Defining and measuring subjective well-being for sport policy. *International Journal of Sport Policy and Politics* 10(4): 815–827. https://doi.org/10.1080/19406940.2018.1518253

Tully, S. and T. Meyvis 2016. Questioning the end effect: Endings are not inherently overweighted in retrospective evaluations of experiences. *Journal of Experimental Psychology: General* 145(5): 630–642. https://doi.org/10.1037/xge0000155

Yap, S. C. Y., J. Wortman, I. Anusic, S. Glenn, L. D. Scherer, M. B. Donnellan and R. E. Lucas 2017. The effect of mood on judgments of subjective well-being: Nine tests of the judgment model. *Journal of Personality and Social Psychology* 113(6): 939–961. https://doi.org/10.1037/pspp0000115

Yeager, D. S., J. A. Krosnick, P. S. Visser, A. L. Holbrook and A. M. Tahk 2019. Moderation of classic social psychological effects by demographics in the U.S. Adult population: New opportunities for theoretical advancement. *Journal of Personality and Social Psychology* 117(6): e84–e99. https://doi.org/10.1037/pspa0000171

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 91–123

Anthropometric history and the measurement of wellbeing

Bernard Harris[1,*]

Abstract

It has often been recognised that the average height of a population is influenced by the economic, social and environmental conditions in which it finds itself, and this insight has inspired a generation of historians to use anthropometric data to investigate the health and wellbeing of past populations. This paper reviews some of the main developments in the field, and assesses the extent to which height remains a viable measure of historical wellbeing. It explores a number of different issues, including the nature of human growth; the impact of variations in diet and exposure to disease; the role of ethnicity; the relationships between height, mortality and labour productivity; and the "social value" of human stature. It concludes that, despite certain caveats, height has retained its capacity to act as a "mirror" of the conditions of past societies, and of the wellbeing of their members.

Keywords: anthropometrics; height; health; wellbeing; standard of living

1 Introduction

As Floud (1997, 1–2) once observed, economic historians have often seemed to place more emphasis on production than consumption. He argued that this emphasis was misplaced, since "the central purpose of economic life and economic growth [is] an improvement in the condition of the people". However, although the reasons for focusing on the condition of the people may seem obvious (especially in a volume devoted to the "demographic aspects of human wellbeing"), the question of how it should be measured has proven much more controversial. This debate has been echoed in the efforts made by students of contemporary societies to examine the

[1]School of Social Work and Social Policy, University of Strathclyde, Glasgow, UK

[*]Correspondence to: Bernard Harris, bernard.harris@strath.ac.uk

DOI: 10.1553/populationyearbook2021.rev02

relationship between the "condition of the people", the "standard of living", and "wellbeing". As Sen (1987, 1) explained:

> Within the general notion of the living standard, divergent and rival views of the goodness of life coexist in an unsorted bundle. There are many fundamentally different ways of seeing the quality of living, and quite a few of them have some immediate plausibility. You could be *well off*, without being *well*. You could be *well*, without being able to lead the life you *wanted*. You could have got the life you *wanted*, without being *happy*. You could be *happy*, without having much *freedom*. You could have a good deal of *freedom*, without achieving much. We can go on.

Economic historians have sought to measure changes in the standard of living in a variety of ways, including through the use of income-based measures, such as real wages or GDP per capita, as well as health-based measures (e.g., Feinstein 1998; Clark 2005; Allen 2009), such as age-specific mortality and life expectancy (e.g., Szreter and Mooney 1998). A number of writers have also sought to construct composite indices, such as measures based on the Human Development Index (Floud and Harris 1997; Crafts 1997; Prados de la Escosura 2019) or other combinations, including measures of income, health, time-use and inequality (Gallardo-Albarrán and DeJong 2021; see also Jones and Klenow 2016). However, many writers have also attempted to investigate changes in living standards using a range of anthropometric measures (Steckel 1995, 2009; Galofré-Vilà 2018). The vast majority of these studies have focused on the analysis of changes in average height, although some have also investigated changes in weight and BMI.[1]

As Steckel (1995) explained more than 20 years ago, the average height of a population offers a number of advantages as a measure of living standards and wellbeing. First, it is itself a form of composite measure, since it reflects the net impact of the dietary resources consumed by human beings, and the demands made on those resources by energy use and, in particular, disease. Second, it has functional consequences that individuals might themselves have "reason to value" (Sen 1999, 14), such as lower mortality, higher earning capacity, and even general happiness (see also Deaton and Arora 2009). Height data also have the advantage of being available for periods and places for which other data, such as wage or mortality data, may be lacking. This has become even more true in recent years, as new investigations have highlighted the availability of skeletal records, which have been used to extend the anthropometric record back into the very distant past (Galofré-Vilà et al. 2018; Steckel et al. 2019).

However, although the use of height data has become increasingly popular in recent years, it has also provoked criticism. Some authors have questioned the value of height data on the grounds that there is "no way of measuring the exchange rate

[1] The BMI, or body mass index, is equal to weight in kilogrammes divided by the square of height in metres.

of height for real income" (Crafts 1992, 428). Others have debated the conclusions drawn from specific studies, either because of disagreements over the use of particular statistical procedures (see, e.g., Komlos 1993a), or because of doubts about the impact of possible selection biases (Bodenhorn et al. 2013, 2015, 2017). A third group of authors have drawn attention to the issues raised by claims that height is particularly sensitive to the consumption of certain foodstuffs (Mokyr and Ó Gráda 1996), while others have suggested that the value of international height comparisons may be vitiated by the impact of ethnicity (A'Hearn 2016). Some of the assumptions that underpinned the initial contributions to this debate have also been thrown into doubt by improvements in our understanding of the physiology of human growth (see, e.g., Wells 2016). Moreover, the relationship between height and the economic and environmental factors to which it has been linked may change over time (Alter 2004).

Given these developments, this paper asks whether the average height of a historical population can continue to be regarded as a valuable measure of its wellbeing. The paper begins by looking at the background to the emergence of "anthropometric history", and then provides a brief summary of the sources that have been used to investigate average height, and of the different factors that are believed to have influenced height. Next, it discusses the literature on the question of "critical periods" and the relationship between child and adult welfare. Section 6 reviews some of the technical debates that have either enlivened or bedevilled the field since the earliest publications, and their relationships to other welfare indicators. Sections 7 and 8 explore the relationship between height and diet and disease, and Section 9 discusses recent attempts to assess the relationship between height and ethnicity. The final sections look at the extent to which height can be used to "predict" aspects of wellbeing in later life, and at the "social" value attached to height in areas such as the marriage market.

2 Background

Over the last 40 years, economic, social, demographic and medical historians have explored the use of height data in historical studies in different ways. During the 1950s, 1960s and 1970s, Thomas McKeown and his collaborators argued that the principal causes of the decline of mortality in Britain and in other European countries were improvements in diet and in the "standard of living" (McKeown and Brown 1955; McKeown and Record 1962; McKeown et al. 1972; McKeown 1976). Yet this conclusion was (and has remained) highly contentious (for a recent summary, see Harris and Helgertz 2019). Fogel et al. (1978, 42) argued that one of the reasons for this uncertainty was "the virtual absence of data bearing on the amount and nutritional adequacy of the food supply", and suggested that height data might provide an indirect way of addressing this knowledge gap. However, it soon became apparent that using height data might be less straightforward than it initially appeared, because height is not simply a measure of dietary input, but also reflects the demands made on people by a wide range of external factors, including their workload and disease environment (see, e.g., Fogel 1986, 446–7).

Although the study of human height first attracted the attention of many historians in the context of debates about the decline in mortality, it also has profound implications for other debates in economic and social history. As Gallardo-Albarrán and DeJong (2021) have recently highlighted, there has been a (very) long-running debate over changes in the standard of living during the British industrial revolution, and this debate has often turned on the relative importance attached to real wages as opposed to a wider range of welfare indicators. Floud (1984a, 14–15) argued that it is precisely because height reflects the impact of a range of indicators that it has the power to contribute to this debate, as the following extract demonstrates:

> To sum up this catalogue of problems, neither computations of the real wage nor those of national income per head seem easily to meet [Daniel] Usher's (1980, 2) criterion that they should be more than "mere numbers with no apparent effect upon our lives and no status as indicators of progress towards goals than people might want the economy to achieve". They give us only a most inadequate idea of the impact which the transformation of European society has had upon the lives of Europeans.
>
> This is a dispiriting conclusion. But there has recently emerged an alternative source of information about the welfare of Europeans in the past which will at least supplement, and for some purposes replace, the traditional measures of welfare. This information lies in the millions of observations held in European archives of the physical height of people in the past. It has long been known – both from economic study and from common observation – that people have been growing taller, but it has only recently become clear that that fact carries with it much information about their welfare (Floud 1984a, 14–15).

This debate is directly related to the debates associated with the introduction of the concept of the "biological standard of living" by John Komlos. Komlos (1987, 921) argued that the average heights of successive birth cohorts of different sections of the United States population did not always move in the same direction as conventional economic indicators during the years leading up to the outbreak of the US Civil War.[2] Although the average value of men's heights was declining, the average value of real wages was increasing. Komlos argued that this indicated that a different conception of the standard of living was needed:

> The above argument suggests that the human biological system can experience periods of stress even as aggregate output per capita grows significantly. Thus anthropometric measurements do not appear to be

[2] This interpretation has been challenged by Bodenhorn et al. (2017); for Komlos's response, see Komlos and A'Hearn (2016, 2019); Komlos (2019) and Section 6 below.

> perfect proxies for the material wellbeing of the population. Rather they might be considered a component of the biological standard of living, which under certain conditions might diverge in significant ways from such conventionally-defined standards as output per capita.

Although the concept of the "biological standard of living" has gained widespread currency in the last 30 years, its value has been challenged. Although height might be regarded as a *biological measure* of the standard of living (see also Harris 2009, 60), it also needs to be placed alongside other biological measures, such as morbidity or mortality. The distinction between the "biological standard of living" and "material wellbeing" also implies that the concept of material wellbeing can itself be reduced to "such conventionally-defined standards as output per capita". This assertion runs directly counter to the claims made by those who believe that *any* meaningful concept of the standard of living must also include such factors as "health, longevity and the quality of life" (Floud et al. 2011, 14).

Although many economic historians have advocated the use of height data based on the claim that they provide a "mirror of the condition of society" (Tanner 1987), some writers have argued that height can also be regarded as an important health indicator in its own right. Oddy (1982, 121–5) observed that height data enable medical historians to measure the health of past generations directly, rather than relying on the inversion of measures such as mortality – which, as he pointed out, measures the absence of health, rather than its presence. Costa and Steckel (1997, 71–2) also advocated the use of height as a health indicator in their attempts to develop a historical version of the Human Development Index. They noted that contemporary efforts to measure human development are often based on a composite index incorporating measures of GDP per capita, life expectancy and literacy. Although they recognised that height and life expectancy are not precisely analogous, they argued that it is possible to apply this measure to populations for whom mortality data are lacking by substituting height for life expectancy.

Although Costa and Steckel's paper played an important role in demonstrating the value of height data in a wider context, it also highlighted a potential source of tension between approaches that seek to incorporate height into other welfare indexes, and approaches that treat height as a (relatively) comprehensive indicator in its own right. In a previous paper, Steckel (1992, 284) had argued that one of the attractions of using height data was the extent to which they capture the influence of a wide range of factors associated with Sen's efforts to measure wellbeing. This tension reflects a longstanding debate between those who regard height as a supplement to conventional measures of wellbeing, and those who believe it offers an alternative to these measures (see also Section 7 below):

> Average height is particularly adept at assessing degrees of deprivation, a feature that places the measure nicely within the basic-needs approach to living standards. While the basic needs approach has been criticised for the conceptual problems associated with ascertaining what is basic, in many ways average height finesses this problem because it is a

> measure of net nutrition. Average height incorporates the extent to which individuals have greater needs created by factors such as a harsher disease environment or greater workloads. In this vein, average height is also conceptually consistent with [Amartya] Sen's framework of functionings and capabilities, though, of course, height registers primarily conditions of health during the growing years as opposed to one's status with respect to commodities more generally.

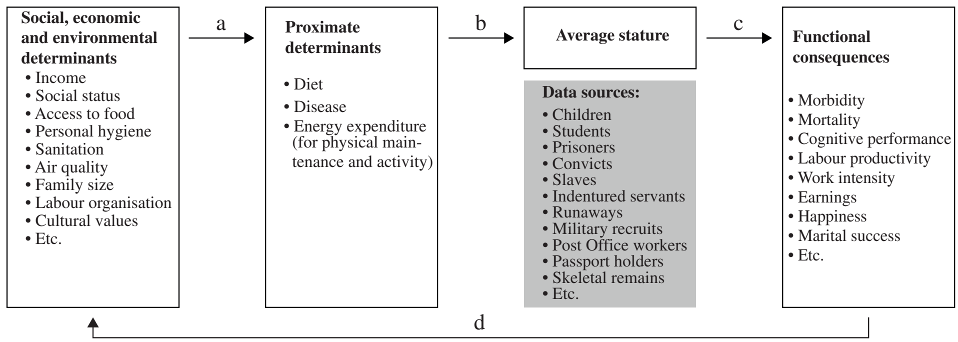
This section has focused on the use of height as a proxy for variables associated with the determination of mortality trends, as a "mirror of the condition of society", and as a health indicator in its own right. However, it could be argued that for height to function as a measure of wellbeing, it also needs to be correlated with other aspects of people's lives, which, in Sen's words, they might have "reason to value" (Sen 1999, 14). Anthropometric historians have sought to explore these dimensions by examining the relationship between height as a summary measure of the living conditions people experience in childhood, and different aspects of people's lives in adulthood. As Floud et al. (2011, 20–5) have argued, height "matters" not only because of what it reveals about the impact of childhood conditions on growth, but also because increases in average height have been associated with improvements in cognitive performance, labour productivity and life expectancy.

3 Height as a measure of wellbeing

As the shaded panel of Figure 1 (*Data sources*) indicates, anthropometric historians have obtained data from a wide range of sources, but the majority of the available data refer to males, and have been derived from populations who were subject to some form of surveillance or control, such as schoolchildren, prisoners and convicts, military recruits, indentured servants and slaves. Other information has been obtained from skeletal remains, passport holders and selected groups of workers, such as postal workers. The nature of these sources has prompted a number of questions about selection biases and representativeness. These issues will be discussed in more detail in Sections 4–6 below.

Although most of the individual variation in stature can be attributed to genetics, some of the earliest investigations highlighted the existence of systematic differences between the heights of different social groups (Villermé 1829, 384–5). As the panel entitled *Social, economic and environmental determinants* shows, these systematic differences are related to a large number of factors – including income, social status, access to food, personal hygiene, sanitation, air quality, family size, labour organisation and cultural values – that reflect the economic, social and environmental conditions that may prevent people from achieving their genetic height potential.

Figure 1:
Height as a measure of the standard of living



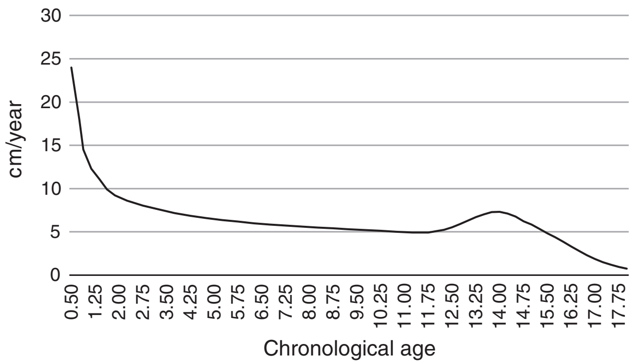
Source: Steckel 1995, 1908, and text.

The next panel in Figure 1 illustrates the *proximate determinants* through which these social, economic and environmental determinants exercise their effects. The key proximate determinants are diet, disease and the energy required for physical maintenance and activity. These factors interact with each other in different ways. As Fogel and others have argued, height is a measure of *net* nutrition. It reflects the balance between the nutrients that a person consumes and the demands made by factors such as the "level of physical activity … climate … and … exposure to various diseases" (Fogel 1986, 446–7). The relationship between diet and disease is particularly important. As Eveleth and Tanner (1990, 191–2) explained, exposure to disease can lead to a loss of appetite, while also increasing the body's need for extra nutrition. Moreover, diarrhoeal infections can prevent the body from retaining the nutrients that are consumed.

The *Functional consequences* panel of Figure 1 highlights some of the functional consequences associated with improvements in stature. These include lower morbidity and mortality, improved cognitive performance and greater work capacity, as well as the prospect of greater happiness and social mobility. These relationships have underpinned the theory of "technophysio evolution", which states that improvements in the height and wellbeing of one generation provide the basis for improvements in the height and wellbeing of subsequent generations (Floud et al. 2011, 1–40).

The theory of technophysio evolution is also illustrated by the arrows linking the different panels. As we have already noted, variations in the average heights of different sections of a population can be linked to systematic differences in a range of social, economic and environmental determinants (*Relationship a*). These factors influence height through their impact on the proximate determinants of diet, disease and energy requirements (*Relationship b*). Differences in height can also be associated with a range of functional consequences (*Relationship c*). These

Figure 2:
Boys' velocity standards



Source: Tanner et al. 1966, 630–631.

functional consequences then exert their own influence on the social, economic and environmental factors that shape the height of the following generation (*Relationship d*).

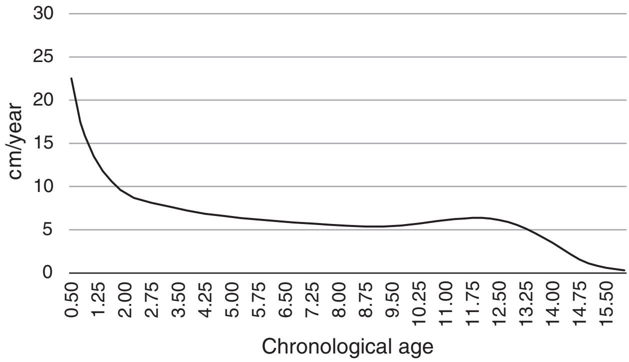
4 Critical periods

Although the study of anthropometric history is ultimately concerned with groups, the field is based on an understanding of the basic pattern of individual growth. As Figures 2 and 3 demonstrate, individuals grow rapidly during infancy and early childhood, and more slowly between early childhood and adolescence. The rate of growth accelerates at adolescence before slowing down. In countries such as modern Britain, the majority of individuals achieve their mature height between the ages of 16 (in the case of girls) and 18 (boys). However, in the past, it seems likely that the onset of adolescence was often delayed, and that individuals continued to grow at higher ages (see, e.g., Tanner 1990, 157–62; Beekink and Kok 2017).[3]

Height and weight have often been used to monitor the health of both individuals and populations. At the individual level, auxologists use height data to identify children who are pathologically short, and they use growth curves to establish whether children are failing to grow at their expected rates (see, e.g., Tanner 1990, 178–9). However, differences in the average heights of different groups of people

[3] Although most anthropometric historians have tended to assume that the onset of the adolescent growth spurt was delayed for historical populations, Gao and Schneider (2020) have recently suggested that it may not have occurred at all. If confirmed, this conclusion could have major implications for our understanding of the pattern of human growth in the past and among deprived populations today.

Figure 3:
Girls' velocity standards



Source: Tanner et al. 1966, 630–631.

can also provide an essential guide to differences in the conditions these groups experienced in childhood. As Eveleth and Tanner (1976, 1; 1990, 1) explained:

> A child's growth rate reflects, perhaps better than any other single index, his [*sic.*] state of health and nutrition, and often indeed his [*sic.*] psychological situation also. Similarly, the average values of children's heights and weights reflect accurately the state of a nation's public health and the average nutritional status of its citizens, when appropriate allowance is made for differences, if any, in genetic potential.... Thus a well-designed growth study is a powerful tool with which to monitor the health of a population, or to identify sub-groups ... whose share in economic and social benefits is less than it might be.

It has often been argued that children's heights are most susceptible to the influence of adverse conditions during the years in which they should be growing most rapidly. Tanner (1990, 131) argued that "in many populations, the period when the child is most at risk from malnutrition, often combined with infection, is six months to three years". Eveleth and Tanner (1990, 194) identified the period from birth to five years as most critical, while also arguing that "a second period when the child may be especially sensitive to the influence of undernutrition is adolescence" (*ibid.*, p. 196). This conclusion has recently been echoed by Depauw and Oxley (2019, 925) in their analysis of changes in the heights of Belgian prisoners during the 19th century. Although they acknowledged that "it is frequently assumed that conditions at birth" had the greatest effect on adult stature, they argued that it was the years between the ages of 11 and 18 that mattered most.[4]

These arguments have important implications for the ways in which we use height and other anthropometric data to measure the health and wellbeing of past societies. If the factors that influence growth only exert a substantial effect during "critical periods", this might limit the extent to which we can use such data to represent social conditions more generally. However, such a conclusion may be premature. For anthropometric historians, two important questions arise: first, to what extent are rates of growth influenced by environmental and nutritional conditions throughout a person's growing years; and, second, what is the relationship between specific periods of infection or nutrition and a person's final (or mature) adult stature?

Unfortunately, we have relatively few data showing concurrent changes in the heights of children at different ages. Harris (1993, 364) examined changes in the heights of children in different parts of the UK during and after the First World War, and concluded that "the differences between the heights of children who were born in 1914 and the heights of children born in 1918 may have owed at least as much to conditions around the time of measurement as … to conditions at the time of birth"; although he also acknowledged that "it is difficult to reach any firm conclusions on the basis of the evidence … currently available". However, the most famous demonstration of this effect was provided by Howe and Schiller (1952) and Tanner (1962, 121–3; 1990, 129–30) in their accounts of changes in the heights of children attending schools in Stuttgart between 1911 and 1953. Tanner found that "there was a uniform increase at all ages from 1920 to 1940, but in both world wars the height dropped as … food intakes … became restricted" (Tanner 1990, 129–30).[5]

The second key question concerns the relationship between childhood experiences and adult stature. Eveleth and Tanner (1990, 195) argued that "the question of whether undernutrition in the first one or two years of life necessarily leads to an adult deficit in body size has been discussed frequently and inconclusively", but their overall conclusion was that "much depends on the circumstances obtaining when the severe episode of undernutrition is over" (see also Tanner 1990, 135). In a very influential historical study, Steckel (1986, 724–5) found that the relative values of the heights of slave children increased very rapidly from the age of 10 onwards. Prentice et al. (2013, 911) have recently argued that "substantial height catch-up

[4] This issue has also been examined by Kopczyński and Rodak (2020) in their study of the impact of the Polish heights between 1918 and 1939. They argued that the deprivations of the First World War had their greatest effects on the adult heights of children who experienced the war as adolescents. This was because these children had less opportunity to benefit from catch-up growth during the more favourable conditions of the 1920s.

[5] Schneider and Ogasawara (2018, p. 65) have recently studied the impact of the disease environment on the growth of children in Japan between 1917 and 1939. They concluded that "the secular increase in height in interwar Japan was more strongly influenced by cumulative responses to the health environment at all ages across development rather than being simply the outcome of improving health conditions in early life".

occurs between 24 mo[nths] and midchildhood and again between midchildhood and adulthood, even in the absence of any interventions". They also cited evidence from their own study of growth patterns in rural Gambia, which showed that "an extended pubertal growth phase allows very considerable height recovery, especially in girls during adolescence" (see also Wells 2016, 373; NCD Risk Factor Collaboration 2020).

One of the main issues raised by this literature is the importance of catch-up growth. If children are exposed to short periods of malnutrition or disease, they may experience a cessation in growth, but they are likely to return to what Tanner (1990, 165) described as a "predetermined growth curve" if conditions improve. This suggests that short-term periods of illness or malnutrition are unlikely to have an effect on adult or mature height, and that adult height will be affected only if the adverse conditions continue. However, catch-up growth may not be "cost-neutral" (Wells 2016, 224). As Singhal (2017, 236–7) has suggested, "there is now compelling evidence that 'accelerated' or too fast growth during critical or sensitive windows in early life has detrimental effects on long-term health, and particularly the risk of obesity and cardiovascular disease"(see also Leon et al. 1996).

These findings have different – and possibly conflicting – implications for our understanding of the relationship between height and wellbeing. If children experience short-term periods of illness or malnutrition, their rate of growth may be temporarily reduced, but they are also likely to return to their predetermined growth curves if conditions improve. However, even if a catch-up process does occur, these periods of disease or malnutrition may be associated with longer-term health deficits. On the other hand, if adverse conditions continue, the children are likely to have a shorter adult height. This suggests that adult height can be a sensitive barometer of the cumulative impact of adverse circumstances throughout the growing years, even if the impact of short-term fluctuations can sometimes be masked (see also Harris 2008a).

5 Children and adults

A second problem is presented by the question already highlighted by Richard Steckel (1992, 284): namely, that average heights reflect the impact of living standards on health during the period from conception to maturity "as opposed to one's status with respect to commodities more generally". There are a number of possible responses to this question. One argument is that the height of a child might capture the impact of some aspects of adult health, given that the height of one generation may be determined, at least in part, by previous standards of maternal nutrition. In other words, child and adult heights may reflect the impact of the maternal environment, as they capture some aspect of the welfare of adult women during pregnancy. However, this point would probably be insufficient to counter the view that heights still only capture aspects of human welfare during particular periods of life.

Steckel's observation also raises an important question about the particular demands placed on people during childhood and the allocation of resources between adults and children. For example, an increase in the demand for child labour might exacerbate the burdens placed on children, but it could also enhance their claims to a larger share of household resources (see, e.g., Floud et al. 2011, 138). Equally, if parents decide to devote more resources to their children, this may enhance the standard of child welfare, without necessarily reflecting a change in the welfare of society as a whole.

Although the majority of anthropometric historians have tended to assume that average heights reflect the experiences of current generations, others have argued that the secular increase in height is an "intergenerational phenomenon, with the offspring of each generation becoming taller than the previous generation, but with the increase in height per generation being tightly controlled" (Cole 2003, 166). However, other evidence suggests that average heights can increase very rapidly when living conditions improve. Gruenwald et al. (1967, 1028) estimated that the average height of six-year-old children in Japan increased by between 3.2 and 3.7 cm from 1945/6 to 1957/8, while Tanner et al. (1982) noted that the average height of Japanese adolescents increased by between 7.9 and 9.7 cm from 1957 to 1977. Ling and King (1987, 187) reported that the average height of Chinese adolescents in Hong Kong increased by between 4.2 cm and 6.7 cm from the early 1960s to the early 1980s (see also Harris 2001, 1432).

Although these studies suggest that average heights can increase very rapidly when circumstances change, this does not mean that the problem of intergenerational effects can be discounted entirely. As Wells (2016, 104) has explained, several studies have shown that the children and grandchildren of mothers who were exposed to undernutrition in their own lifetimes seem to face a higher risk of obesity. This suggests that even if these children grow to be taller than their forebears, they may still be exposed to medical risks that originated in the experiences of earlier generations (see also *ibid.*, 313).

The use of height as a measure of wellbeing may also be complicated by the relationship between height and mortality. As we shall see, it has often been noted that shorter people are more likely to die at any given age than taller people, leading some observers to point out that this pattern will necessarily inflate the average height of survivors (Alter 2004, 545). Friedman (1982, 502) argued that "the mean height of adult males in Trinidad (aged 26–45) who died [during the early-nineteenth century] … was 0.61 inches [1.5 cm] less than for those who survived"; with the result that the "mean height of the survivors was 0.03 inches [0.01 cm] taller than that of the initial population". However, these results referred to the impact of deaths that occurred after mature height had been achieved. If mortality occurs at ages when children are still growing, any selection effect is likely to be outweighed by the impact of adverse conditions on the growth of those who survive (Hatton 2011, 2014; see also Prentice et al. 2013, 914).

Regardless of one's views on these questions, there are clearly limits to the extent to which heights provide direct evidence of standards of living after mature height has been attained. Some historians have attempted to compensate for this limitation by using evidence of adult weights. Weight is a useful indicator in this context because it continues to change after mature height has been achieved (Horrell et al. 2009, 96). However, it is also more ambiguous, partly because of its subjective element, and partly because being underweight and being overweight can both be indicators of ill-being (Harris 2014, 128).

It is also essential to consider the question of gender. As Osmani and Sen (2003) explained, the study of female heights is important not only in its own right (as an indicator of female wellbeing), but also because of the potential impact of women's nutritional status on the health of their offspring. However, there are relatively few studies of changes in female stature, and many of the studies that do exist have focused on the heights of convicts and prisoners. While these sources have also been used to examine changes in male height, the heights of male and female convicts may not be entirely comparable because of differences in patterns of male and female criminality, as well as in policing and sentencing procedures (Harris 2008b, 165).

Although researchers are beginning to examine a wider range of sources, a consistent picture has yet to emerge. Koepke et al. (2018) have recently compared changes in the heights of men and women in Switzerland using data obtained from passport applicants, convicts, female auxiliaries and mothers attending the Basel maternity hospital. They argued that a secular increase in female heights began a generation earlier than the onset of a similar increase in male heights. However, the evidence from other studies is more mixed. Ezzati et al. (2016, 7) argued that "at the turn of the twentieth century, men seem to have had a relative advantage over women in under-nourished compared to better-nourished societies"; whereas Perkins et al. (2016, 152) concluded that "sexual dimorphism in stature is more pronounced when undernutrition and childhood disease are mitigated".

One of the main challenges involved in comparing male and female heights is the need to take into account social and economic factors on the one hand, and physiological differences on the other. Tanner (1962, 127–8) argued that girls are more resistant to the effect of adverse circumstances because they show greater powers of canalisation, or homeorrhesis. Although the concept of canalisation has been questioned, it might help to explain why other studies have also found that the heights of past generations of girls appeared to compare more favourably with the heights of current generations than was the case for their male counterparts (see also Harris 2009).[6]

[6] The concept of canalisation was introduced by the geneticist and developmental biologist, C. H. Waddington, who argued that "developmental reactions, *as they occur in organisms submitted to natural selection*, are in general canalized. That is to say, they are adjusted so as to bring about one definite end-result regardless of minor variations in conditions during the course of the reaction" (Waddington 1942, 563). This implies that children should grow at the same rate, relative to the rest of the population, throughout their growing years. Although it is widely used, it has not been accepted

6 Truncated samples and selected populations

Although many authors would agree that, in principle, height data provide an important index of wellbeing, there have been a number of significant disagreements over the main trends in height in particular contexts. There is, for example, a longstanding argument about the correct representation of the main trends in height in Britain during the late 18th and early 19th centuries, and there have also been important debates about the representativeness of the height data used to measure changes in wellbeing in both Britain and the United States during the second and third quarters of the 19th century.

In their original study, Floud et al. (1990) argued that there was a slow and irregular improvement in the average heights of successive birth cohorts of British males between the mid-18th century and the end of the first quarter of the 19th century, but this interpretation of the data has been challenged by Komlos (1993a,b) and other authors (see, e.g., Cinnirella 2008; Komlos and Küchenhoff 2012; Meredith and Oxley 2014).[7] Although some of the debate has focused on Floud et al.'s analysis of the heights of children who joined the Marine Society between *circa* 1770 and 1859, most of the controversy has revolved around their analysis of the heights of military recruits. Critics have challenged Floud et al.'s decision to pool data from different military sources, and they have also questioned the estimation procedures the authors used to extrapolate from the heights of military recruits to the (male) population as a whole.[8]

In addition to raising these statistical questions, different authors have attempted to support their interpretations of the data by comparing them with indicators that can be regarded as proximate determinants of welfare, such as real wages or food availability; or with other indicators, such as mortality, which could be regarded either as alternatives to stature or as indirect determinants of it. Komlos (1993b, 365) sought to reinforce his initial critique of Floud et al. by highlighting evidence suggesting that average food supplies had deteriorated. However, this argument has also been challenged. Indeed, in their latest contribution, Harris et al. (2015)

---

uncritically (see, e.g., Hermanussen et al. 2001). These authors also found, *contra* Tanner, that "growth in boys appeared slightly more canalised than in girls" (pp. 163, 166).

7 Galofré-Vilà et al. (2018, 87–8) also found evidence of a small decline in average heights during the course of the 18th century in their study of skeletal remains. However, they were careful to point out that their conclusions were based on very small sample sizes, and may have been distorted by changes in the composition of the population from which their samples were drawn.

8 As Floud et al. (1990, 118) explained, the British Army imposed a number of different minimum height standards to control the flow of volunteers at different times. They advocated the use of Reduced Sample Maximum Likelihood Estimation to compensate for the effects of this variation when estimating the average height of the underlying population. Komlos (1993a,b) claimed that this procedure was flawed, and led to erroneous estimates. He provided a fuller discussion of the use of different methods for correcting for the effects in truncation in 2004 (Komlos 2004). For further responses, see Floud et al. (1993a), Floud et al. (1993b) and Floud et al. (2011, 66–7, 136–8).

presented evidence that average food supplies rose overall between 1750 and 1850 (see also Harris 2016).

The point of reviewing these debates is not to revisit the technical aspects for their own sake, but to highlight their broader implications for the study of anthropometric history. If we regard stature as an entirely separate measure of the standard of living – which may or may not be implied by the use of such terms as the "biological standard of living" – then there is perhaps less need for it to be correlated with other indicators. However, if we regard stature as an alternative measure of "the" standard of living, then it is more important to show how trends in height may or may not be related to more conventional indicators, where evidence for such indicators exists.

The interpretation of the height data drawn from volunteer samples, including both military recruits and prisoners, has also been called into question by Howard Bodenhorn and his co-authors (Bodenhorn et al. 2013, 2017). Anthropometric historians have tended to organise the heights of mature adults by birth cohorts, either because they believe that the factors that influence height exert their strongest influence during infancy and early childhood, or because they believe that the factors that shape average heights operate from conception to maturity (see also Section 4 above). Bodenhorn et al. argued that we also need to consider the ways in which economic fluctuations might influence the propensity of men and women of different heights either to volunteer for the army or commit crimes. This led them to conclude that the declines in average height observed by students of both antebellum America and mid-19th-century Britain were both artefacts of selection bias.

Although Bodenhorn et al. were undoubtedly correct to highlight the importance of selection bias, it also important to note at least two caveats. First, it is necessary to acknowledge that their concerns were not new – Floud et al. (1990, 115–8) discussed the impact of labour market fluctuations on the pool of army volunteers at some length – and their efforts to rework both the UK data and the US data have themselves been challenged (Zimran 2019; Komlos 2019; Komlos and A'Hearn 2016, 2019). Second, Bodenhorn et al.'s conclusions also raise questions about how they see the relationship between height and other welfare indicators. Their work appeared to pose a direct challenge to the role played by anthropometric studies in supporting "pessimistic" accounts of the impact of industrialisation on the standard of living (see, e.g., Bodenhorn et al. 2013, 6–8; *ibid.* 2015, 8–10; *ibid.* 2017, 174–5). However, in the conclusion to their 2017 paper, they acknowledged that "mortality rates remained stubbornly high through the early decades of industrialisation … and in some cases actually increased, as cities became larger and less healthy.… Real wages rarely fell, but there is reason to doubt that feeble nominal wage growth protected the lowest strata from the consequences of food-price shocks" (Bodenhorn et al. 2017, 202). If this is correct, it is legitimate to ask why they believed that these factors should not have also been reflected in the anthropometric record.

7 Dietary influences on stature

As the anatomist, David Sinclair (1978, 140) once explained, "the best way of growing tall and heavy is to have tall and heavy parents". This insight has been reinforced by more recent studies, which have suggested that approximately 80 per cent of the variation in height between individuals of European descent (Visscher 2008) or living in modern western societies (Silventoinen 2003) is genetic in origin, with the remainder being attributed to environmental conditions, of which the most important are nutrition and infection. As Eveleth and Tanner (1976, 246) explained:

> An ill child is a poorly-nourished child, although the extent of slowing down depends on a number of factors. Poorly-nourished children are more susceptible to and more severely-affected by infection than well-nourished children.... Infection in turn lowers the nutritional intake of the child and the vicious spiral continues.

Anthropometric historians have often considered the question of whether some foods may be more nutritious, and therefore growth-promoting, than others. As Floud et al. (1990, 298) acknowledged, it has often been assumed that at the end of the 18th century, the Scots and Irish were poorer than the English or Welsh, and yet their children also grew to be taller. They speculated that this pattern may have reflected the extent to which "potatoes and milk, and perhaps also the oatmeal of the Scots, seem to have provided a healthy and balanced diet for those who did not have to fight off urban disease".

The question of whether either the Scots or the Irish owed their relative tallness to the particular characteristics of their staple diets has also been examined by Mokyr and Ó Gráda (1996, 163–4). They suggested that the specific claims made on behalf of the nutritional value of particular foods meant that anthropometric historians needed to "tone down" claims that height provided an alternative, or even superior, guide to the "standard of living". However, they also argued that some of the height advantage enjoyed by Irish recruits may have been caused by selection effects, thus clouding the issue further.[9]

The importance of different kinds of foodstuffs has also been discussed elsewhere. Floud et al. (2011, 162) noted that people who were over-dependent on cereal-based diets needed to consume more food in order to obtain the same nutritional benefits. Dasgupta and Ray (1990, 215–6) argued that individuals who consumed high-fibre diets under pre-industrial conditions needed to increase their total consumption by around 35 per cent in order to derive the same nutritional benefits from their diets as people living under more favourable circumstances.

[9] Grasgruber et al. (2016, 164–5) have recently examined the relationship between a wide range of variables and differences in height in 105 countries. They found that potato consumption was correlated positively with height ($r = 0.68$; $p < 0.001$), but were unable to explain why. They noted "the significance of potatoes ... is unexpected because of the poor quality of potato proteins, their low consumption rate and a very low 'nutrient density'".

Other authors have also highlighted the possible importance of meat and, in particular, dairy products. Jörg Baten and his co-authors have argued that variations in milk consumption help to explain regional variations in stature in 19th-century Bavaria, Prussia and France (Baten 2009; Baten and Murray 2000), and in other parts of Europe during the time of the Roman Empire (Koepke and Baten 2005; 2008). These findings have been reinforced by contemporary surveys that have also highlighted a positive association between height and milk consumption (Perkins et al. 2016, 153–4; Wells 2016, 302, 362). However, Baten has also suggested that the impact of milk consumption on variations in stature in the latter case may have been mediated by the specific economic situation of the Roman Empire and its high population densities (Meinzer et al. 2019, 238–9).

Although the significance of this issue should not be exaggerated, it does pose a challenge to the idea that height provides a *general* guide to the “standard of living”. As Mokyr and Ó Gráda (1996, 163–4) suggested, if one population was shorter than another because its staple diet was based on wheat rather than potatoes, one would not automatically conclude that its standard of living was poorer. It is, perhaps, for this reason that both Mokyr and Ó Gráda (1996, 163–4) and Baten and Blum (2014, 568) argued that height data should be used to complement conventional welfare indicators, rather than being cited as replacements for them.

8 Height and disease

As we have already seen, one of the key arguments in favour of using anthropometric approaches is the claim that height is a measure of *net* nutrition – i.e., that in addition to capturing the effect of dietary inputs, it also reflects the impact of the external demands placed on the body by the environment in which the person lives. These demands might include not only those created by the need for calories for work and physical maintenance, but also the effects of disease. This argument is particularly important in the context of debates about the impact of industrialisation on the standard of living. Proponents of the anthropometric approach have contended that height captures not only the effects of changes in real wages, but also the impact of urbanisation on the disease environment (Floud 1984a).

Auxologists and anthropometric historians have identified a number of different ways in which stature may be affected by disease. First, illness can both suppress appetite and increase the number of calories required to ward off infection. It can also interfere with the absorption of essential nutrients. Children who are subjected to repeated bouts of diarrhoea are likely to grow more slowly if they lack access to nutritional supplementation. As Floud et al. (1990, 245) explained:

> Infection affects nutritional status by a number of different mechanisms. These include loss of appetite, energy lost as heat during fever and loss of other nutrients in sweat, vomiting, decreased absorption of nutrients, protein catabolism, and reduced food intake resulting from cultural

> factors. Malabsorption can also result from infestation with intestinal parasites, but the most common and severe cause is diarrhoea, which causes food to pass through the intestine too quickly to be absorbed.

Historians have also highlighted the impact on stature of malaria (Hong 2007) and of hookworm infestation (Coelho and McGuire 1999), while other scholars have pointed to the impact on stature of atmospheric pollution. Sharpe (2012) identified a number of different diseases that can cause stunting, including measles, whooping cough, bronchopneumonia, diarrhoea and rickets. The last of these diseases was found to be associated not only with poor diet, but also with reduced exposure to sunlight due to the dense fog enveloping Britain's cities. Bailey et al. (2018) have recently built on this foundation in their study of the impact of air pollution on the heights of men who completed their army service in Britain between 1914 and the early 1920s. They concluded that coal smoke reduced adult height by almost half an inch (1.2 cm), and that efforts to clean up Britain's air accounted for approximately 25 per cent of the gain in average male stature between the birth cohorts of the 1890s and the 1980s (*ibid.*, 33).

Although most anthropometric historians would accept the premise that height reflects the impact of a wide range of influences, some observers have suggested that this assumption is less of a strength than a weakness. Crafts (1992, 428) complained that it is "unclear how to incorporate height in a welfare index" because "no way of measuring the exchange rate of height for real income has yet been devised". It might, however, be argued that in some ways at least, this misses the point. It is precisely because height captures a wider range of influences in addition to real income that its proponents regard it as a superior measure, even if this also means that it is more difficult to relate changes or variations in height to any single cause (see also Floud et al. 2011, 12–13; Steckel 2008, 136; Steckel 2016, 38).

9 Height and ethnicity

As we have already seen, anthropometric historians have devoted a great deal of effort to the challenge of identifying trends in average height within individual countries. However, some of the earliest (and most foundational) papers also sought to draw comparisons between countries. After comparing data from 20 countries in different parts of the world, Steckel (1983) concluded that international differences in mean stature were particularly sensitive to variations in per capita income and the degree of income inequality. Floud (1984b, 23; see also *ibid.* 1994, 23) examined trends in the average heights of men in 11 European countries between the mid-19th and late-20th centuries, and concluded that "western European heights have responded entirely systematically, over the past hundred years, to changes in income and disease, just as heights in the modern world respond to similar differences between countries".

In recent years, it has become possible to study the impact of environmental and genetic factors on height in much more detail, and this has led some authors to suggest that genetic factors may play a larger role in influencing cross-national variations in adult stature. Brian A'Hearn (2016) has recently argued that the distribution of heights in southern Europe correlates quite closely with variations in real wages before 1850, but less so thereafter. He suggested that this pattern could be explained in part by changes in work intensity (i.e., workers were compensated for reductions in hourly wages by working longer), and in part by changes in the disease environment (including the eradication of malaria). However, when looking at variations in the heights of adult males born during the 1980s, he argued that genetic differences also played an important role.

As A'Hearn acknowledged, this analysis was subject to a number of limitations. In order to measure the impact of variations in the disease environment, he examined differences in life expectancy. It is, however, likely, that much of the variation in life expectancy reflected the impact of cross-national differences in mortality from non-communicable diseases at higher ages. Other writers have preferred to study the correlation between height and child mortality rates (see, e.g., Grasgruber et al. 2014; Grasgruber et al. 2016). These measures also have their limitations, but the correlation between child mortality rates and height in A'Hearn's sample was greater than the correlation between height and life expectancy.[10]

However, other writers have also concluded that genetic factors have played an increasingly important part in explaining contemporary height variations. Grasgruber et al. (2014) explored the impact of a range of factors that have contributed to variations in adult male stature in 45 European territories and their "offshoots" in Australia, New Zealand and the United States. Although they concluded that height was quite closely correlated with a number of economic and socio-economic factors, including nutritional quality, GDP per capita, health expenditure, child mortality and income inequality, they also detected relationships with a number of genetic variables, including the distribution of Y-haplogroup I-M170, the combined frequencies of Y-haplogroup I-M170 and R1B-U106, and the phenotypic distributions of lactose tolerance. They also reported similar, though not identical, conclusions after extending their analysis to include countries in Africa, Asia and Oceania (Grasgruber et al. 2016).

While these findings suggest that genetic factors may play a larger role in determining international variations in stature than was previously supposed, economic, environmental and nutritional factors also continue to be important. Even if the

[10] We have data on adult male stature, life expectancy at birth and children's mortality rates for 27 of the countries that A'Hearn examined. The correlation between height and life expectancy was 0.487 ($p = 0.01$), and the correlation between height and child mortality was −0.569 ($p = 0.002$). For data on height and life expectancy, see A'Hearn (2016, 769–70); for data on child mortality rates, see https://datacatalog.worldbank.org/dataset/world-development-indicators. The child mortality rate is the average for the period 1980–85, with the exceptions of the Czech Republic (1981–5), Serbia (1984–5) and Slovenia (1981–5). A'Hearn's original study also included results for "Med-Yugoslavia".

impact of these factors on stature has declined over time, it has not been eliminated altogether. As Grasgruber et al. (2016, 194) have argued, "the factors leading to the increase in the average height intertwine with public policies that improve the overall quality of life". Thus, even if the importance of genetic factors has increased, the evidence suggests that height continues to be an important tool for monitoring the success of public policies. Baten and Blum (2012) reviewed the historical evidence on changes in average male stature in 156 countries between *circa* 1810 and 1989. After "taking into account … protein availability, disease environment, lactose intolerance and food preferences", they concluded that "the height impact of 'race' seems rather small" (p. S69; see also NCD Risk Factor Collaboration 2020, 1520).

10 Height, wages and mortality

The majority of this paper has been concerned with the use of height data as summary reflections of the aggregated impact of factors such as real income, food consumption and disease – all of which have traditionally been associated with the "standard of living". However, an important part of the argument in favour of using height data as indicators of wellbeing is that they also capture aspects of human development that have functional consequences for other welfare measures, including both labour productivity and longevity.

An early illustration of the relationship between height and both longevity and productivity was provided by Friedman's study of the heights of Trinidadian slaves. As we have already seen, he showed that slaves who survived the initial registration period were 0.61 inches (1.5 cm) taller than non-survivors, and that craftsmen were 0.5 inches (1.25 cm) taller than fieldhands (Friedman 1982, 488–9). Similar results were reported by Margo and Steckel (1982) and Costa (1993). These findings have also been echoed in studies of more recent societies. Waaler (1984) showed that shorter people are more likely to die at younger ages, and Schultz (2002) demonstrated that taller people enjoy higher wages. Deaton and Arora (2009) concluded that height is positively correlated with both income and education, as well as with happiness and wellbeing.

As we have already seen, height is an important marker of the effects of environmental and nutritional conditions on childhood growth. However, in modern western societies, approximately 80 per cent of the individual variation in height is likely to be due to genetic effects (see Section 7 above). It is, therefore, important to ask what dimensions of stature are being captured by the association between height and other indicators. This question is particularly important when considering historical data. It seems reasonable to assume that environmental factors accounted for more of the variation in individual stature in the past, when a greater proportion of the population was subjected to conditions that were likely to restrict their growth. If this was the case, we might expect the association between height and other measures to have been stronger if this association was caused by the impact of early life conditions on adult stature.

One way of approaching this issue is to compare the association between height and mortality in historical and contemporary populations. Both Costa (1993) and Alter (2004) have compared the association between height and mortality within historical populations with Waaler's (1984) analysis of the relationship between height and mortality in mid-20th-century Norway. In all three cases, mortality declined as height increased, up to a height of approximately 73 inches (183 cm), at which point the relationship was reversed. If the association between height and mortality had been caused primarily by genetic factors, the curve should have shifted to the right as average heights increased. The fact that it appears to have remained constant suggests that environmental factors also played a role (see also Harris 1997).

This is a reassuring conclusion for those who believe that improvements in the environmental and nutritional conditions that are associated with adult stature will also lead to reductions in mortality. However, it is important to recognise that the association between height and mortality is also linked to specific causes of death, such as cancer, coronary heart disease and chronic obstructive pulmonary disease; and that there are other causes of death for which height and mortality are correlated positively (Floud et al. 2014, xxxiii; Perkins et al. 2016, 155–7). As the cause structure of mortality changes, we might expect to see further changes in the relationship between height and mortality.

11 The social value of height

The preceding sections have focused on the extent to which height can be correlated with other factors associated with the measurement of wellbeing. This has led us to examine factors that influence height and rates of growth, as well as other indicators that may be affected by them. However, it is worth asking whether height also possesses a more "subjective" value as either an index or a dimension of wellbeing. If height is correlated with measures of happiness or "subjective wellbeing", is this because it is also associated with other correlates, or because it is valued in its own right?

In an intriguing paper, Deaton and Arora (2009) compared the heights and weights of more than 450,000 men and women with their positions on a "self-anchoring sliding scale". Respondents were asked to rate their quality of life on a sliding scale of 1–10, with one representing "the worst possible life", and 10 representing the best. They found that "men who are above average height … report that they are a little more than one-seventh of a step … above men who are below average height". Moreover, they reported that women whose height was above average were a little less than one-tenth of a step higher on the scale than women whose height was below the average. However, the authors also found that almost all of the difference disappeared after controlling for the effects of income and education. In other words, they concluded that if taller people were happier, it was not because they were taller, but because they were wealthier and better educated.

There are, however, some indications that height does have a value of its own. As Floud et al. (2011, 13) observed, there is some evidence for this assumption in the "immense effort and expenditure undertaken by some parents of particularly short children to lengthen their children by persuading surgeons to break their leg bones and then … stretch their legs as they heal". There is also evidence to suggest that height is often a quality we value in others. Several studies have shown that taller men are more likely to marry at younger ages, and are more likely to ever marry, although the reasons why this is the case remain unclear. Murray (2000, 518) argued that while women may prefer taller men on cultural or aesthetic grounds, it is also possible that women recognise that such men are likely to be healthier or have better earnings prospects (see also Weitzman and Conley 2014; Yamamura and Tsutsui 2017). Sohn (2015, 111) examined the "trade-off" between height and income in Indonesia, and found that while much of the association between male height and marriage was related to income, "there are still other attributes that women look for in male height, and this indicates that there are marriage market penalties for short men".

Although much of the available evidence suggests that height does possess a social value, it does not necessarily follow that societies will always behave in ways that seek to maximise stature. As we have already suggested, height is a measure of net nutritional status "from conception to maturity", and may therefore reflect, at least in part, the extent to which parents are either able or willing to behave in ways that maximise their children's nutritional status. Variations in stature may also reflect the extent to which societies "trade off" some aspects of wellbeing against others. Eltis (1982, 474–5) argued that many historical populations may have engaged in "nutritional satisficing" or "aim[ing] for a nutritional target lower than maximum". He argued that this was because they valued "self-employment" over the achievement of more material goals. He also suggested that in many cases, improvements in average stature only occurred when labour was either partially or entirely coerced.

12 Conclusions

Eveleth and Tanner (1976, 1; see also *ibid.* 1990, 1) argued that "a well-designed growth study" was not only "a powerful tool with which to monitor the health of a population", but also a tool "to pinpoint sub-groups … whose share in economic and social benefits is less than it might be". The link between height, health and "economic and social benefits" was the basis for Tanner's subsequent description of height as a "mirror of the condition of society" (Tanner 1987). To what extent does this claim continue to apply to the study of historical populations?

This paper began by examining the development of anthropometric history, and of research into the impact of economic, social and environmental factors on variations in human stature. It then explored some of the issues associated with the identification of "critical periods", and efforts to extrapolate from the analysis of male heights to the understanding of differences in welfare between different genders and age groups. It also provided a brief account of the impact on height of both diet and disease, before looking at some recent debates about the impact on stature of "race" and ethnicity. It concluded by examining the relationship between height, wages and mortality, and between height and "subjective wellbeing".

In general, this brief survey suggests that height has retained its capacity to shed light on the welfare of past populations, with some possible caveats. One problem concerns the question of "catch-up" growth among people who experience adverse conditions early in life. There is increasing evidence that even though children – especially at the youngest ages – are able to "catch up" if their environmental or nutritional circumstances improve, this may still come at the cost of future health problems. If this is the case, then the ability of these children to return to their "predetermined growth curves" may still mask significant health deficits. If, on the other hand, the children's circumstances do not improve, their final, or mature, height is likely to be lower, and will thus reflect the cumulative impact of the adverse circumstances they experienced throughout their growing years.

The value of height as a measure of welfare is also affected by the role played by particular foodstuffs. Various authors have argued that height is especially sensitive to the consumption of oats, potatoes, milk or other sources of animal protein. Some of these findings may, of course, have been affected by selection issues. However, insofar as growth is affected by the consumption of particular foodstuffs, it is also important to consider the extent to which access to the most nutritious – and growth-promoting – foods is itself related to broader conditions.

Another controversial issue concerns the relationship between height and ethnicity. It has recently been suggested that ethnicity may play an increasingly important role in determining international variations in stature, although this may depend in part on the particular nature of the variables that have been used as proxies for economic and health conditions. However, it still seems likely that international variations in stature would have been more closely associated with variations in economic and environmental circumstances in the past, when the impact of these factors on individual height differences may also have been greater.

The use of height as a measure of wellbeing depends in part on the argument that it can be correlated with a range of causal factors that are broadly associated with the "standard of living", and in part on the claim that it is also correlated with other dimensions of wellbeing that individuals might have "reason to value" (Sen 1999, 14). Two issues that have received particular attention in this context are wages and mortality. However, it is important to remember that the correlations between height and both of these indicators may change over time. Costa (2015, 551) has argued that "returns to height in the United States *circa* the mid-nineteenth century were very low" because they were generally associated with low status occupations, but that these returns have increased over time because of the correlation between height and cognitive performance, and the transition "from a brawn- to a brain-based economy".

The relationship between height and mortality is also, in some senses, time contingent. As we have seen, it seems reasonable to suppose that environmental factors had a greater influence on variations in stature in the past than they do today. If the relationship between stature and mortality reflected the relationship between stature and living conditions, we might expect the relationship between stature and mortality to weaken over time. This effect is compounded by changes in the cause structure of mortality. If height is negatively correlated with diseases that are having a declining impact on mortality, we would expect the relationship between height and overall mortality to change accordingly.

Although increases in height have generally reflected improvements in living standards, this relationship may be complicated in other ways. As we have already seen, height is a cumulative measure of the conditions experienced by children from conception to maturity. As a result, in addition to indicating a household's general living conditions, a child's height may also reflect the way in which resources are distributed within the household. In many contemporary societies, it is still common for girls to receive a smaller share of the available resources than their brothers. However, historians have continued to disagree over the extent to which such a gender divide in the treatment of siblings may have existed in the past (see, e.g., Harris 2008b). This is one reason why the relative paucity of data on female heights remains a significant challenge for anthropometric historians.

The long-term history of anthropometric change may also be affected by changes in household size and structure. Previous studies have shown that height can vary with birth order, and that children who grow up in (numerically) large families have often been shorter than children who grow up in smaller families (Öberg 2017). These findings have led a number of authors to conclude that one of the major causes of the improvement in heights during the 20th century was the decline in fertility – even though not everyone would see the decision to have fewer children as a form of "wellbeing" (Hatton and Martin 2010; see also Hatton and Bray 2010).

This issue also highlights the potential significance of parental choice in determining children's growth. Although parents do not necessarily "purchase" height (Floud et al. 2011, 13), it is generally assumed that they will seek to obtain the living conditions with which height is often correlated. However, as Eltis (1982, 474–5) argued, many adult populations may – no doubt unintentionally – have reduced their children's growth by prioritising leisure over work. Although he did not discuss this issue at great length, he also speculated that this practice only ceased under conditions in which labour was either partially or entirely coerced.

This is not the only perplexing conclusion that the history of human height has sometimes evoked. In general, as material conditions have improved, children have grown taller and heights have increased. However, a number of scholars have also suggested that some of the largest improvements in stature in earlier periods were preceded by significant disasters. Clark (2007, 101) argued that the Black

Death of 1347–51 “raised living standards all across Europe” in the following years. In a similar vein, some anthropometric historians have argued that “the tremendous reduction of the population during the Justinian Plague and the end of the West Roman Empire had a strong impact on health, because the much smaller population [in number] was likely better nourished” (Baten et al. 2019, 394). Although this interpretation has not gone unchallenged (see Galofré-Vilà et al. 2018, 81–2), it nevertheless provides a rather sobering coda to the long-term history of anthropometric change.

Acknowledgements

Previous versions were presented to the Annual Conference of the Economic History Society at Queen’s University, Belfast, on 6 April 2019; and at a conference on “The demographic aspects of human wellbeing” at the Wittgenstein Centre, Vienna, on 11 November 2019. I am grateful to the editors of this journal and to Gregori Galofré-Vilà, Daniel Gallardo Albarrán and three anonymous referees for their helpful comments.

References

A’Hearn, B. 2016. The anthropometric history of the Mediterranean world. In *The Oxford handbook of economics and human biology*, eds J. Komlos and I. Kelly, 765–796. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199389292.013.45

Allen, R. 2009. Engels’ pause: Technical change, capital accumulation and inequality in the British industrial revolution. *Explorations in Economic History* 38(4): 411–447. https://doi.org/10.1016/j.eeh.2009.04.004

Alter, G. 2004. Height, frailty and the standard of living: Modelling the effects of diet and disease on declining mortality and increasing height. *Population Studies* 58(3): 265–279. https://doi.org/10.1080/0032472042000272339

Bailey, R., T. Hatton and K. Inwood 2018. Atmospheric pollution, health and height in late-nineteenth century Britain. *Journal of Economic History* 78(4): 1210–1247. https://doi.org/10.1017/S0022050718000578

Baten, J. 2009. Protein supply and nutritional status in nineteenth-century Bavaria, Prussia and France. *Economics and Human Biology* 7(2): 165–180. https://doi.org/10.1016/j.ehb.2009.02.001

Baten, J. and M. Blum 2012. Growing tall but unequal: New findings and new background evidence on anthropometric welfare in 156 countries, 1810–1989. *Economic History of Developing Regions* 27(Supplement): S66–S85. https://doi.org/10.1080/20780389.2012.657489

Baten, J. and M. Blum 2014. Why are you tall while others are short? Agricultural production and other proximate determinants of global heights. *European Review of Economic History* 18(2): 144–165. https://doi.org/10.1093/ereh/heu003

Baten, J. and J. Murray 2000. Heights of men and women in nineteenth-century Bavaria: Economic, nutritional and disease influences. *Explorations in Economic History* 37(4): 351–369. https://doi.org/10.1006/exeh.2000.0743

Baten, J., R. Steckel, C. S. Larsen and C. Roberts 2019. Multidimensional patterns of European health, work and violence over the past two millennia. In *The backbone of Europe: Health, diet, work and violence over two millennia*, eds R. Steckel, C. S. Larsen, C. Roberts and J. Baten, 381–396. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108379830.015

Beekink, E. and J. Kok 2017. Temporary and lasting effects on childhood deprivation on male stature. *History of the Family* 22(2–3): 196–213. https://doi.org/10.1080/1081602X.2016.1212722

Bodenhorn, H., T. Guinnane and T. Mroz 2013. Problems of sample-selection bias in the historical heights literature: A theoretical and econometric analysis. Center Discussion Papers 148749, Yale University, Economic Growth Center. https://doi.org/10.22004/ag.econ.148749

Bodenhorn, H., T. Guinnane and T. Mroz 2015. Sample-selection biases and the "industrialization puzzle". NBER Working Paper Series no. 21249. https://doi.org/10.3386/w21249

Bodenhorn, H., T. Guinnane and T. Mroz 2017. Sample-selection biases and the industrialization puzzle. *Journal of Economic History* 77(1): 171–207. https://doi.org/10.1017/S0022050717000031

Cinnirella, F. 2008. Optimists or pessimists? A reconsideration of nutritional status in Britain, 1740–1865. *European Review of Economic History* 12(3): 325–354. https://doi.org/10.1017/S136149160800227X

Clark, G. 2005. The condition of the working class in England, 1209–2004. *Journal of Political Economy* 113(6): 1307–1340. https://doi.org/10.1086/498123

Clark, G. 2007. *A farewell to alms: A brief economic history of the world*. Princeton: Princeton University Press.

Coelho, P. and R. McGuire 1999. Biology, disease and economics: An epidemiological history of slavery in the American South. *Journal of Bioeconomics* 1: 151–190. https://doi.org/10.1023/A:1010034830415

Cole, T. 2003. The secular trend in human physical growth: A biological view. *Economics and Human Biology* 1(2): 161–168. https://doi.org/10.1016/S1570-677X(02)00033-3

Costa, D. 1993, Height, weight, wartime stress, and older-age mortality: Evidence from the union army records. *Explorations in Economic History* 30(4): 424–449. https://doi.org/10.1006/exeh.1993.1018

Costa, D. 2015. Health and the economy in the United States from 1750 to the present. *Journal of Economic Literature* 53(3): 503–579. https://doi.org/10.1257/jel.53.3.503

Costa, D. and R. Steckel 1997. Long-term trends in health, welfare and economic growth in the United States. In *Health and welfare during industrialisation*, eds R. Steckel and R. Floud, 47–89. Chicago: University of Chicago Press.

Crafts, N. 1992. Review of R. Floud, K. Wachter and A. Gregory, Height, health and history: Nutritional status in the United Kingdom. *Economic History Review* 45(2): 427–428. https://doi.org/10.1111/j.1468-0289.1992.tb01309.x

Crafts, N. 1997. Dimensions of the "quality of life" during the British industrial revolution. *Economic History Review* 50(4): 617–639. https://doi.org/10.1111/1468-0289.00071

Dasgupta, P. and D. Ray 1990. Adapting to undernourishment: The biological evidence and its implications. In *The political economy of hunger. Vol. 1: Entitlement and wellbeing*, eds J. Drèze and A. Sen, 191–246. Oxford: Clarendon Press.

Deaton, A. and R. Arora 2009. Life at the top: the benefits of height. *Economics and Human Biology* 7(2): 133–136. https://doi.org/10.1016/j.ehb.2009.06.001

Depauw, E. and D. Oxley 2019. Toddlers, teenagers and terminal heights: The importance of puberty for male adult stature, Flanders 1800–76. *Economic History Review* 72(3): 925–952. https://doi.org/10.1111/ehr.12745

Eltis, D. 1982. Nutritional trends in Africa and the Americas: Heights of Africans, 1819–1839. *Journal of Interdisciplinary History* 12(3): 453–475. https://doi.org/10.2307/203269

Eveleth, P. and J. Tanner 1976. *Worldwide variation in human growth*. First edition. Cambridge: Cambridge University Press.

Eveleth, P. and J. Tanner 1990. *Worldwide variation in human growth*. Second edition. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511629105

Ezzati, M. et al. 2016. A century of trends in adult human height. *eLife*, 5: e13410. https://doi.org/10.7554/eLife.13410

Feinstein, C. H. 1998. Pessimism perpetuated: Real wages and the standard of living in Britain during and after the industrial revolution. *Journal of Economic History* 58(3): 625–658. https://doi.org/10.1017/S0022050700021100

Floud, R. 1984a. Measuring the transformation of the European economies: Income, health and welfare. CEPR Discussion Papers 33. London: Centre for Economic Policy Research.

Floud, R. 1984b. The heights of Europeans since 1750: A new source for European economic history'. *NBER Working Papers* No. 1318. https://doi.org/10.3386/w1318

Floud, R. 1994. The heights of Europeans since 1750: A new source for European economic history. *Stature, living standards and economic development*, ed J. Komlos, 9–24. Chicago: University of Chicago Press.

Floud, R. 1997. *The people and the British economy 1830–1914*. Oxford: Oxford University Press.

Floud, R. and B. Harris 1997. Health, height and welfare: Britain 1700–1980. In *Health and welfare since industrialisation*, eds R. Steckel and R. Floud, 91–126. Chicago: Chicago University Press.

Floud, R., R. Fogel, B. Harris and S. C. Hong 2011. *The changing body: Health, nutrition and human development in the western world since 1700.* Cambridge: Cambridge University Press.

Floud, R., B. Harris and S. C. Hong 2014. Introduction. In *Health, mortality and the standard of living in Europe and North America since 1700*, eds R. Floud, R. Fogel, B. Harris and S. C. Hong, vol. 1, xiii–xlii. Cheltenham: Edward Elgar.

Floud, R., K. Wachter and A. Gregor 1990. *Height, health and history: Nutritional status in the United Kingdom 1750–980*, Cambridge: Cambridge University Press.

Floud, R., K. Wachter and A. Gregory 1993a. Measuring historical heights: Short cuts or the long way round? A reply to Komlos. *Economic History Review* 46(1): 145–154. https://doi.org/10.1111/j.1468-0289.1993.tb01326.x

Floud, R., K. Wachter and A. Gregory 1993b. Further thoughts on the nutritional status of the British population. *Economic History Review* 46(2): 367–368. https://www.jstor.org/stable/2598023

Fogel, R. 1986. Nutrition and the decline in mortality since 1700: Some preliminary findings. In *Long-term factors in American economic growth*, eds S. Engerman and R. Gallman, 439–555. Chicago: University of Chicago Press. https://doi.org/10.7208/9780226209319-010

Fogel, R., S. Engerman, J. Trussell, R. Floud, C. Pope and L. Wimmer 1978. The economics of mortality in North America, 1650–1910: A description of a research project. *Historical Methods* 11(2): 75–108. https://doi.org/10.1080/01615440.1978.9955221

Friedman, G. 1982. The heights of slaves in Trinidad. *Social Science History* 6(4): 482–515. https://doi.org/10.2307/1170973

Gallardo-Albarrán, D. and H. DeJong 2021. Optimism or pessimism? A composite view on English living standards during the industrial revolution. *European Review of Economic History* 25(1): 1–19. https://doi.org/10.1093/ereh/heaa002

Galofré-Vilà, G. 2018. Growth and maturity: A quantitative systematic review and network analysis in anthropometric history. *Economics and Human Biology* 28: 107–118. https://doi.org/10.1016/j.ehb.2017.12.003

Galofré-Vilà, G., A. Hinde and A. Guntupalli 2018. Heights across the last 2000 years in England. *Research in Economic History* 34: 67–98. https://doi.org/10.1108/S0363-326820180000034003

Gao, P. and E. Schneider 2020. The growth pattern of British children, 1850–1975. *Economic History Review*. https://doi.org/10.1111/ehr.13002

Grasgruber, P., J. Cacek, T. Kalina and M. Sebera 2014. The role of nutrition and genetics as key determinants of the positive height trend. *Economics and Human Biology* 15(1): 81–100. https://doi.org/10.1016/j.ehb.2014.07.002

Grasgruber, P., M. Sebera, Hrazdíra, E., J. Cacek and T. Kalina 2016. Major correlates of male height: a study of 105 countries. *Economics and Human Biology* 21: 172–195. https://doi.org/10.1016/j.ehb.2016.01.005

Gruenwald, P., H. Funakawa, S. Mitani, T. Nishimura and S. Takeuchi 1967. Influence of environmental factors on foetal growth in man. *Lancet* 289(7498): 1026–1029. https://doi.org/10.1016/S0140-6736(67)91541-3

Harris, B. 1993. The demographic impact of the First World War: An anthropometric perspective. *Social History of Medicine* 6(3): 343–66. https://doi.org/10.1093/shm/6.3.343

Harris, B. 1997. Growing taller, living longer? Anthropometric history and the future of old age. *Ageing and Society* 17(5): 491–512. https://doi.org/10.1017/S0144686X97006594

Harris, B. 2001. Height and nutrition. In *The Cambridge world history of food*, eds K. Kiple and K. C. Ornelas, vol. 2, 1427–1438. New York: Cambridge University Press. https://doi.org/10.1017/CHOL9780521402156.032

Harris, B. 2008a. Review of S. Kunitz, The health of populations: General theories and particular realities. *Economics and Human Biology* 6(3): 492–494. https://doi.org/10.1016/j.ehb.2008.07.005

Harris, B. 2008b. Gender, health and welfare in England and Wales since industrialisation. *Research in Economic History* 26: 157–204. https://doi.org/10.1016/S0363-3268(08)26003-9

Harris, B. 2009. Anthropometric history, gender and the measurement of wellbeing. In *Gender and wellbeing in Europe: Historical and contemporary perspectives*, eds B. Harris, L. Gálvez and H. Machado, 59–84. Farnham: Ashgate.

Harris, B. 2014. Health and welfare. In *The Cambridge economic history of modern Britain*, eds R. Floud, J. Humphries and P. Johnson, Vol. 2. 1870–2010, 122–150. Cambridge: Cambridge University Press.

Harris, B. 2016. Food supply, health and economic development in England and Wales during the eighteenth and nineteenth centuries. *Scientia Danica, Series H, Humanistica* 4(7): 139–152. http://publ.royalacademy.dk/backend/web/uploads/2020-02-14/AFL%206/SDH4_7_00_00_2016_5868/SDH4_7_00_00_2016_5868.pdf.

Harris, B. and J. Helgertz 2019. Urban sanitation and the decline of mortality. *History of the Family* 24(2): 207–26. https://doi.org/10.1080/1081602X.2019.1605923

Harris, B., R. Floud and S. C. Hong 2015. How many calories? Food availability in England and Wales in the 18th and 19th centuries. *Research in Economic History* 31: 111–191. https://doi.org/10.1108/S0363-326820150000031003

Hatton, T. 2011. Infant mortality and the health of survivors: Britain, 1910–50. *Economic History Review* 64(3): 951–972. https://doi.org/10.1111/j.1468-0289.2010.00572.x

Hatton, T. 2014. How have Europeans grown so tall?. *Oxford Economic Papers* 66(2): 349–372. https://doi.org/10.1093/oep/gpt030

Hatton, T. and B. Bray 2010. Long run trends in the heights of European men, 19th–20th centuries. *Economics and Human Biology* 8(3): 405–413. https://doi.org/10.1016/j.ehb.2010.03.001

Hatton, T. and R. Martin 2010. The effects on stature of poverty, family size and birth order: British children in the 1930s. *Oxford Economic Papers* 62(1): 157–184. https://doi.org/10.1093/oep/gpp034

Hermanussen, M., R. Largo and L. Molina 2001. Canalisation in human growth: A widely accepted concept reconsidered. *European Journal of Pediatrics* 160(3): 163–167. https://doi.org/10.1007/s004310000706

Hong, S. C. 2007. The burden of early exposure to malaria in the United States, 1850–1860: malnutrition and immune disorders. *Journal of Economic History* 67(4): 1001–1035. https://doi.org/10.1017/S0022050707000472

Horrell, S., D. Meredith and D. Oxley 2009. Measuring misery: Body mass, ageing and gender inequality in Victorian London. *Explorations in Economic History* 46(1): 93–119. https://doi.org/10.1016/j.eeh.2007.12.001

Howe, P. and M. Schiller 1952. Growth responses of the school child to changes in diet and environmental factors. *Journal of Applied Physiology* 5(2): 51–61. https://doi.org/10.1152/jappl.1952.5.2.51

Jones, C. and P. Klenow 2016. Beyond GDP: Welfare across countries and time. *American Economic Review* 106(9): 2426–2457. https://doi.org/10.1257/aer.20110236

Koepke, N. and J. Baten 2005. The biological standard of living in Europe during the last two millennia. *European Review of Economic History* 9(1): 61–95. https://doi.org/10.1017/S1361491604001388

Koepke, N. and J. Baten 2008. Agricultural specialisation and height in ancient and medieval Europe. *Explorations in Economic History* 45(2): 127–146. https://doi.org/10.1016/j.eeh.2007.09.003

Koepke, N., J. Floris, C. Pfister, F. Rühli and K. Staub 2018. Ladies first: Female and male adult height in Switzerland, 1770–1930. *Economics and Human Biology* 29: 76–87. https://doi.org/10.1016/j.ehb.2018.02.002

Komlos, J. 1987. The height and weight of West Point cadets: Dietary change in anetebellum America. *Journal of Economic History* 47(4): 897–927. https://doi.org/10.1017/S002205070004986X

Komlos, J. 1993a. The secular trend in the biological standard of living in the United Kingdom, 1730–1860. *Economic History Review* 46(1): 115–144. https://www.jstor.org/stable/2597683

Komlos, J. 1993b. Further thoughts on the nutritional status of the British population. *Economic History Review* 46(2): 363–366. https://doi.org/10.1111/j.1468-0289.1993.tb01338.x

Komlos, J. 2004. How to and how not to analyse deficient height samples. *Historical Methods* 17(4): 160–173. https://doi.org/10.3200/HMTS.37.4.160-173

Komlos, J. 2019. Shrinking in a growing economy is not so puzzling after all. *Economics and Human Biology* 32: 40–55. https://doi.org/10.1016/j.ehb.2018.11.005

Komlos, J. and B. A'Hearn 2016. The decline in the nutritional status of the U.S. antebellum population at the onset of modern economic growth. NBER Working Papers No. 21845. https://doi.org/10.3386/w21845

Komlos, J. and B. A'Hearn 2019. Clarifications of a puzzle: The decline in nutritional status at the onset of modern economic growth in the USA. *Journal of Economic History* 79(4): 1129–53. https://doi.org/10.1017/S0022050719000573

Komlos, J. and H. Küchenhoff 2012. The diminution of the physical stature of the English male population in the eighteenth century. *Cliometrica* 6(1): 45–62. https://doi.org/10.1007/s11698-011-0070-7

Kopczyński, M. and M. Rodak 2020. The Polish *interbella* puzzle: The biological standard of living in the Second Polish Republic. *Economic History Review* 74(1):181–203. https://doi.org/10.1111/ehr.13046

Leon, D., I. Koupilova, H. Lithell, L. Berglund, R. Mohsen, D. Vågerö, U.-B. Lithell and P. McKeigue 1996. Failure to realise growth potential in utero and adult obesity in relation to blood pressure in 50-year old Swedish men. *British Medical Journal* 312(7028): 401–406. https://doi.org/10.1136/bmj.312.7028.401

Ling, J. and N. King 1987. Secular trends in stature and weight in southern Chinese children in Hong Kong. *Annals of Human Biology* 14(2): 187–190. https://doi.org/10.1080/03014468700006912

Margo, R. and R. Steckel 1982. The heights of American slaves: New evidence on slave nutrition and health. *Social Science History* 6(4): 516–538. https://doi.org/10.2307/1170974

McKeown, T. 1976. *The modern rise of population*. London: Edward Arnold.

McKeown, T. and R. Brown 1955. Medical evidence related to English population changes in the eighteenth century. *Population Studies* 9(2): 119–141. https://doi.org/10.1080/00324728.1955.10404688

McKeown, T. and R. Record 1962. Reasons for the decline of mortality in England and Wales during the nineteenth century. *Population Studies* 16(2): 94–122. https://doi.org/10.1080/00324728.1962.10414870

McKeown, T., R. Brown and R. Record 1972. An interpretation of the modern rise of population in Europe. *Population Studies* 26(3): 345–382. https://doi.org/10.1080/00324728.1972.10405908

Meinzer, N., R. Steckel and J. Baten 2019. Agricultural specialisation, urbanisation, workload and stature. In *The backbone of Europe: Health, diet, work and violence over two millennia*, eds R. Steckel, C. S. Larsen, C. Roberts and J. Baten, 231–252. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108379830.010

Meredith, D. and D. Oxley 2014. Food and fodder: Feeding England, 1700–1900. *Past and Present* 222: 163–214. https://doi.org/10.1093/pastj/gtt020

Mokyr, J. and Ó Gráda, C. 1996. Height and health in the United Kingdom, 1815–60: evidence from the East India Company Army. *Explorations in Economic History* 33(2): 141–168. https://doi.org/10.1006/exeh.1996.0007

Murray, J. 2000. Marital protection and marital selection: Evidence from a historical-prospective sample of American men. *Demography* 37(4): 511–521. https://doi.org/10.1353/dem.2000.0010

NCD Risk Factor Collaboration (NCD-RisC) 2020. Height and body-mass index trajectories of school-aged children and adolescents from 1985 to 2019 in 200 countries and territories: A pooled analysis of 2181 population-based studies with 65 million participants. *Lancet* 396: 1511–1524. https://doi.org/10.1016/S0140-6736(20)31859-6

Öberg, S. 2017. Too many is not enough: Studying how children are affected by their number of siblings and resource dilution in families. *History of the Family* 22(2–3): 157–174. https://doi.org/10.1080/1081602X.2017.1302890

Oddy, D. 1982. The health of the people. In *Population and society in Britain, 1850–1980*, eds T. Barker and M. Drake, 121–140. London: Batsford.

Osmani, S. and Sen A. 2003. The hidden penalties of gender inequality: Fetal origins of ill-health. *Economics and Human Biology* 1(1):105–121. https://doi.org/10.1016/S1570-677X(02)00006-0

Perkins, J., S. Subramanian, G. Davey Smith and E. Özaltin 2016. Adult height, nutrition and population health. *Nutrition Reviews* 74(3): 149–165. https://doi.org/10.1093/nutrit/nuv105

Prados de la Escosura, L. 2019. Human development in the age of globalisation. European Historical Economics Society Working Paper Series 157. http://www.ehes.org/EHES_157.pdf

Prentice, A. M., K. Ward, G. Goldberg, M. J. Landing, S. Moore, A. Fulford and A. Prentice 2013. Critical windows for nutritional interventions against stunting. *American Journal of Clinical Nutrition* 97(5): 911–918. https://doi.org/10.3945/ajcn.112.052332

Schneider, E. and K. Ogasawara 2018. Disease and child growth in industrialising Japan: Critical windows and the growth pattern, 1917–39. *Explorations in Economic History* 69: 64–80. https://doi.org/10.1016/j.eeh.2018.05.001

Schultz, T. P. 2002. Wage gains associated with height as a form of health human capital. *American Economic Review* 92(2): 349–353. https://doi.org/10.1257/000282802320191598

Sen, A. 1987. *The standard of living*. Cambridge: Cambridge University Press.

Sen, A. 1999. *Development as freedom*. Oxford: Oxford University Press.

Sharpe, P. 2012. Explaining the short stature of the poor: Chronic childhood disease and growth in nineteenth-century England. *Economic History Review* 65(4): 1475–1494. https://doi.org/10.1111/j.1468-0289.2011.00629.x

Silventoinen, K. 2003. Determinants of variation in adult body height. *Journal of Biosocial Science* 35(2): 263–285. https://doi.org/10.1017/S0021932003002633

Sinclair, D. 1978. *Human growth after birth*. Third edition. London: Oxford University Press.

Singhal, A. 2017. Long-term adverse effects of early growth acceleration of catch-up growth. *Annals of Nutrition and Metabolism* 70: 236–240. https://doi.org/10.1159/000464302

Sohn, K. 2015. The value of male height in the marriage market. *Economics and Human Biology* 18: 110–124. https://doi.org/10.1016/j.ehb.2015.05.004

Steckel, R. 1983. Height and *per capita* income. *Historical Methods* 16(1): 1–7. https://doi.org/10.1080/01615440.1983.10594092

Steckel, R. 1986. A peculiar population: The nutrition, health and mortality of American slaves from childhood to maturity. *Journal of Economic History* 46(3): 721–741. https://doi.org/10.1017/S0022050700046842

Steckel, R. 1992. Stature and living standards in the United States. In *American economic growth before the civil war*, eds R. Gallman and J. Wallis, 265–308. Chicago: University of Chicago Press. https://doi.org/10.7208/9780226279473-008

Steckel, R. 1995. Stature and the standard of living. *Journal of Economic Literature* 33(4): 1903–1940. https://www.jstor.org/stable/2729317

Steckel, R. 2008. Biological measures of the standard of living. *Journal of Economic Perspectives* 22(1): 129–152. https://doi.org/10.1257/jep.22.1.129

Steckel, R. 2009. Height and human welfare: Recent developments and new directions. *Explorations in Economic History* 46(1): 1–23. https://doi.org/10.1016/j.eeh.2008.12.001

Steckel, R. 2016. Biological measures of wellbeing. In *The Oxford handbook of economics and human biology*, eds J. Komlos and I. Kelly, 32–51. New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199389292.013.25

Steckel, R., C. S. Larsen, C. Roberts and J. Baten (eds) 2019. *The backbone of Europe: Health, diet, work and violence over two millennia*. Cambridge: Cambridge University Press.

Szreter, S. and G. Mooney 1998. Urbanisation, mortality and the standard of living debate: New estimates of the expectation of life at birth in nineteenth-century British cities. *Economic History Review*. 51(1): 84–112. https://doi.org/10.1111/1468-0289.00084

Tanner, J. 1962. *Growth at adolescence, with a general consideration of the effects of hereditary and environmental factors upon growth and maturation from birth to maturity*. Second edition. Oxford: Blackwell.

Tanner, J. 1987. Growth as a mirror of the condition of society: Secular trends and class distinctions. *Pediatrics International* 29(1): 96–103. https://doi.org/10.1111/j.1442-200X.1987.tb00015.x

Tanner, J. 1990. *Fetus into man: Physical growth from conception to maturity*. Second edition. Cambridge, MA: Harvard University Press.

Tanner, J., T. Hayashi, M. Preece and N. Cameron 1982. Increase in length of leg relative to trunk size in Japanese children from 1957 to 1977: Comparisons with British and with Japanese Americans. *Annals of Human Biology* 9(5): 411–423. https://doi.org/10.1080/03014468200005951

Tanner, J., R. Whitehouse and M. Takaishi 1966. Standards from birth to maturity for height, weight, height velocity, weight velocity: British children, 1965: Part II. *Archives of Disease in Childhood* 41(220): 613–635. http://doi.org/10.1136/adc.41.220.613

Usher, D. 1980. *The measurement of economic growth*. Oxford: Blackwell.

Villermé, L. R. 1829. Mémoire sur la taille de l'homme en France. *Annales d'Hygiène Publique* 1: 351–399.

Visscher, P. 2008. Sizing up human height variation. *Nature Genetics* 40(5): 489–490. https://doi.org/10.1038/ng0508-489

Waaler, H. 1984. Height, weight and mortality: The Norwegian experience. *Acta Medica Scandinavica* 215(S679): 1–56. https://doi.org/10.1111/j.0954-6820.1984.tb12901.x

Waddington, C. H. 1942. Canalization of development and the inheritance of acquired characteristics. *Nature* 150: 563–565. https://doi.org/10.1038/150563a0

Weitzman, A. and D. Conley 2014. From assortative to ashortative coupling: Men's height, height heterogamy, and relationship dynamics in the United States. NBER Working Papers No. 20402. https://doi.org/10.3386/w20402

Wells, J. C. K. 2016. *The metabolic ghetto: An evolutionary perspective on nutrition, power relations and chronic diseases*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511972959

Yamamura, E. and Y. Tsutsui 2017. Comparing the role of the height of men and women in the marriage market. *Economics and Human Biology* 26: 42–50. https://doi.org/10.1016/j.ehb.2017.02.006

Zimran, A. 2019. Sample-selection bias and height trends in the nineteenth-century United States. *Journal of Economic History* 79(1): 99–138. https://doi.org/10.1017/S0022050718000694

Research Articles

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 127–140

Going beyond GDP with a parsimonious indicator: Inequality-adjusted healthy lifetime income

***David E. Bloom*[1], *Victoria Y. Fan*[2], *Vadim Kufenko*[3], *Osondu Ogbuoji*[4], *Klaus Prettner*[5,6,*] *and Gavin Yamey*[4]**

Abstract

Per capita GDP has limited use as a well-being indicator because it does not capture many dimensions that imply a "good life", such as health and equality of opportunity. However, per capita GDP has the virtues of being easy to interpret and to calculate with manageable data requirements. Against this backdrop, there is a need for a measure of well-being that preserves the advantages of per capita GDP, but also includes health and equality. We propose a new parsimonious indicator to fill this gap, and calculate it for 149 countries. This new indicator could be particularly useful in complementing standard well-being indicators during the COVID-19 pandemic. This is because (i) COVID-19 predominantly affects older adults beyond their prime working ages whose mortality and morbidity do not strongly affect GDP, and (ii) COVID-19 is known to have large effects on inequality in many countries.

Keywords: beyond GDP; well-being; health; inequality; human development; lifetime income; COVID-19

[1]Harvard T.H. Chan School of Public Health, Boston, MA, USA

[2]University of Hawai'i at Mānoa, Office of Public Health Studies, Myron B. Thompson School of Social Work, Honolulu, HI, USA

[3]University of Hohenheim, Stuttgart, Germany

[4]Duke University, Duke Global Health Institute, Durham, NC, USA

[5]Vienna University of Economics and Business, Vienna, Austria

[6]Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

[*]Correspondence to: Klaus Prettner, Klaus.Prettner@wu.ac.at

DOI: 10.1553/populationyearbook2021.res1.1

1 Introduction

Per capita GDP has limited use in measuring well-being. The reasons why this is the case are well known:[1] for example, per capita GDP does not capture negative externalities of production, catastrophes increase GDP because of reconstruction efforts and GDP does not take into account the quality of the natural environment. Most importantly – and partly as a consequence of these shortcomings – per capita GDP disregards generally desired aspects such as living a long and healthy life, or providing larger parts of the population the opportunity to share in the gains of economic prosperity (Fitoussi et al. 2009; Jones and Klenow 2016; Fan et al. 2018; Lutz et al. 2018; Bloom et al. forthcoming).

The following comparison between Germany and Iceland in 2013 illustrates the consequences of these omissions. According to the World Bank (2019), the two countries had comparable levels of GDP per capita (USD42,914 and USD42,372, respectively, adjusted for purchasing power). Thus, based on these numbers, one would surmise that the typical German was slightly better off than the typical Icelander. However, this misses the point that life expectancy at birth was 80.49 years in Germany, compared with 82.06 years in Iceland. Even if we disregard the intrinsic value of health, this implies that the *lifetime income* of the average Icelander under current conditions (calculated as life expectancy multiplied by GDP per capita) was higher than that of the average German. Similarly, the Gini index of income inequality in Germany stood at 31.1 (expressed in per cent) and at 25.4 in Iceland. Considering that income distributions are skewed towards higher incomes, this implies that the median Icelander was likely to be better off financially than the median German, even when disregarding differences in lifetime horizons.

Despite the problems outlined above, per capita GDP does have the virtues of being easy to interpret and to calculate with manageable data requirements. Against this backdrop, there is a need for a measure that combines the advantages of per capita GDP with the virtues of including health and equality. In the following, we propose inequality-adjusted healthy lifetime income (IHLI) as such an indicator. The resulting number, expressed in PPP-adjusted dollars, refers to the amount that a new-born in a certain economy can expect to earn over the years in which she is in good health under the given economic and health conditions, and adjusted for the level of inequality.

Our new indicator, IHLI, could be a useful complement to standard well-being indicators, particularly during periods in which levels of health and inequality change drastically. The current COVID-19 pandemic represents just such a situation, because (i) COVID-19 predominantly affects older adults beyond their prime working ages whose mortality and morbidity do not strongly affect GDP, and (ii) COVID-19 is known to have large effects on inequality in many countries.

---

[1] Simon Kuznets, who is credited with the original formulation of GDP, already warned against its use as a welfare measure.

Even though IHLI is able to address some of the shortcomings of standard well-being measures, it also has limitations: for example, (i) it does not include measures related to education; (ii) it is more difficult to compute and requires more data than per capita GDP; and (iii) it depends on the quality of the underlying health measure, which is known to be difficult to estimate across different countries. Thus, of course, we do not propose IHLI as a substitute for other indicators. We simply hope that IHLI provides additional valuable information for policymakers and the general public.

2 Inequality-adjusted healthy lifetime income

Several indicators have been proposed to address the problems of per capita GDP as a well-being measure, including the Human Development Index (HDI), the Happy Planet Index, the Resource-Infrastructure-Environment (RIE) index, Gross National Happiness and the Better Life Index (see ul Haq 2003; Natoli and Zuhair 2011; Fan et al. 2018; OECD 2019; New Economics Foundation 2019; for detailed information). For example, the HDI considers three components, gross national income (GNI), life expectancy and schooling, and merges the sub-indices constructed out of these components into a single overall index. This index is then used to create country league tables with a country's rank being a proxy for its development level. A similar approach has been proposed by Natoli and Zuhair (2011) in their RIE, with the sub-components comprising indices for the availability of resources, the availability of infrastructure and the quality of the environment.

While these alternative indicators represent highly valuable improvements over GDP along several dimensions, they also tend to have high data requirements (on issues such as schooling, environmental quality, housing, civic engagement and work-life balance). Moreover, these indicators are often based on subjective evaluations of life satisfaction or happiness that are collected by polling a small subset of the population, and they are often hard to interpret because different components with incompatible units of measurement are meshed together to construct an overall index. As an illustration, consider the HDI (ul Haq 2003; Klugman et al. 2011). Because this indicator's components of income, life expectancy and education are measured in different units, an overall index ranging from zero to one must be constructed out of the different components' sub-indices. The resulting index lacks an economic interpretation, and has an upper limit of one by construction. The upper limit implies that further development cannot change the index value appreciably, and that well-developed countries tend to cluster at index values close to one. Thus, at the upper end of the distribution, only marginal differences across countries may be evident in the index value, while the underlying fundamental data on life expectancy, income and schooling could differ substantially.

To address the trade-off of including additional dimensions of economic well-being, while ensuring that the indicator remains easy to calculate, easy to interpret and based on readily available data, we propose using inequality-adjusted healthy lifetime income (IHLI) as a novel and complementary measure of economic well-being. This measure consists of the following components: (i) GDP per capita adjusted for purchasing power (pppGDPpc) to capture a country's material living standard; (ii) healthy life expectancy at birth (HALE) to capture health-related aspects such as environmental quality and access to high-quality medical facilities; and (iii) an inverse measure of the Gini coefficient (1 – Gini) as a proxy for an average person's opportunities to benefit from economic progress (Sen 1976).[2] The following straightforward formulation of the indicator

$$IHLI_i = pppGDPpc_i \times HALE_i \times (1 - Gini_i) \tag{1}$$

implies that it should be interpreted as reflecting the amount that a new-born in economy *i* can expect to earn over the years in which she is in good health under the given economic and health conditions, and adjusted for the level of inequality. Note that the unitary weights of the different components in this formulation follow mathematically from the underlying units of the different components: since the outcome is inequality-adjusted healthy lifetime income, it does not make sense to use a different weighting scheme in Equation (1).[3]

Using the World Bank's (2019) World Development Indicators on pppGDPpc in international dollars with a base year of 2011, the WHO's (2014) Global Health Observatory database on HALE in years and Solt's (2019) Standardized World Income Inequality Database on the Gini coefficient of disposable income, we calculate this indicator for the year 2010 for all countries for which the necessary data inputs are available. Table A.1 in the appendix displays the results. The first column shows each country's rank as measured by the IHLI indicator; the second column provides each country's name; the third to fifth columns show the three components of IHLI; and the sixth column reports the IHLI value.

Interestingly, the IHLI indicator alters some standard rankings that are based solely on per capita GDP. For example, among high-income countries, the United States and Saudi Arabia have comparatively low IHLI values despite their high per capita GDP because of their rather low values for healthy life expectancy and rather high inequality levels. By contrast, some European countries, such as Austria, Belgium, Denmark and Sweden, have comparatively high IHLI values despite their lower per capita GDP because they have rather high healthy life expectancy values and rather low inequality levels.

---

2 While there are discussions on the issue of whether a single measure of inequality can capture the entire distribution reasonably well, the Gini coefficient possesses some desirable properties in this regard, and it also tends to be strongly correlated with other inequality measures. For discussions, see Cowell (2011) and Ferreira (2020).

3 To see why this is the case, consider the following analogue. Assume that we would like to calculate the distance travelled by a car within a certain time span. Then we need to multiply the speed of the car by the duration of the travel. Applying a different weight to the speed of the car and the duration of the travel would be meaningless.

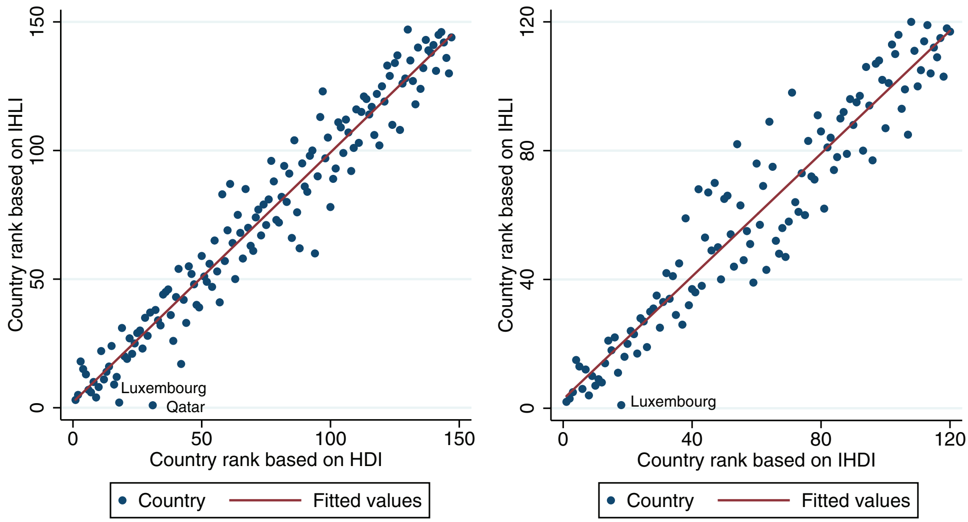
The differences between the traditional ranking based on per capita GDP values and the ranking based on the IHLI indicator can be decomposed into the contributions of the two additional components: the correlation between the differences in the country league table based on per capita GDP and the country league table based on IHLI is −0.654 for the Gini index and 0.299 for HALE, with both coefficients being significant at the 1% level. Thus, changes in the Gini coefficient contribute relatively more in changing the ranking, which is due to the fact that there is more cross-country variation in this variable than in HALE. While it is not uncommon for Gini indices between countries to differ by a factor of two, this is generally not the case for HALE. Overall, considering the additional dimensions of healthy life expectancy and inequality may contribute substantially to revising the ranking of the countries to account for well-being.

While the adjustments discussed above seem reasonable, we compare IHLI with some obvious alternatives – the HDI and its inequality-adjusted version (Hicks 1997; UNDP 2017) – to assess whether the different well-being indicators yield country rankings that are generally in line with one another. This is important, because if the ranking based on IHLI differs substantially from the ranking based on an established indicator such as the HDI and its inequality-adjusted version, it is likely that IHLI is not well suited for assessing cross-country differences in well-being. In our comparison, we need to restrict the dataset to 120 countries because the inequality-adjusted HDI is only available for this subset.

Figure 1 displays the correlation between the country rankings based on the IHLI and the HDI (left diagram; correlation coefficient: 0.9710) and between the IHLI and the inequality-adjusted HDI (right diagram; correlation coefficient: 0.9536). Overall, this analysis shows a strong positive correlation between the rankings based on the different indicators. This result is reassuring, as a different outcome might indicate that we are not measuring well-being accurately, and might therefore be missing important well-being dimensions that the HDI captures. Even though the different measures lead to consistent country rankings, our proposed indicator has the following advantages:

- It has a direct and immediate economic interpretation.
- It does not depend on aggregating different sub-indicators that are based on incompatible units of measurement.
- It is not restricted to a value between zero and one, and is thus not bounded from above, which allows for further development to be measured accurately, and avoids a clustering of countries at the upper bound of the index.
- It is parsimonious in terms of its computation and data input requirements.
- It can be obtained for more countries.
- The weights of its components follow directly from the interpretation of the indicator.

Figure 1:
Comparison between IHLI and HDI rankings, 2010 (left diagram: HDI; right diagram: inequality-adjusted HDI)



3 Variants of IHLI

We view the formulation of IHLI in Equation (1) as an important improvement over per capita GDP. However, using gross national income (GNI) per capita instead of GDP per capita to measure income may prove useful because particularly small, open economies such as Luxembourg are highly dependent on commuters. In this case, GNI might capture income better than GDP because GNI counts only residents' income. In this case, our indicator would change to

$$IHLI_{i,gni} = pppGNIpc_i \times HALE_i \times (1 - Gini_i), \tag{2}$$

where *pppGNIpc* is the ppp-adjusted GNI, and the subscript in the indicator name signifies the use of GNI instead of GDP in the calculation.

Moreover, HALE itself is subject to criticism because it is defined as weighted life expectancy over a complete set of health states, with the weights being between zero (dead) and one (optimal health), as described in Sullivan (1971), Murray et al. (2002) and WHO (2014).[4] Thus, detailed morbidity data (on years of healthy life lost due to disability, YLD) are required to construct HALE, and these data might

[4] We are grateful to two referees who suggested delving more deeply into the problems with using HALE as an indicator of healthy life expectancy.

not be available for all countries over long time spans. In addition, the morbidity data are very difficult to estimate and the methods for their calculation change over time. Thus, while it would be ideal to construct our indicator with a measure of healthy life expectancy, another variant of the proposed indicator could rely on life expectancy at birth (LEXP) instead of HALE. Life expectancy arguably captures health less well than HALE, but it might be available for more countries and more time periods, which could allow for an extension of the sample when reconstructing the index over the past decades. Furthermore, life expectancy is easier to calculate, and is not subject to changes in the method of its calculation over time. In this case, the indicator would need to be renamed inequality-adjusted lifetime income (ILI), and the formula would change to

$$ILI_i = pppGDPpc_i \times LEXP_i \times (1 - Gini_i). \tag{3}$$

As yet another version, we could use life expectancy conditional on reaching a certain age, such as age 20, instead of life expectancy at birth in the construction of the indicator. In this case, we would have the formula

$$IRLI_i^{a=20} = pppGDPpc_i \times LEXP_i^{a=20} \times (1 - Gini_i), \tag{4}$$

where the interpretation of the indicator changes to inequality-adjusted remaining lifetime income (IRLI), and the superscript indicates conditioning on age 20. In addition, we would need to keep in mind that this version of the indicator disregards potentially important information on child mortality.

4 Limitations

Of course, no welfare measure is perfect, and the IHLI also has many limitations. For example, (i) per capita GDP might not be the best proxy of lifetime income because it does not reflect the future growth potential of an economy. If a country has a strong potential for economic growth over the coming decades, then a child born into this economy could expect to earn a much higher lifetime income than a child born into a stagnant or even shrinking economy. However, adjusting for the growth potential of an economy is very complicated, because economic growth can sometimes stall rather unexpectedly (witness Japan in the 1990s, or the global economy during the COVID-19 crisis). (ii) Period healthy life expectancy might not be a good proxy of cohort healthy life expectancy. If mortality continues to decrease for all age groups, a child born today can reasonably expect to outlive current period life expectancy. For this reason, our indicator is on the conservative side. (iii) Income inequality is assumed to be stable over the life cycle, which is generally not the case. (iv) Income inequality is assumed to be a good proxy for health inequality, which might not always be true. (v) Like standard aggregate macroeconomic indicators, this indicator disregards race, gender, migration status, etc. Finally, (vi) in contrast to the HDI and the IHDI, the IHLI lacks the education dimension, which would be difficult to add in terms of interpretation. Overall, of course, we do not claim that our indicator eliminates all of the problems of established indicators, and that it should replace them. Rather, we merely suggest that IHLI solves some of the problems of standard well-being measures, and that it might be worthwhile to report IHLI alongside other indicators.

5 Conclusions

We have proposed a novel indicator for measuring economic well-being that accounts for income, health and inequality, and that can be readily interpreted as inequality-adjusted healthy lifetime income. Although this indicator captures more dimensions of well-being than per capita GDP, it nevertheless remains easy to calculate and easy to interpret, and it requires limited data. A country ranking for the year 2010 shows some reasonable deviations from a ranking based on per capita GDP. While the IHLI-based country rankings are consistent with the rankings based on other established indicators such as the HDI, the IHLI does not share the HDI's shortcomings, and it is available for more countries (149 countries instead of 120 countries for the inequality-adjusted HDI). It might be useful to develop different versions of the indicator in order to further increase the availability of data, and to adjust for distortions that are caused by commuting into small, open economies. Overall, our proposed indicator might allow for better comparisons of well-being across countries and over time.

However, we do not advocate abandoning per capita GDP, as it continues to be a highly valuable gauge of the economic strength of a country. Per capita GDP has many advantages over other proposed indicators because it includes both capital income and wage income. If only the wage level is considered, capital income, which is a substantial source of income for many people, is neglected. If only consumption measures are considered, saving/investment and borrowing are disregarded. In addition, per capita GDP is available over long periods and for many countries, which is typically not the case for alternative income measures.

Acknowledgements

Vadim Kufenko gratefully acknowledges the support provided by the Faculty of Business, Economics, and Social Sciences at the University of Hohenheim within its research focus "Inequality and Economic Policy Analysis (INEPA)". We would like to thank the managing editor of the Vienna Yearbook of Population Research, Maria Winkler-Dworak; two anonymous referees; Vanessa di Lego; Tora Estep; Bernhard Hammer; Michael Kuhn; Wolfgang Lutz; Erich Strießnig; Joshua Wilde; the participants of the Wittgenstein Centre conference "Demographic Aspects of Human Well-Being"; and the participants of the 19th meeting of German, Austrian and Swiss demographers (DACH19) for their valuable comments and suggestions.

References

Bloom, D. E., M. Kuhn and K. Prettner 2020. Modern infectious diseases: Macroeconomic impacts and policy responses. *Journal of Economic Literature* (forthcoming).

Cowell, F. 2011. *Measuring inequality*. London School of Economics Perspectives in Economic Analysis. Oxford, UK: Oxford University Press.

Fan V. Y., D. E. Bloom, O. Ogbuoji, K. Prettner and G. Yamey 2018. Valuing health as development: Going beyond gross domestic product. *The BMJ* 363: k4371. https://doi.org/10.1136/bmj.k4371

Ferreira, F. 2020. In defense of the Gini coefficient. URL: https://blogs.worldbank.org/developmenttalk/defense-gini-coefficient, accessed on 31 July 2020.

Fitoussi, J. P., A. Sen and J. Stiglitz 2009. Report by the commission on the measurement of economic performance and social progress. The Commission on the Measurement of Economic Performance and Social Progress. Paris, France.

Hicks, D. A. 1997. The inequality-adjusted human development index: A constructive proposal. *World Development* 25(8): 1283–1298. https://doi.org/10.1016/S0305-750X(97)00034-X

Jones, C. I. and P. J. Klenow 2016. Beyond GDP? Welfare across countries and time. *American Economic Review* 106(9): 2426–2457. https://doi.org/10.1257/aer.20110236

Klugman, J., F. Rodriguez and H. J. Choi 2011. The HDI 2010: New controversies, old critiques. *The Journal of Economic Inequality* 9(2): 249–288. https://doi.org/10.1007/s10888-011-9178-z

Lutz, W., A. A. Lijadi, E. Striessnig, A. Dimitrova and M. Caldeira Brant de Souza Lima 2018. Years of good life (YoGL): A new indicator for assessing sustainable progress. IIASA Working Paper WP-18-007, Laxenburg, Austria.

Murray, C. J. I., J. A. Salomon, C. D. Mathers and A. D. Lopez 2002. *Summary measures of population health. Concepts, ethics, measurement, and applications*. Geneva, Switzerland: World Health Organization.

Natoli, R. and S. Zuhair 2011. Measuring progress: A comparison of the GDP, HDI, GS and the RIE. *Social Indicators Research* 103(1): 33–56. https://doi.org/10.1007/s11205-010-9695-3

New Economics Foundation 2019. Happy planet index. http://happyplanetindex.org/countries, accessed on 18 August 2019.

OECD (Organisation for Economic Co-operation and Development) 2019. Better life index. http://www.oecdbetterlifeindex.org/#/11111111111, accessed on 8 August 2019.

Sen, A. 1976. Real national income. *The Review of Economic Studies* 43(1): 19–39.

Solt, F. 2019. Measuring income inequality across countries and over time: The standardized world income inequality database. SWIID Version 8.1, May 2019. https://fsolt.org/swiid/, accessed on 20 May 2019.

Sullivan, D. F. 1971. A single index of mortality and morbidity. *HSMHA Health Reports* 86: 347–354.

ul Haq, M. 2003. The birth of the human development index. In: *Readings in human development*, Oxford. https://sites.hks.harvard.edu/cid/events/papers/ulhaq_humandevelopindexchap_030201.pdf

UNDP (United Nations Development Programme) 2017. Human development reports: data. http://hdr.undp.org/en/data, accessed on 20 May 2019.

World Bank 2019. World development indicators. http://datatopics.worldbank.org/world-development-indicators/, accessed on 20 May 2019.

WHO (World Health Organization) 2014. WHO methods for life expectancy and healthy life expectancy. https://www.who.int/healthinfo/statistics/LT_method_1990_2012.pdf, accessed on 31 July 2020.

WHO (World Health Organization) 2019. Global health observatory (GHO) data. https://www.who.int/gho/mortality_burden_disease/life_tables/hale/en/, accessed on 20 May 2019.

Appendix

Table A.1:
Ranking of countries according to the IHLI indicator as of 2010

Rank	Country	pppGDPpc	HALE	Gini	IHLI
1	Qatar	125,141	66.7	0.397	5,033,177
2	Luxembourg	91,743	71.7	0.280	4,736,156
3	Norway	62,350	71.8	0.246	3,375,477
4	Singapore	72,116	74.8	0.393	3,274,319
5	Switzerland	55,866	72.5	0.290	2,875,718
6	Netherlands	46,102	71.4	0.265	2,419,374
7	Denmark	43,998	70.4	0.253	2,313,825
8	Sweden	42,989	71.9	0.254	2,305,808
9	Austria	43,336	71.4	0.282	2,221,639
10	Ireland	43,515	71.5	0.298	2,184,136
11	Belgium	41,086	70.8	0.257	2,161,300
12	Iceland	40,137	72.6	0.260	2,156,306
13	United States	49,479	68.7	0.370	2,141,511
14	Finland	39,848	70.4	0.254	2,092,760
15	Germany	40,429	70.9	0.287	2,043,741
16	Canada	40,699	72.2	0.311	2,024,622
17	Oman	45,336	65.5	0.319	2,022,228
18	Australia	41,464	72.2	0.329	2,008,775
19	France	36,815	72.2	0.294	1,876,574
20	Japan	35,750	73.8	0.314	1,809,896
21	Italy	36,201	72.8	0.331	1,763,112
22	United Kingdom	36,509	71.3	0.335	1,731,046
23	Cyprus	33,913	72.4	0.298	1,723,636

Continued

Table A.1: Continued

Rank	Country	pppGDPpc	HALE	Gini	IHLI
24	New Zealand	32,119	72.0	0.319	1,574,869
25	Spain	32,507	72.9	0.337	1,571,156
26	Saudi Arabia	45,421	64.4	0.478	1,526,916
27	Slovenia	28,678	69.2	0.247	1,494,361
28	Malta	28,359	71.6	0.272	1,478,194
29	Czech Republic	28,353	68.1	0.253	1,442,334
30	Greece	28,726	71.5	0.332	1,372,015
31	Israel	29,665	72.4	0.369	1,355,242
32	Portugal	27,238	70.8	0.337	1,278,584
33	Slovak Republic	25,159	66.8	0.257	1,248,705
34	Bahamas, The	29,222	66.4	0.439	1,088,531
35	Hungary	22,405	65.6	0.271	1,071,469
36	Estonia	22,741	66.7	0.323	1,026,889
37	Croatia	20,758	67.9	0.277	1,019,027
38	Poland	21,771	67.3	0.313	1,006,568
39	Lithuania	21,071	64.5	0.335	903,772
40	Kazakhstan	20,097	60.2	0.261	894,052
41	Malaysia	21,107	65.6	0.412	814,166
42	Seychelles	20,365	64.6	0.410	776,196
43	Belarus	16,261	62.3	0.245	764,841
44	Romania	17,469	65.2	0.331	761,983
45	Latvia	18,252	64.6	0.357	758,145
46	Argentina	18,712	67.3	0.399	756,852
47	Chile	19,442	68.9	0.453	732,738
48	Iran, Islamic Rep.	17,943	63.9	0.382	708,566
49	Uruguay	17,082	68.0	0.393	705,093
50	Lebanon	16,452	65.2	0.348	699,371
51	Turkey	17,959	64.4	0.404	689,319
52	Venezuela, RB	16,545	65.7	0.380	673,943
53	Bulgaria	15,283	65.7	0.332	670,746
54	Mauritius	15,938	64.5	0.366	651,770
55	Montenegro	14,038	67.1	0.312	648,082
56	Barbados	16,425	66.2	0.469	577,372
57	Panama	15,629	68.3	0.473	562,537
58	Mexico	15,716	66.5	0.463	561,225
59	Algeria	12,871	64.5	0.324	561,184
60	Serbia	12,688	66.2	0.339	555,208
61	Iraq	12,718	59.6	0.302	529,062

Continued

Table A.1:
Continued

Rank	Country	pppGDPpc	HALE	Gini	IHLI
62	Thailand	13,487	65.5	0.406	524,722
63	Gabon	15,356	54.8	0.393	510,789
64	Brazil	14,539	64.5	0.463	503,583
65	North Macedonia	11,355	66.5	0.347	493,099
66	Costa Rica	13,000	69.7	0.456	492,918
67	Maldives	12,006	67.6	0.400	486,943
68	Tunisia	10,436	65.1	0.377	423,271
69	St. Lucia	11,788	65.7	0.459	419,006
70	Albania	9,927	66.4	0.383	406,705
71	Jordan	9,473	64.7	0.346	400,829
72	Bosnia and Herzegovina	9,720	66.7	0.394	392,885
73	Dominican Republic	11,133	63.8	0.455	387,090
74	China	9,526	67.6	0.430	367,049
75	Colombia	10,791	65.7	0.498	355,903
76	Ukraine	7,824	62.2	0.270	355,279
77	Egypt, Arab Rep.	9,859	60.0	0.404	352,555
78	Peru	9,957	66.3	0.469	350,524
79	Timor-Leste	8,861	57.3	0.314	348,289
80	Ecuador	9,352	66.8	0.443	347,978
81	Paraguay	9,801	64.5	0.463	339,455
82	Jamaica	7,999	66.1	0.409	312,489
83	Mongolia	7,709	60.2	0.333	309,526
84	Sri Lanka	8,530	66.1	0.488	288,668
85	Botswana	13,053	52.9	0.589	283,793
86	Armenia	6,703	65.2	0.360	279,696
87	Indonesia	8,433	60.4	0.456	277,104
88	Georgia	6,982	64.8	0.401	270,989
89	Fiji	7,352	60.4	0.393	269,548
90	Morocco	6,443	63.5	0.398	246,306
91	South Africa	11,888	50.5	0.594	243,736
92	El Salvador	6,301	64.2	0.407	239,874
93	Bhutan	6,420	58.7	0.395	227,985
94	Guatemala	6,714	62.1	0.467	222,230
95	Samoa	5,400	64.4	0.414	203,783
96	Philippines	5,597	60.6	0.416	198,074
97	Tonga	4,984	63.7	0.377	197,798
98	Cabo Verde	5,828	63.1	0.481	190,869
99	Vietnam	4,408	66.5	0.376	182,921
100	Namibia	8,461	53.4	0.600	180,724

Continued

Table A.1: Continued

Rank	Country	pppGDPpc	HALE	Gini	IHLI
101	Bolivia	5,407	61.0	0.453	180,431
102	Congo, Rep.	5,186	54.2	0.422	162,476
103	Pakistan	4,284	56.2	0.342	158,406
104	Yemen, Rep.	4,479	54.3	0.359	155,888
105	Lao PDR	4,219	56.0	0.351	153,327
106	Moldova	3,911	61.3	0.362	152,951
107	Nicaragua	4,029	65.5	0.429	150,676
108	Myanmar	3,721	56.3	0.328	140,787
109	India	4,463	57.4	0.469	136,035
110	Nigeria	5,083	46.2	0.421	135,969
111	Honduras	3,971	65.5	0.498	130,576
112	Mauritania	3,317	54.5	0.358	116,058
113	Micronesia, Fed. Sts.	3,298	60.3	0.417	115,929
114	Vanuatu	2,948	62.0	0.374	114,419
115	Kyrgyz Republic	2,790	61.1	0.341	112,346
116	Sao Tome and Principe	2,642	59.0	0.302	108,783
117	Bangladesh	2,443	60.7	0.339	98,009
118	Ghana	3,059	54.1	0.412	97,321
119	Cambodia	2,523	58.5	0.343	96,967
120	Senegal	2,725	56.1	0.392	92,931
121	Cameroon	2,930	48.3	0.431	80,520
122	Kenya	2,476	55.5	0.442	76,677
123	Zambia	3,279	50.1	0.540	75,574
124	Nepal	1,986	59.2	0.362	75,010
125	Tajikistan	2,106	62.4	0.440	73,603
126	Côte d’Ivoire	2,690	46.2	0.410	73,331
127	Solomon Islands	1,871	60.6	0.419	65,888
128	Comoros	2,426	54.7	0.529	62,507
129	Afghanistan	1,694	51.6	0.302	61,004
130	Benin	1,819	51.7	0.447	51,999
131	Lesotho	2,366	45.7	0.523	51,568
132	Chad	1,925	45.1	0.414	50,880
133	Guinea	1,574	50.0	0.369	49,645
134	Gambia, The	1,644	52.5	0.426	49,549
135	Madagascar	1,386	55.8	0.417	45,078
136	Uganda	1,516	50.4	0.413	44,847
137	Zimbabwe	1,738	46.7	0.453	44,396

Continued

Table A.1: Continued

Rank	Country	pppGDPpc	HALE	Gini	IHLI
138	Burkina Faso	1,423	49.7	0.396	42,728
139	Rwanda	1,368	56.2	0.472	40,597
140	Ethiopia	1,074	54.0	0.324	39,199
141	Guinea-Bissau	1,400	48.8	0.436	38,542
142	Togo	1,226	51.1	0.421	36,285
143	Liberia	1,086	51.6	0.356	36,076
144	Sierra Leone	1,200	43.9	0.383	32,503
145	Malawi	1,033	50.4	0.435	29,415
146	Niger	814	49.4	0.359	25,782
147	Mozambique	918	47.6	0.440	24,464
148	Burundi	726	50.3	0.368	23,082
149	Haiti	1,502	32.3	0.529	22,851

Source: Solt (2019); World Bank (2019), WHO (2019) and own calculations.

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 141–167

Global improvements in Years of Good Life since 1950

***Erich Striessnig*[1,*], *Claudia Reiter*[1] *and Anna Dimitrova*[2]**

Abstract

Human well-being at the national aggregate level is typically measured by GDP per capita, life expectancy or a composite index such as the HDI. A more recent alternative is the Years of Good Life (YoGL) indicator presented by Lutz et al. (2018; 2021). YoGL represents a refinement of life expectancy in which only those person-years in a life table are counted that are spent free from material (1), physical (2) or cognitive limitations (3), while being subjectively perceived as satisfying (4). In this article, we present the reconstruction of YoGL to 1950 for 140 countries. Since life expectancy – as reported by the UN World Population Prospects in five-yearly steps – forms the basis of our reconstruction, the presented dataset is also available on a five-yearly basis. In addition, like life expectancy, YoGL can be flexibly calculated for different sub-populations. Hence, we present separate YoGL estimates for women and men. Due to a lack of data, only the material dimension can be reconstructed based directly on empirical inputs since 1950. The remaining dimensions are modelled based on information from the more recent past.

Keywords: Years of Good Life; well-being indicator; human development; survival; basic needs

1 Introduction

Researchers trying to shed light on the question of whether there have been global improvements in the quality of life (QOL) are typically limited to studying either (1) just the most recent past, as the available data allow for more comprehensive international comparisons; or (2) just a small number of selected countries, mostly

[1]Department of Demography, University of Vienna, Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

[2]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

[*]Correspondence to: Erich Striessnig, erich.striessnig@univie.ac.at

DOI: 10.1553/populationyearbook2021.res1.2

in the developed world; or (3) just a narrow set of often not very sophisticated indicators that can be calculated based on data that have already been collected over longer periods of time. The QOL indicators for which longer time series do exist, and that are available in harmonized form over a large number of countries, include life expectancy and GDP per capita. Both of these indicators have their shortcomings (Stiglitz et al. 2010; Ghislandi et al. 2019), and can give us only limited insight into how QOL has actually changed over time. In addition, neither of the two indicators include information on national differences in subjective well-being, which is not represented well in measures of affluence or longevity (Easterlin 1974; Helliwell et al. 2012).

To overcome these challenges, researchers have recently proposed a new indicator that focuses on the number of Years of Good Life (YoGL) that people can expect to live as members of different populations (Lutz et al. 2021). The YoGL indicator is based on period life tables that are available for most countries of the world starting in 1950 (United Nations 2019). Yet estimates of how many years people can expect to live do not tell us about the quality of those years in terms of human well-being. Thus, to derive YoGL, the person-years lived in different age brackets according to the life table are weighted by the age-specific proportions of people living years of life without limitations in the dimensions of (1) material, (2) physical, (3) cognitive and (4) subjective well-being (SWB). The first three of these dimensions can be assessed based on objective standards, whereas the fourth dimension is, by definition, subjectively assessed. While exclusively objective assessments of QOL fail to address individual differences in preferences (different people have different utility functions), and exclusively subjective assessments overemphasize those personal preferences, the combination of the two approaches stipulates that a life year lived can only be seen as "a good year" if objective criteria are met and – borrowing from Diener's "hallmarks" of the SWB area (Diener 1984) – if the individual confirms that the assessment is based on his or her own subjective set of weights. Since both the subjective and the objective criteria have to be met, there is no trade-off between the different dimensions at the individual level.

In this article, we present a new dataset containing the reconstruction of YoGL to 1950 for 140 countries of the world. Naturally, not all of the dimensions required for the orthodox calculation of YoGL are available from internationally harmonized datasets for that many countries; thus, we had to relax some of the assumptions made in the original definition of YoGL. In addition, we lack information for all four dimensions for the same set of individuals that goes far enough back into the past. Having this kind of information is important to account for possible correlations between the different dimensions, as in many cases, individuals who fail to fulfill the YoGL criteria on objective grounds will also be dissatisfied with their life from their own subjective point of view. Therefore, after outlining the conceptual approach and relating it to the literature, we discuss our methodological approach aimed at circumventing the problem of individual-level correlation in the absence of individual-level data. We then present the results derived from applying that methodology to the available data, and close with a discussion of the new dataset's implications with regard to changes in QOL since 1950.

2 Conceptualization and related work

The YoGL indicator was first introduced by Lutz et al. (2018) to help with the assessment of the progress made in reaching the Sustainable Development Goals. Viewed from a micro-level perspective, YoGL aims to estimate the remaining years of life an individual can expect to live in a "good" state. However, from a macro-level perspective, its explicit purpose is to serve as a basis for judging whether long-term development trajectories are sustainable. Therefore, YoGL focuses on the changing composition of populations with regard to characteristics that constitute a society's human well-being in the aggregate (Lutz et al. 2021).

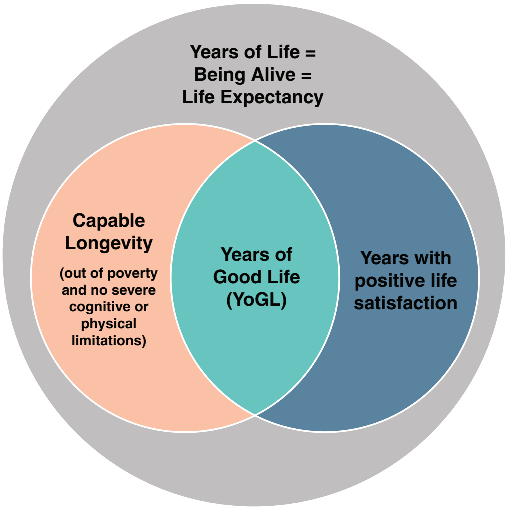
The structure of YoGL is based on different dimensions that are subject to a clear hierarchy, as depicted in Figure 1. First and foremost, survival is considered the essential prerequisite for enjoying any QOL. However, since mere survival cannot sufficiently capture well-being, "good" years of life are counted conditional on meeting minimum standards in both objective and subjective dimensions. Following Desai et al. (1992), the objectively assessable conditions measuring "capable longevity" consist of three sub-dimensions: being out of poverty, being cognitively enabled and having no serious physical disabilities. Life years are considered as good years only if people are above the critical levels in all three of these objective dimensions and in their self-reported overall life satisfaction. This indicator is graphically depicted in Figure 1 as the intersection of capable longevity and years with positive life satisfaction as a subset of the overall years of life. More detailed information about the theoretical foundations of the indicator, as well as a comparison between YoGL and other existing well-being indicators, can be found in Lutz et al. (2018; 2021) and Reiter and Lutz (2019).

When seeking to reconstruct YoGL on a global scale back to 1950, we face problems related to data availability. Internationally harmonized data on subjective well-being are not available before the 1980s, when the World Values Survey was first conducted (Inglehart 1990). Consistent international surveys of people's cognitive and physical skills are not available before the early 2000s, with the start of SHARE (Börsch-Supan et al. 2005) and the WHO's SAGE survey (Kowal et al. 2012). Only the poverty sub-dimension of capable longevity can be reconstructed based on the available economic data.

2.1 Poverty: conceptualization and application

Poverty is a multi-dimensional concept that can be understood in either absolute or relative terms. Absolute poverty is concerned with the deprivation of basic human needs, such as nutrition and shelter. In its strongest form, which is also known as "indigence," survival itself is put at risk (Lok-Dessallien 2000). If a person is not able to afford the minimum nutritional requirements necessary for her daily activities, or if she is lacking access to safe drinking water and sanitation, she would be considered to be indigent or in extreme poverty. The concept of relative

Figure 1:
Dimensions of Years of Good Life indicator of human well-being



Source: Lutz et al. (2021).

poverty, on the other hand, is concerned with the distribution of resources within a society (Atkinson 1975). In contrast to absolute poverty, relative poverty is not determined by a fixed standard of living, but instead compares different segments of a population, such as people at different ends of the income distribution. The usual definition of relative poverty is living on less than 60% of the median income after discounting social transfers (European Commission and Statistical Office of the European Union 2018). Thus, relative poverty can vary greatly from country to country, even after adjusting for purchasing power standards (PPS).

Developments in absolute and relative poverty do not necessarily follow the same trend. Therefore, it is possible for a member of a rich society to suffer from relative poverty, while not being in absolute poverty. Moreover, an increase in relative poverty may be accompanied by a decrease in absolute poverty. For example, when a society becomes wealthier, richer households may benefit more from the economic growth than poorer households. In such a case, more households would end up being above the absolute poverty line, even as the gap between the higher and lower income groups becomes larger. Similarly, changes in socioeconomic conditions can affect the absolute and relative aspects of poverty differently (Lok-Dessallien 2000). For example, if prices increase faster than wages, individuals might fall into absolute poverty without changing their relative position, since everyone's standard of living has declined proportionally. Changes in cultural values or status symbols over time may have similar results (Streeten 1990).

Both absolute and relative poverty are important dimensions to consider when observing social and economic trends and conditions. However, within this paper and in the context of YoGL, we focus exclusively on the dimension of absolute poverty, since our goal is to make cross-country comparisons over time. As Sen has argued (1983, 159), there is "an irreducible absolutist core in the idea of poverty," which is why he identified basic material subsistence as one of three "basic capabilities" that jointly determine a person's freedom to achieve well-being (the other two being basic health and cognitive functioning).

While the concept of absolute poverty should be universally applicable, what constitutes a minimum standard of living can, in practice, vary considerably from place to place and over time, as people's underlying value judgements change over time. Hence, in assessing poverty, another major distinction can be made between objective and subjective approaches to poverty (Siposne Nandori 2014). The objective perspective, also known as the welfare approach, is based on the notion that the conditions under which a person would be considered poor can be defined externally; i.e. independently of the value judgements of the person to whom the conditions apply. The subjective approach is, by contrast, not universally applicable, and emphasizes the individual's own perception of poverty. In other words, individuals decide how poor they are based on their subjective beliefs and perceptions. This notion of poverty depends heavily on how much value a person puts on different goods and services, and has at its core a feeling of deprivation about which intersubjective agreement cannot be reached, rather than an externally-prescribed condition of poverty. This feeling is, of course, strongly affected by other members of society against whom individuals measure themselves; i.e. their reference group (Siposne Nandori 2014).

Despite a growing interest in the measurement of subjective poverty, and in integrating the perspective of the poor into poverty research, objective measures of poverty have so far dominated the development literature, and have shaped most poverty reduction policies. Likewise, in YoGL, we also account for poverty as an objective criterion, while differences in individual preferences are entirely accounted for by the subjective well-being dimension.

3 Data and methods

To reconstruct YoGL for 140 countries since 1950, we rely on various data sources. Of those sources, the most essential are the UN World Population Prospects (2019), which provide information on longevity in the form of data on age- and sex-specific survival for all countries of the world from 1950 onward. As shown in Lutz et al. (2021), one of the strengths of YoGL is that it can be derived for all kinds of sub-groups of populations, such as by level of educational attainment. However, as consistent education-specific longevity information is available only for a small number of countries – e.g. from Eurostat (2020) – we have chosen to exclude this dimension here.

As Figure 1 highlights, longevity is not the only precondition for living a good life. For the reconstruction of YoGL, we need additional information on the proportion of life years that meet certain minimum standards in the aforementioned objective dimensions of being out of absolute poverty and enjoying physical and cognitive health, as well as the subjective dimension of life satisfaction. Only one of these four YoGL dimensions – i.e. being out of absolute poverty – can be reconstructed based on the available economic data. For the other three dimensions, we have to apply additional assumptions in order to provide a historical dataset of YoGL around the world.

3.1 Reconstructing the poverty dimension

As we outlined above, several attempts have been made to reconstruct past poverty prevalence. Since our goal beyond the scope of the present paper in reconstructing YoGL is to study not just its past development, but also, eventually, its implications for future human well-being under different scenarios, we base our reconstruction of the poverty dimension on the most comprehensive source of information available today: i.e. the World Bank's *World Development Indicators* (WDI, World Bank 2017). The WDI provide aggregated information based on household surveys on the proportion of people living below $1.90 a day for a large number of countries. The longest of those national time series derived from household survey data start in 1981. However, earlier reconstruction efforts go back even further. In the following, we will explain how we make use of the reconstruction conducted by Bourguignon and Morrison (2002) to extend our poverty time series back to 1950. By applying this approach, we can make use of the UN time series on longevity in their entirety.

3.1.1 Inequality among world citizens (1820–1992)

Following Roser and Ortiz-Ospina (2013), we combine the WDI data with the information provided by Bourguignon and Morrisson (2002). In their article published in *American Economic Review*, Bourguignon and Morrison reconstructed global wealth inequality for a large number of territories over the past 200 years. However, for the reasons explained above, our interest lies in the information provided on the 1950–1992 period, which is not fully covered by the contemporary WDI estimates. Bourguignon and Morrison's contributions to our work are income distributions (in deciles) reconstructed for different countries and territories from historical national accounts, as well as from the scarce information available on past distributions of wealth and population size. The individual steps that lead us from those reconstructed income distributions to the proportions of the population living in poverty – i.e. those living on less than $1.90 a day – are the following:

1. Apply the income distributions to distribute overall GDP (as reported by Bourguignon and Morrison) across the respective population in any given year

and country. This gives us the annual income of a representative individual in each income decile.
2. Divide the annual income of a representative individual by 365 to derive the daily disposable income of a representative decile member.
3. Apply linear interpolation to determine the proportion of the population within each income decile who are living above or below the poverty line.

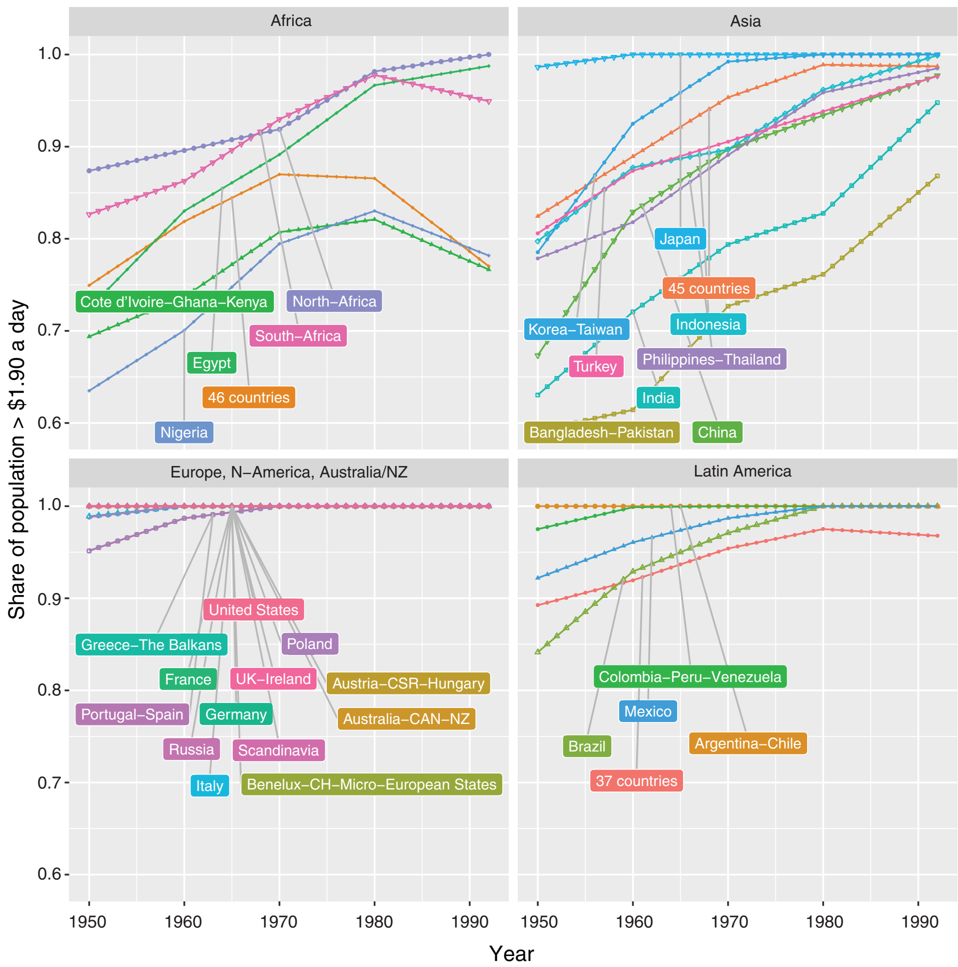
For the population in the higher income deciles, daily disposable income will be far above $1.90 a day; thus, everyone in those income deciles will be out of poverty. However, for the population in the lower income deciles, either everyone in the entire decile (if those in the next-higher income decile also live on less than $1.90 a day), or at least some individuals in the decile (if those in the next-higher income decile already earn more than $1.90 a day), will be living in poverty. By using this approach, we ensure that a disposable income only slightly below the poverty threshold of $1.90 a day leads to a lower proportion of the population in that decile living in poverty than if the income level was far below the threshold.

The results are shown in Figure 2 below, in which countries and territories are grouped by world regions. While vast parts of the populations of Europe and North America had escaped from poverty as early as in 1950, many countries in Africa were still far below that target by 1992. The rapid reduction in poverty among the Asian Tigers is clearly visible, as is the delayed development in South Asia.

Deriving poverty headcount ratios in this way is far from ideal. The limitations that have already been conceded by Bourguignon and Morrison are automatically carried over, especially those pertaining to the poor quality of the historical wealth estimates. Moreover, by using linear interpolation between deciles, we are ignoring potential non-linearities, particularly in the distribution of income within the lower deciles. Still, our method represents the best alternative approach for estimating empirically justified time series that are harmonized over a large number of national territories going back to 1950. In addition, any potential bias induced by our method will be consistent for all territories for which income distributions have been reconstructed by Bourguignon and Morrison. Moreover, going forward from 1992, when the reconstructed time series end, we are not taking the numbers displayed in Figure 2 at face value. Rather, building on the argument of a consistent bias, we rescale these time series to match the more precise WDI headcount ratios in whichever year they first become available for a given country, and apply local polynomial fitting to smooth the transition between the two components of each national series from 1950 to the present. Thus, the reconstruction by Bourguignon and Morrison is only used as an empirical foundation for back-projecting headcount ratios to 1950 based on the more solid evidence from more recent household surveys used by WDI.

Before we can take this additional step, we need to assign the countries of today to the territories that existed in 1820, and for which Bourguignon and Morrison provide information. Not all of these territories still exist as nation states today, and in the case of some formerly unified territories, we can only assume that each of

Figure 2:
Proportion of population out of poverty; i.e. living on more than $1.90 a day. 1950–1992



Source: Own calculations based on Bourguignon and Morrisson (2002).

their separate pieces has emerged out of poverty along the same trajectory. While this assumption might be unrealistic – e.g. in the case of the Hapsburg Empire, as the Austrian and Czech parts industrialized earlier than the Hungarian and Slovak parts of the Empire – by 1950, these differences no longer play a big role in terms of the proportion of the population living below the $1.90 threshold. In addition, this lack of precision is mitigated by rescaling to the WDI headcount ratios, which already differentiate between these different territories.

Previous research has shown that gender and poverty are intrinsically linked (Nieuwenhuis et al. 2018). However, as neither the WDI nor Bourguignon and Morrison provide their results disaggregated by gender (or age), for the reconstruction of YoGL, we assume the same poverty rates for males and females, as well as for different age groups. Hence, the gender differences in YoGL reported in this paper are derived from gender-specific life tables, as well as from gender-specific differences in the reconstruction of other sub-dimensions of YoGL, which are described in more detail in Section 3.2.

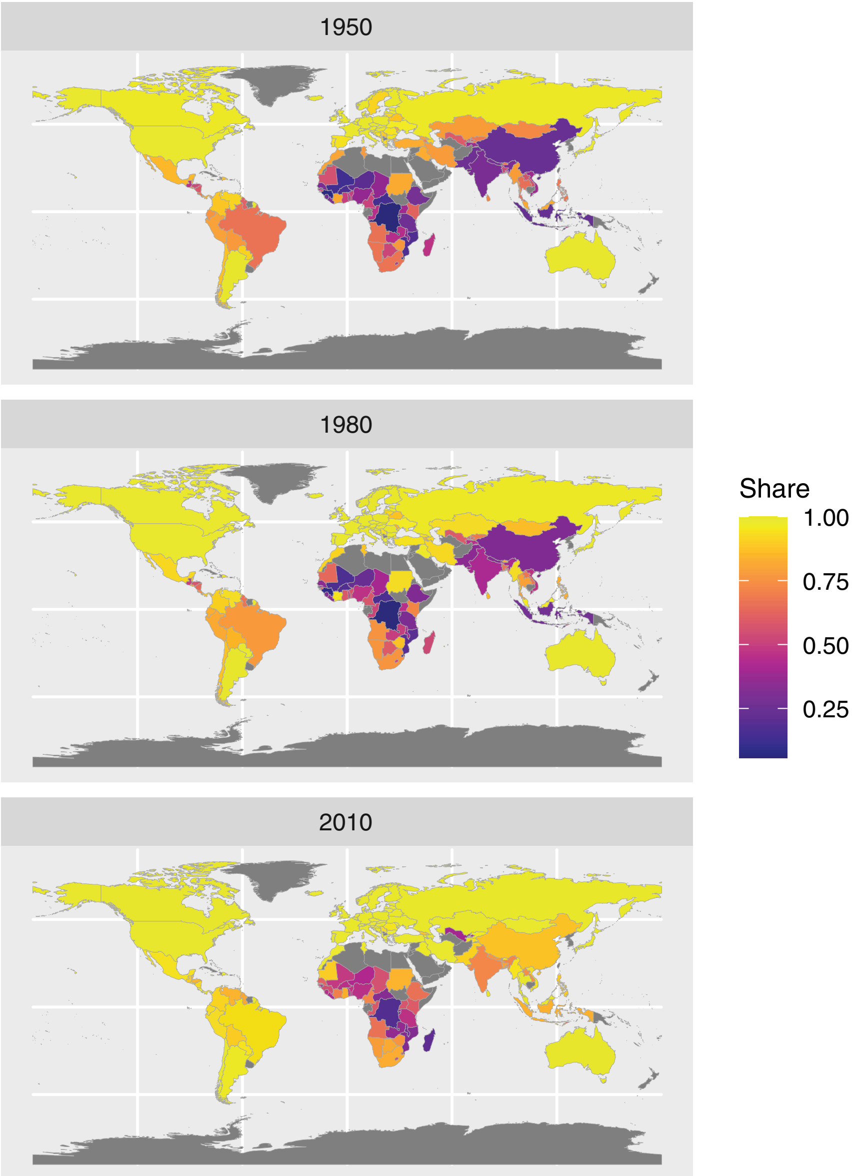
Reconstructed proportions out of poverty for 140 countries in the world are displayed in Figure 3 for 1950, 1980 and 2010. Besides showing strong improvements in people's living conditions over time, our reconstruction points to large regional differences. While much of Eastern Asia was still experiencing abject poverty in the 1950s, poverty had diminished noticeably in this part of the world by 2010. On the other hand, poverty alleviation has not been nearly as successful in most of Sub-Saharan Africa, where widespread poverty has persisted into the 21st century, with strong implications for the number of good life years that are lost due to poverty.

3.2 Reconstructing YoGL

While information about the remaining YoGL dimensions is available for individual countries, it is not available for the vast majority. We therefore resort to reconstructing YoGL based on available survey information from more recent years. In particular, we build our global time series estimates going back to 1950 on existing estimates of YoGL for the 2010–2014 period from 38 countries at various stages of development (Lutz et al. 2021). More specifically, we apply the following procedure to estimate worldwide YoGLs over time:

1. For 38 countries, where data are available for all YoGL dimensions, we estimate YoGL without the poverty dimension. This version of YoGL (henceforth referred to as "YoGL3") is solely based on years lived with positive life satisfaction, with basic cognitive ability and without severe activity limitations. Estimates of YoGL3 provide the basis for the next steps in the procedure.
2. Based on the relationship between the proportion of YoGL3 in life expectancy and total life expectancy at age 20 observed for those 38 countries (see Figure 5), we fit a logistic model to predict YoGL3 for each country and each point in time.
3. In order to consider individual correlations between being out of poverty and the other three YoGL dimensions, we calculate an adjustment factor representing the ratio between the complete YoGL indicator and YoGL3 times the share out-of-poverty for the 38 benchmark countries. As this ratio varies strongly with poverty prevalence (see Figure 6), we estimate the adjustment factor based on the country-specific poverty rates.

Figure 3:
Fitted proportion of population living above $1.90 a day. 1950, 1980, 2010



Source: Own calculations based on Bourguignon and Morrisson (2002) and WDI (World Bank 2017).

4. Finally, YoGL at age 20 is calculated by multiplying the predicted YoGL3 by the estimates of the proportions out-of-poverty displayed in Figure 3, corrected by the adjustment factor to consider individual correlations.

In the following sections, each step will be described in more detail.

3.2.1 YoGL3 for 38 benchmark countries

The first step in our estimation, the calculation of YoGL3 for 38 benchmark countries, follows the same steps as the calculation of YoGL. Relying on survey data, we calculate country-specific proportions of individuals above the three remaining dimensions of YoGL by five-year age groups and gender. YoGL3 is then derived based on demographic life table methods (Sullivan method) in which person-years lived at each age are multiplied by age-specific prevalence rates. For more details on the calculation of YoGL, see Reiter and Lutz (2019), as well as Lutz et al. (2018; 2021).

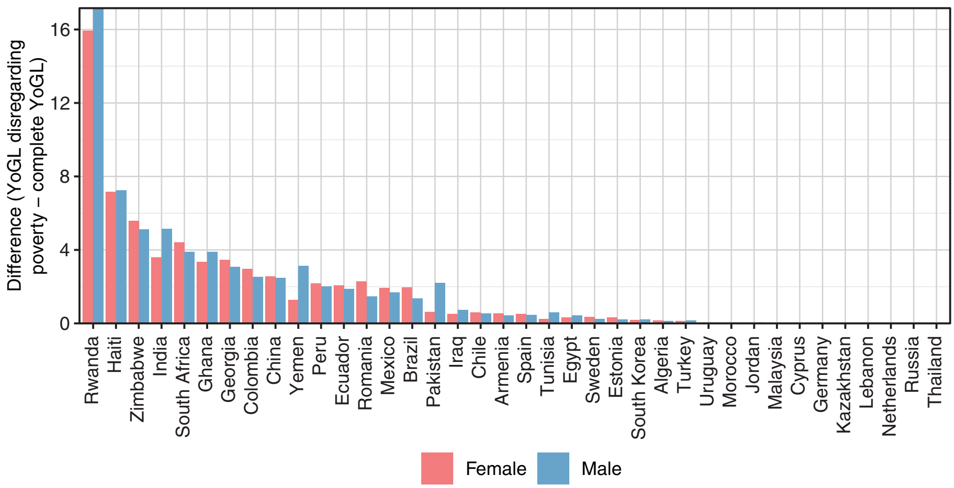
Figure 4 shows the difference between YoGL3 and YoGL at age 20 for males and females in all 38 benchmark countries. Irrespective of gender, the difference is larger in poorer countries, where more good years of life are lost due to a large share of the population living in poverty. In some countries (e.g. Thailand, Russia, Netherlands), the difference is close to zero, which suggests that everyone living in poverty is also below the threshold in at least one additional dimension.

3.2.2 Relationship between YoGL3 (disregarding poverty) and life expectancy

In the next step, we analyze the relationship between YoGL3 and life expectancy. By looking at how the proportion of "good" years (YoGL3/Life expectancy) changes with increasing life expectancy, we can estimate YoGL3 as a function of a country's development stage (proxied by life expectancy). As the proportion of good years is a fraction of total life expectancy, we use a fractional logit model for this purpose. Figure 5 depicts the resulting relationship by gender. In both cases, the fractional logit model fits the 38 observations quite well. In terms of gender differences, it is worth noting that the proportion of good years of total life expectancy at age 20 is consistently lower for females than for males. Based on these estimates, in a next step we predict YoGL3 for all countries and over time.[1]

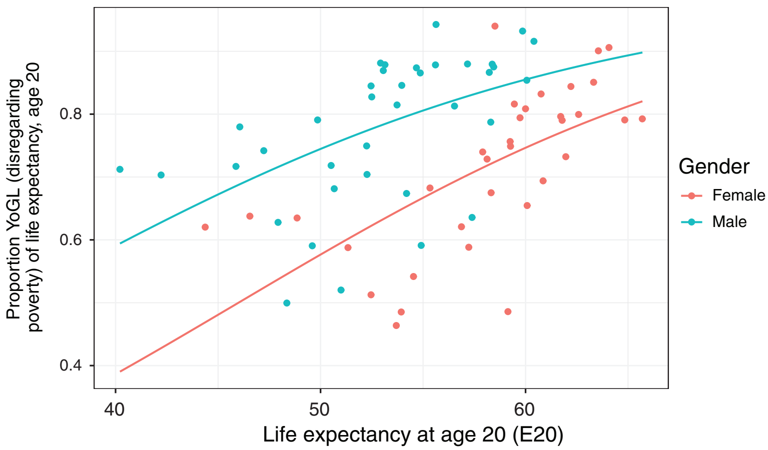
[1] Countries with a proportion of good years below 0.4 were excluded from the analysis. The unusually low YoGLs of these statistical outliers are mostly due to extremely high proportions of the population not passing the threshold in terms of basic numeracy and literacy. This can be explained by inter-cohort differences in access to education – particularly for women – rather than by actual cognitive inability. Including these cases would likely distort the results.

Figure 4:
Difference in Years of Good Life between YoGL3 (disregarding poverty) and complete YoGL at age 20, 2010–14, 38 benchmark countries



Source: Own calculations based on Lutz et al. (2021).

Figure 5:
Gender-specific relationship between the proportion of YoGL3 and overall life expectancy at age 20 in 2010–14 for 38 benchmark countries. Curves fitted based on a fractional logit model



Source: Own calculations based on Lutz et al. (2021).

Our aim of deriving time series estimates from data observed for 38 benchmark countries in the 2010–2014 period no doubt represents a challenge for our methodology. However, given the constraints imposed on us by data availability, basing our reconstruction of YoGL on data from countries that fulfill the data requirements set for YoGL is a reasonable solution. Our reconstructed time series continue to capture the understanding of human well-being enshrined in the original design of YoGL, while accounting for possible correlations between the YoGL sub-dimensions at the individual level. Nevertheless, we are aware that this solution does not fully reflect the impact of period- and country-specific events (e.g. wars, economic crises, etc.) on YoGL dimensions, apart from their impact on life expectancy and material well-being when the available economic data were used. Basing our estimates on countries that were at a similar stage of development (i.e. with a similar life expectancy) as the respective country in the respective year, and assuming that these estimates reflect similar YoGL characteristics, only partly resolves this issue.

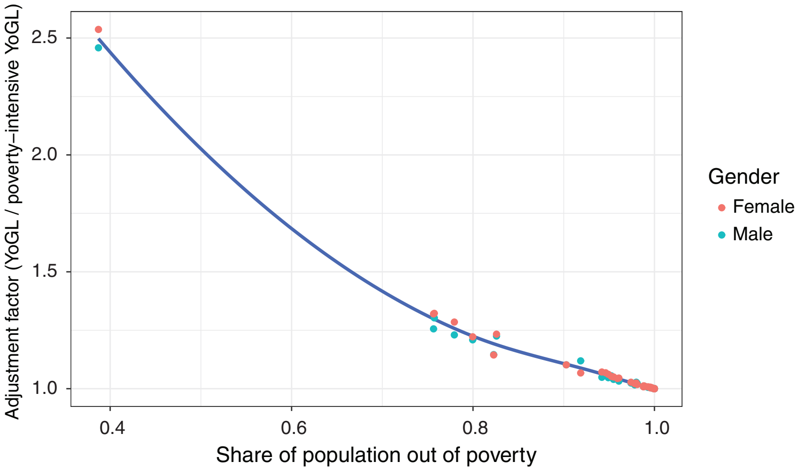
3.2.3 Adjusting for individual correlations

When combining the YoGL3 estimates with the historical data on poverty, we have to make sure that we are not double-counting the loss in good years due to poverty. Many people who fail to meet the out-of-poverty criterion are most likely also below the threshold in at least one of the other three dimensions, and are thus already considered in the calculation of YoGL3. Therefore, we would unnecessarily bias YoGL downward by ignoring this possibility. Ideally, the entire derivation would rely on individual observations of all four YoGL dimensions for one and the same sample of individuals. However, as such historical, internationally harmonized time series do not exist, we have to derive YoGL on the aggregate level.

In order to account for the unobserved individual-level correlation, for each of the 38 benchmark countries where YoGL can be assessed empirically, we first multiply YoGL3 by the proportion of the population out of poverty (henceforth referred to as "poverty-intensive YoGL").[2] As shown in Figure 6, the gender-specific ratio between YoGL and poverty-intensive YoGL (both at age 20) is a declining function of the share of the population that is out of poverty (coefficient of correlation = −0.94). The higher the proportion of people living in poverty, the more likely it is that poverty is the only dimension responsible for a loss in good years in a country. This is also reflected in Figure 4, which shows that the differences between YoGL3 and the complete YoGL are largest in poor countries. Exploiting this strong association, we therefore scale YoGL upward for each country and period using local polynomial regression in order to account for individual correlations between YoGL dimensions.

2 The calculation of "poverty-intensive YoGL" again follows the Sullivan method.

Figure 6:
Relationship between the adjustment factor (YoGL/poverty-intensive YoGL) and the share of the population out of poverty in 2010–14 for 38 benchmark countries as fitted by local polynomial regression



Source: Own calculations based on Lutz et al. (2021).

3.2.4 Estimating YoGL at age 20

The final step in our reconstruction of YoGL at age 20 from 1950 to 2015 for all countries where poverty estimates are available is described by the following formula:

$$YoGL_{c,y}^{20} = E_{c,y}^{20} * \frac{YoGL3_{c,y}^{20}}{E_{c,y}^{20}} * frac\_oop_{c,y} * adj\_factor_{c,y}$$

where $YoGL_{c,y}^{20}$ represents the predicted YoGL at age 20 for a given country c in a given year y; $E_{c,y}^{20}$ denotes life expectancy at age 20 for the same country and year; $YoGL3_{c,y}^{20}/E_{c,y}^{20}$ is the estimated proportion of good years (disregarding poverty) of total life expectancy as a function of $E_{c,y}^{20}$; $frac\_oop_{c,y}$ denotes the estimated country- and time-specific out-of-poverty fraction; and $adj\_factor_{c,y}$ represents the estimated adjustment factor (to account for individual-level correlations) as a function of $frac\_oop_{c,y}$.

Table 1:
Years of Good Life (YoGL) at age 20 for women in the top-five and bottom-five countries in 1950, 1980 and 2010, sorted by the 2010 ranking

	1950	1980	2010
Japan	29.48	45.61	55.67
Spain	33.98	45.37	53.90
France	35.07	44.74	53.55
Switzerland	35.68	45.34	53.11
Italy	35.11	43.78	52.87
Democratic Republic of the Congo	4.14	5.91	19.04
Mozambique	9.47	14.49	18.03
Central African Republic	7.97	14.10	15.14
Eswatini	10.83	17.49	13.67
Lesotho	22.88	30.41	12.74

Source: Own calculations.

4 Results

The result of applying this methodology is a global dataset of YoGL at age 20 for 140 countries from 1950 to 2015[3] in five-year time intervals, which allows us to look at human development from a different angle. In this section, we present our estimates of YoGL, and compare them to other indicators of human well-being: namely, life expectancy and GDP per capita. While life expectancy is available for the majority of countries around the world as far back as 1950, only 53 countries provide sufficient information to enable us to reconstruct time series of GDP per capita this far into the past. This limits our understanding of how human development evolved prior to the 1970s, when consistent information on GDP per capita was becoming available through the Penn World Tables. Being based on life expectancy, YoGL can be reconstructed for a far larger number of countries back to 1950, while covering a much wider range of dimensions of human well-being than GDP per capita. Table 1 shows female YoGL at age 20 in 1950, 1980 and 2010 for the top-five and bottom-five countries, respectively, according to the YoGL ranking in 2010. World maps of female life expectancy and YoGL at age 20, as well as the log of GDP per capita, are shown in the Appendix.

While life expectancy and YoGL are generally correlated, we do find important differences between them. In China, for example, female life expectancy at age 20 rose from 40.6 years in 1950 to 54 years in 1980. However, this increase was far more noticeable in YoGL terms. While Chinese women who were 20 years

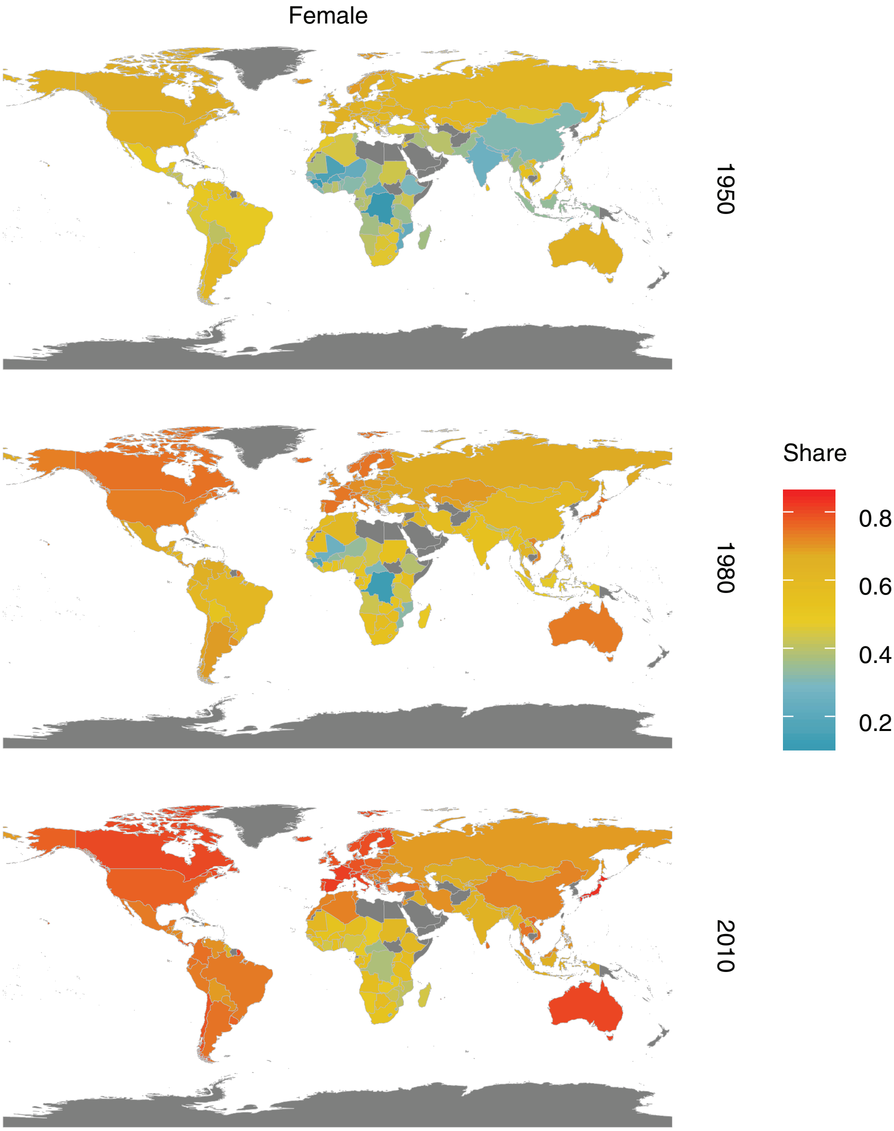
[3] For some countries, poverty rates could not be estimated for the entire period.

old in 1950 could, on average, expect to live another 12.8 good years; by 1980, this figure had increased by a staggering 150% to 32 years, meaning that most of the increase in life expectancy experienced over that period indeed represented an increase in good years. Meanwhile, in India, female life expectancy at age 20 had increased from 35.2 years to 47.7 years. Over the same period, the increase in YoGL (16 years) exceeded the number of years that 20-year-old Indian women could expect to live in 1950 in total (9.5 years). By contrast, in some countries in the Global South, such as in Guinea and the Democratic Republic of the Congo, there were hardly any improvements in YoGL over the first three decades of the observation period, whereas life expectancy was steadily increasing, albeit at a slow pace. This stagnation in YoGL was predominantly due to the poverty dimension. The World Bank data that are available for 1980 onward show that there has been a much steeper decline in the proportion of people living in poverty in those countries since 1980 than the information reconstructed from the Bourguignon and Morrison data would suggest. Hence, assuming that the more recent information provided by the World Bank is more reliable, the initial 1950 poverty rates as reconstructed from Bourguignon and Morrison might be too high, and could lead to the overestimation of YoGL in these countries for the 1950–1980 period.

Another way of looking at this relationship is by mapping YoGL as a share of total life expectancy (Figure 7). Whereas in North America, Western Europe and Australia, up to almost 80% of all remaining life years at age 20 could be expected to be good years as far back as in 1950; in most countries in Sub-Saharan Africa, this level of quality of life still had not been reached by 2010. The lowest value of this share, of only 10%, was recorded for women in the Democratic Republic of Congo in 1950, when the country was still under Belgian colonial rule. By 2010, this value had increased to 38.6%, but it still lagged far behind that in Japan, which was the global leader in terms of YoGL share. Thus, in 2010, 20-year-olds in Japan could expect to spend 84.3% of their remaining life years free from material, physical and cognitive deprivation, while subjectively perceiving these years as satisfying.

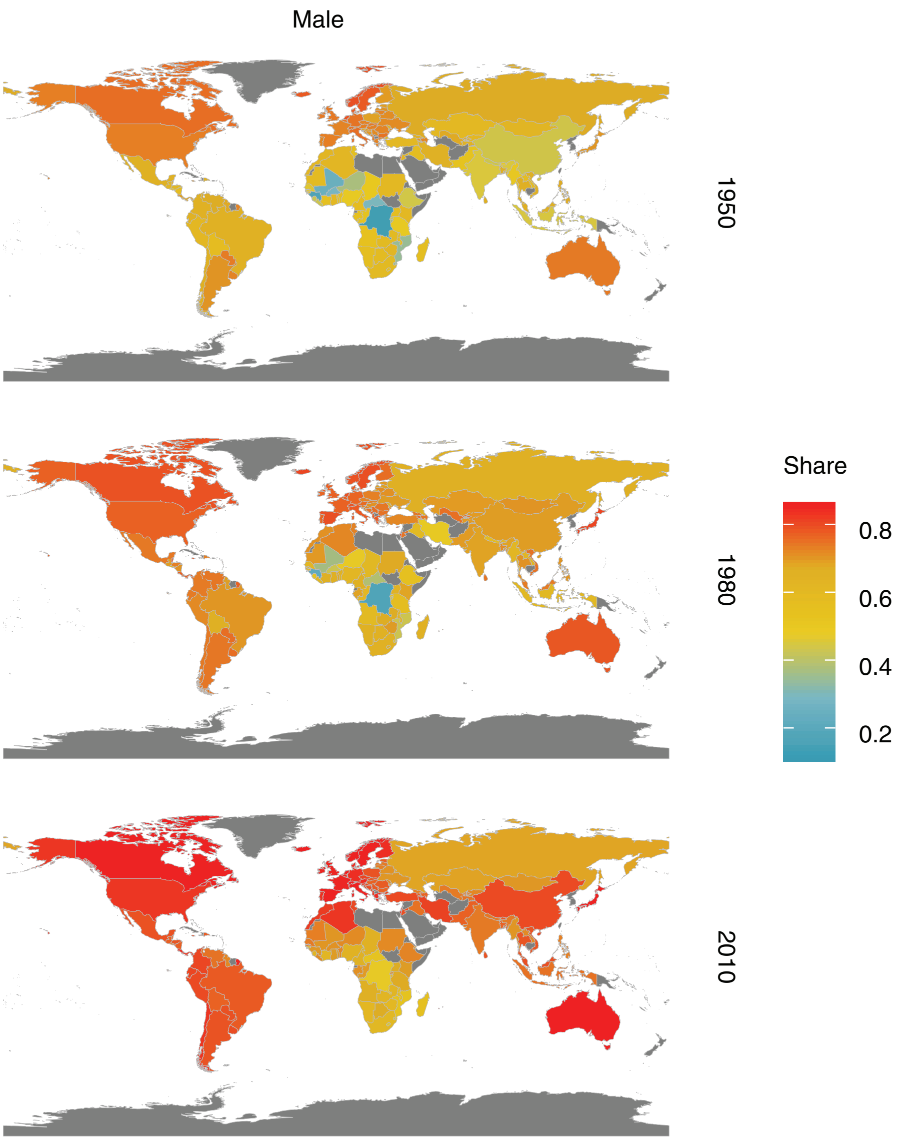
Because it is based on life expectancy, another major advantage of YoGL over GDP per capita is that it can be derived separately for men and women. The gender differences in YoGL, and how they change over time, can shed considerable light on the development process. While both YoGL and life expectancy increased globally over the observation period, and the global gender gap in YoGL (average over all countries) has been decreasing over time, strong gender disparities remain. In India, for example, YoGL at age 20 increased markedly as a share of life expectancy for both sexes. However, the increase for men from 46% in 1950 to 75% in 2010 was exceeded by that for women, as the YoGL gender gap decreased by almost eight percentage points over the same time period, which resulted in 64% of women's remaining life years at age 20 in 2010 being YoGL. Similarly in China, women have been catching up to men in recent decades. Still, as they started from a much lower level, women's share of YoGL in life expectancy at age 20 in 2010 was still 9.2% lower than that of men.

Figure 7:
Share of remaining good years in life expectancy at age 20 by gender, 1950, 1980, 2010



Further telling comparisons can be made by world regions. As shown in Figure 8, which depicts YoGL at age 20 for men vs. women for individual countries grouped into world regions, significant progress has been made in all world regions, not just toward increases in the share of YoGL in life expectancy, but toward increases

Figure 7:
Continued



Source: Own calculations.

in overall YoGL for both sexes. The countries that are still characterized by relatively low YoGL at age 20 in 2010 can be found predominantly in Africa. The disadvantages for females in YoGL are largest in Africa and Southwest Asia. Among the highly developed countries, most of which are in Europe and North America, we

Figure 8:
Male vs. female YoGL at age 20 by world region

Source: Own calculations.

find both smaller differences between YoGL and life expectancy, and slight gender disparities in favour of women. While major gender pay gaps continue to persist in many of these highly developed countries, the female disadvantage in YoGL observed at lower stages of development seems to have reversed in these countries, following the emergence of a slightly positive relationship between life expectancy and the gender gap in YoGL (female minus male YoGL).

Similar results on gender disparities were reported in a separate project on estimating YoGL at age 50 for Europeans (Wittgenstein Centre for Demography and Global Human Capital 2020). While YoGL in 2015 was found to be quite similar for women and men, with a slight advantage for women (consistent with the findings presented in Figure 8); due to higher female life expectancy, women still had fewer “good years” relative to men. These results are related to the male-female health-mortality paradox (di Lego et al. 2019), which describes the phenomenon that women live longer than men, but spend a larger proportion of their life in poorer health (Luy and Minagawa 2014; Pongiglione et al. 2015).

5 Discussion and Outlook

Criticizing conventional indicators of human well-being for how they assess improvements in quality of life over time has become a common refrain in the literature (Stiglitz et al. 2010; Laurent 2017). Yet for lack of better alternatives, particularly when looking into the more distant past, they continue to be widely used. In this article, we present an alternative. The Years of Good Life (YoGL) indicator, as recently introduced by Lutz et al. (2018; 2021), represents the number of years a person can expect to live as part of a specific sub-population without serious limitations in her material, physical, cognitive and subjective well-being. YoGL addresses many of the shortcomings of GDP per capita, life expectancy and the human development index (HDI). But while the merits of this new, demography-based approach to evaluating QOL are described in more detail elsewhere (Lutz et al. 2021), here, we have presented the first attempt to consistently reconstruct gender-specific YoGL for 140 countries of the world going back to 1950.

This time span is, of course, not nearly as long as the recent reconstruction of YoGL for Finland to 1860 by Reiter and Lutz (2019). Yet since no other country in the world provides data ranging back more than 150 years, our reconstruction – based on available international time series on longevity from the World Population Prospects (United Nations 2019), reconstructed poverty headcount ratios as provided by Bourguignon and Morrisson (2002), as well as previous derivations of YoGL for a much smaller group of countries in the more recent past (Lutz et al. 2021) – marks the boldest effort so far to enable country comparisons of QOL going back 70 years.

These reconstructed time series provide a basis for reevaluating past development trajectories with respect to their sustainability. Only if YoGL does not decline for a specific (sub-) population over time – i.e. for subsequent generations – can development be referred to as sustainable. While in terms of GDP per capita, the exploitative extraction of natural resources would invariably indicate an improvement in human well-being, our newly reconstructed time series of YoGL can help us detect the possible negative effects on human well-being that the related losses in biodiversity and environmental quality may have had due to their effects on life expectancy, as well as through the ensuing deterioration in the health dimension of YoGL. Likewise, while GDP per capita neglects distributional aspects, growing inequality would result in lower YoGL due to lower subjective well-being among a society's poorer members. In practice, of course, data quality remains an issue. But our reconstruction of YoGL represents an important first step toward gaining an improved understanding of the extent to which human development in the past evolved sustainably.

Going forward, our intention is to use these reconstructed time series of YoGL as the dependent variable in a well-being production function (Levin and Clark 2010; Clark 2012; Polasky et al. 2015; Irwin et al. 2016). Similar to the conventional growth-accounting framework that is used to explain why some nations surge ahead in terms of GDP while others fall behind, we aim to find the reasons why human well-being is so much higher in some countries than in others. Whereas growth-accounting typically fails to account for the potentially negative feedback effects of economic growth, our reconstructed time series of YoGL allow us to investigate how human well-being has indeed been affected by the collateral damage of economic growth since the second half of the 20th century. Once we have estimated the well-being production function, we can further use it to make predictions about future human well-being under different socioeconomic scenarios; e.g. the Shared Socioeconomic Pathways (SSPs, O'Neill et al. 2017; KC and Lutz 2017).

The major limitations of our reconstruction are mostly associated with a lack of data. In comparison to the indicators necessary for monitoring progress in the full set of Sustainable Development Goals (Maurice 2016), YoGL is an extremely parsimonious alternative. However, despite its rather austere data requirements, sticking to the original YoGL recipe remains a challenge. Internationally consistent efforts to collect data for individual dimensions of YoGL, let alone for all of them at once, have so far hardly been undertaken. Given the strengths of the YoGL indicator, we very much hope that future surveys will include the important dimensions of human well-being needed for its construction. At the same time, our reconstruction was able to rely on the most essential dimension of human well-being, which is being alive; and we were able to provide estimates of the poverty dimension, while accounting for its correlation with the remaining unobserved YoGL dimensions, and thus maintaining the critical relations YoGL aims to account for.

Another important limitation of our analysis lies in the scope and the quality of the data that we were using. For the reconstruction of three out of the four YoGL dimensions, we relied on a relatively narrow set of countries for which data are available from the recent past. In the case of the fourth dimension (being out of poverty), no recent attempts have been made to provide historical time series consistent with the information provided by the World Bank based on household survey data starting in 1980. For a few developing countries that started out with very high poverty prevalence rates in 1950, our reliance on the poverty headcount ratios provided by Bourguignon and Morrisson (2002) led to regime breaks in the pace of development when moving from the older to the more recent and more reliable data sources on poverty. Furthermore, as no information on the age- and gender-specific prevalence of poverty is available, we used country-wide poverty rates for both males and females.

Nonetheless, it is important to note that the idea of human well-being, as embodied by the YoGL indicator, focuses mostly on the tail end of the distribution in each of the four well-being dimensions. Hence, not knowing the precise shape of the complete distribution for each of the YoGL dimensions is less likely to have affected the accuracy of our reconstruction than if the underlying longevity estimates had been inaccurate, since YoGL is primarily a refined version of life expectancy, and the extra dimensions only serve the purpose of identifying which of the total years lived in a population actually qualify to be counted as good years.

In addition, as representative global survey data on the components of YoGL are not available as far back as 1950, our reconstruction also failed to meet the requirement that YoGL should be derived "from the bottom up"; i.e. from individual-level observations. It is this requirement that makes YoGL flexibly applicable, not just to the aggregate country level, but also to specific, well-defined sub-populations for whom life table information does exist, such as men vs. women. However, we provide a reasonable solution for the resulting problem of unobserved individual-level correlations between the different YoGL dimensions by rescaling YoGL based on the empirical relationship between the loss of good years solely due to poverty and the overall share of the population above the poverty line. Given the strength of this relationship, we are confident that we are maintaining the essential characteristics of YoGL.

References

Atkinson, A. B. 1975. *The economics of inequality*. Clarendon Press.

Börsch-Supan, A., K. Hank and H. Jürges 2005. A New Comprehensive and International View on Ageing: Introducing the 'Survey of Health, Ageing and Retirement in Europe.' *European Journal of Ageing* 2(4): 245–53. https://doi.org/10.1007/s10433-005-0014-9

Bourguignon, F. and C. Morrisson 2002. Inequality among World Citizens: 1820-1992. *The American Economic Review* 92(4): 727–44. https://doi.org/10.1257/00028280260344443

Clark, W. C. 2012. Sustainability Science: Its Origins, Evolution and Challenges. Presented at the IIASA 40th Anniversary Conference, Vienna, Austria, 25 October 2012. http://conference2012.iiasa.ac.at/person.html?code=clark&scode=s2s7

Desai, M., A. Sen and J. Boltvinik 1992. *Social progress index: A proposal*. Bogota: United Nations Development Programme.

Diener, E. 1984. Subjective well-being. *Psychological Bulletin* 95(3): 542–75. https://doi.org/10.1037/0033-2909.95.3.542

di Lego, V., P. Lazarevič and M. Luy 2019. The male-female health-mortality paradox. In *Encyclopedia of Gerontology and Population Aging*, eds D. Gu and M. E. Dupre, 1–8. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-69892-2_798-2

Easterlin, R. A. 1974. Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In *Nations and Households in Economic Growth*, eds P. A. David, M. W. Reder, 89–125. Elsevier. https://doi.org/10.1016/B978-0-12-205050-3.50008-7

European Commission, and Statistical Office of the European Union 2018. *Living Conditions in Europe: 2018 Edition*. https://doi.org/10.2785/39876

Eurostat 2020. Life Expectancy by Age, Sex and Educational Attainment Level. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_mlexpecedu&lang=en

Ghislandi, S., W. C. Sanderson and S. Scherbov 2019. A simple measure of human development: The human life indicator. *Population and Development Review* 45(1): 219–33. https://doi.org/10.1111/padr.12205

Helliwell, J. F., R. Layard and J. Sachs (eds) 2012. *World Happiness Report*. The Earth Institute - Columbia. https://worldhappiness.report/ed/2012/

Inglehart, R. 1990. *World Values Survey, 1981-1983*. Inter-university Consortium for Political and Social Research.

Irwin, E. G., S. Gopalakrishnan and A. Randall 2016. Welfare, wealth, and sustainability. *Annual Review of Resource Economics* 8(1): 77–98. https://doi.org/10.1146/annurev-resource-100815-095351

KC, S. and W. Lutz 2017. The human core of the shared socioeconomic pathways: population scenarios by age, sex and level of education for all countries to 2100. *Global Environmental Change* 42(January): 181–92. https://doi.org/10.1016/j.gloenvcha.2014.06.004

Kowal, P., S. Chatterji, N. Naidoo, R. Biritwum, Fan Wu, R. Lopez Ridaura, T. Maximova, et al. 2012. Data resource profile: the world health organization Study on Global AGEing and Adult Health (SAGE). *International Journal of Epidemiology* 41(6): 1639–49. https://doi.org/10.1093/ije/dys210

Laurent, É. 2017. *Measuring tomorrow: Accounting for well-being, resilience, and sustainability in the twenty-first century*. Princeton University Press.

Levin, S. A. and W. C. Clark 2010. Toward a science of sustainability. CID Working Papers 196. Center for International Development at Harvard University. https://ideas.repec.org/p/cid/wpfacu/196.html

Lok-Dessallien, R. 2000. *Review of Poverty Concepts and Indicators*. Chicago: Academic Press. https://pdfs.semanticscholar.org/a358/eb2139bf8c50b338863d0ecb63d4c6dedb21.pdf

Lutz, W., A. A. Lijadi, E. Striessnig, A. Dimitrova and M. Caldeira Brant de Souza Lima 2018. Years of Good Life (YoGL): A New Indicator for Assessing Sustainable Progress. IIASA Working Paper WP-18-007. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://pure.iiasa.ac.at/id/eprint/15402/

Lutz, W., E. Striessnig, A. Dimitrova, S. Ghislandi, A. A. Lijadi, C. Reiter, S. Spitzer and D. Yildiz 2021. Years of Good Life (YoGL) is a Wellbeing Indicator Designed to

Serve Research on Sustainability. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1907351118

Luy, M. and Y. Minagawa 2014. Gender gaps–life expectancy and proportion of life in poor health. *Health Reports* 25(12): 12–19.

Maurice, J. 2016. Measuring progress towards the SDGs—a new vital science. *The Lancet* 388(10053): 1455–58. https://doi.org/10.1016/S0140-6736(16)31791-3

Nieuwenhuis, R., T. Munzi, J. Neugschwender, H. Omar and F. Palmisano 2018. Gender equality and poverty are intrinsically linked: A contribution to the continued monitoring of selected sustainable development goals. United Nations entity for gender equality and the empowerment of women. https://www.unwomen.org/digital-library/publications/2018/12/discussion-paper-gender-equality-and-poverty-are-intrinsically-linked

O'Neill, B. C., E. Kriegler, K. L. Ebi, E. Kemp-Benedict, K. Riahi, D. S. Rothman, B. J. van Ruijven, D. P. van Vuuren, J. Birkmann, K. Kok, M. Levy and W. Solecki 2017. The roads ahead: narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change* 42(January): 169–80. https://doi.org/10.1016/j.gloenvcha.2015.01.004

Polasky, S., B. Bryant, P. Hawthorne, J. Johnson, B. Keeler and D. Pennington 2015. Inclusive wealth as a metric of sustainable development. *Annual Review of Environment and Resources* 40(1): 445–66. https://doi.org/10.1146/annurev-environ-101813-013253

Pongiglione, B., B. L. De Stavola and G. B. Ploubidis 2015. A systematic literature review of studies analyzing inequalities in health expectancy among the older population. *PLoS ONE* 10(6): e0130747. https://doi.org/10.1371/journal.pone.0130747

Reiter, C. and W. Lutz 2019. Survival and years of good life in Finland in the very long run. *Finnish Yearbook of Population Research* 54: 1–27. https://doi.org/10.23979/fypr.87148

Roser, M. and E. Ortiz-Ospina 2013. Global extreme poverty. *Our World in Data*, May. https://ourworldindata.org/extreme-poverty

Sen, A. 1983. Poor, relatively speaking. *Oxford Economic Papers* 35(2): 153–69. https://doi.org/10.1093/oxfordjournals.oep.a041587

Siposne Nandori, E. 2014. Objective and subjective poverty. In *Encyclopedia of quality of life and well-being research*, ed A. C. Michalos, 4424–28. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_3779

Stiglitz, J. E., J. P. Fitoussi and A. Sen 2010. *Mismeasuring our lives: Why GDP doesn't add up*. New York: The New Press. https://thenewpress.com/books/mismeasuring-our-lives

Streeten, P. 1990. Poverty concepts and measurement. *The Bangladesh Development Studies* 18(3): 1–18. https://www.jstor.org/stable/41968775

United Nations 2019. World Population Prospects: The 2019 Revision. New York, NY: Department of Economic and Social Affairs, Population Division. http://esa.un.org/unpd/wpp/

Wittgenstein Centre for Demography and Global Human Capital. 2020. *European Demographic Data Sheet 2020*. Vienna: Wittgenstein Centre (IIASA, OeAW, University of Vienna). http://www.populationeurope.org/en/download/EDS2020_POSTER.pdf

World Bank. 2017. World Development Indicators. Washington D.C.: The World Bank. https://data.worldbank.org/products/wdi

Appendix

Figure A1(a):
World maps of female life expectancy at age 20 (a), female YoGL at age 20 (b) and the log of GDP per capita (c) in 1950, 1980 and 2010.

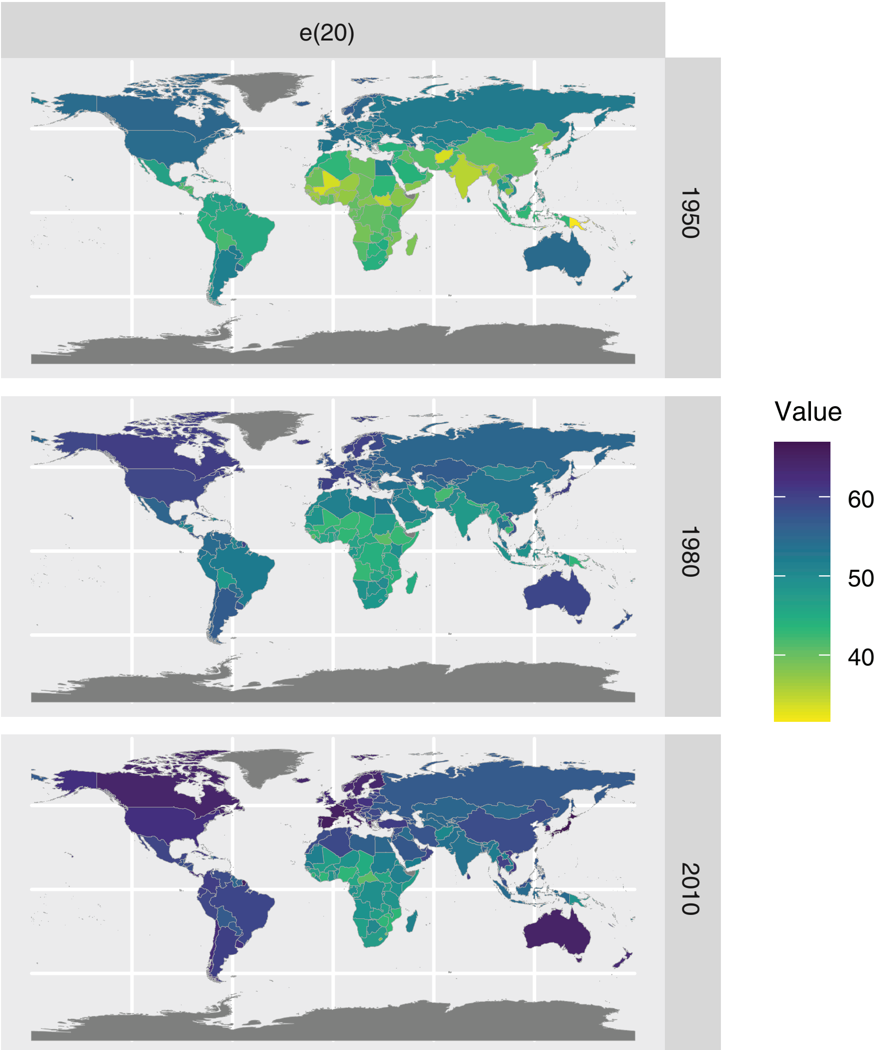


Figure A1(b):

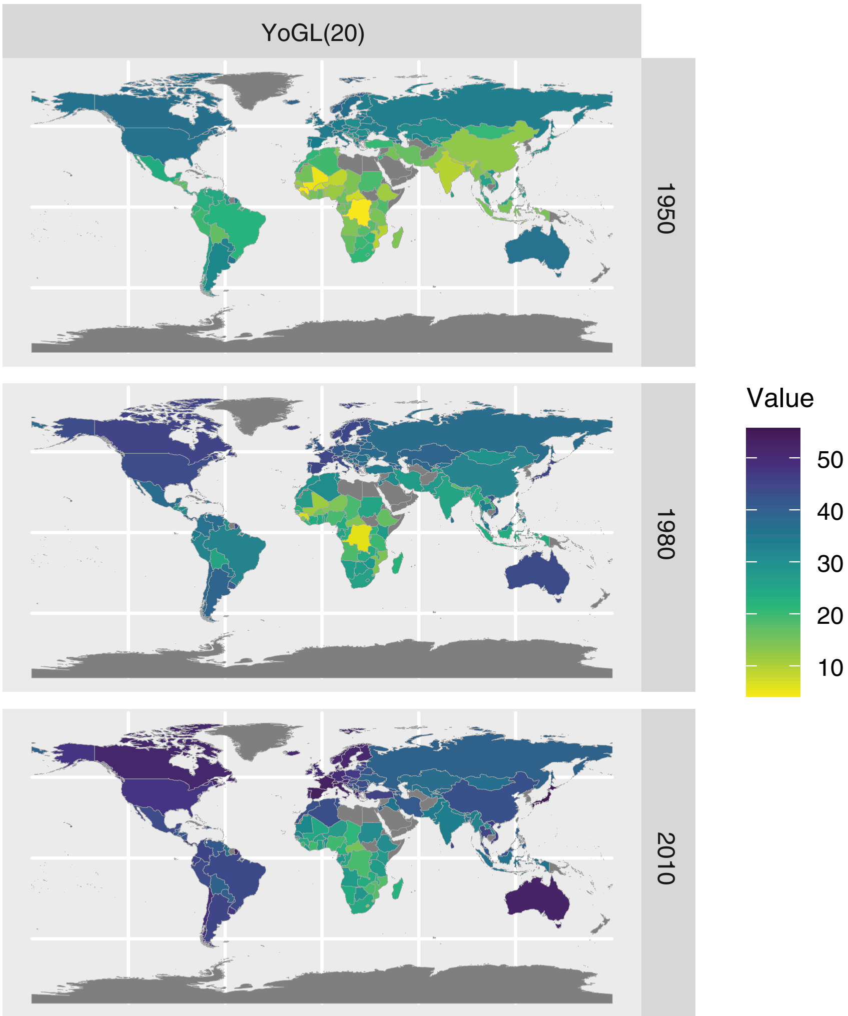
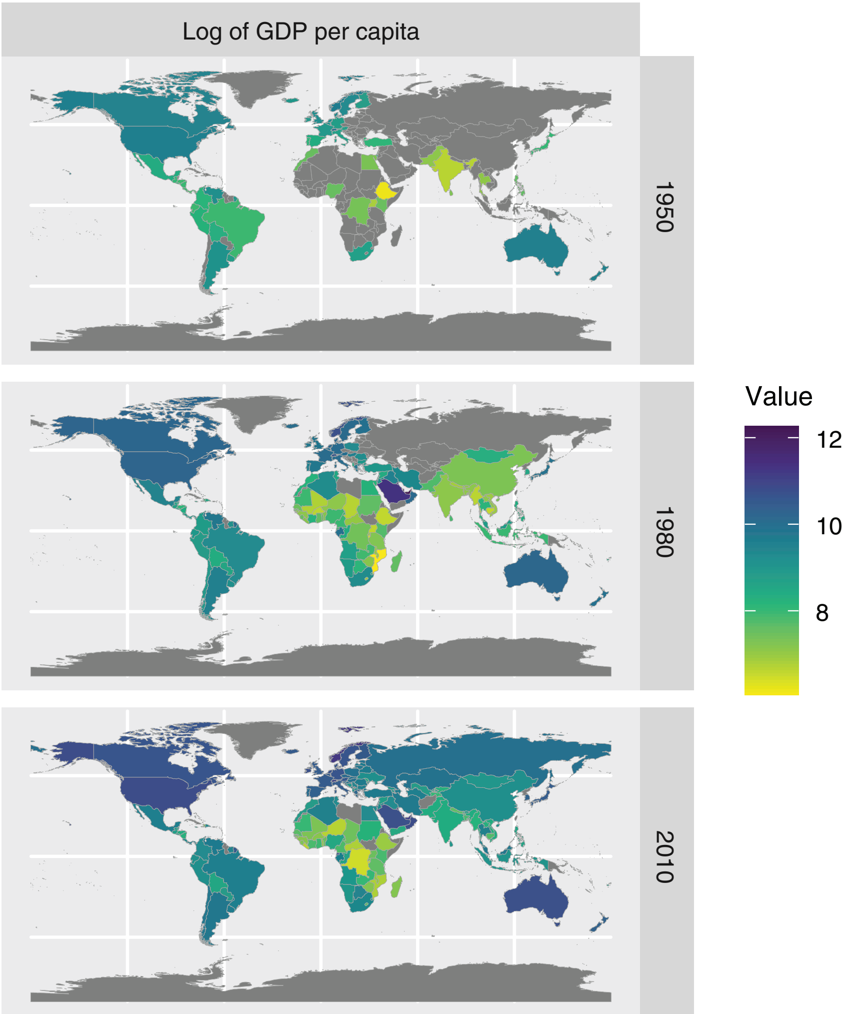


Figure A1(c):



Source: UN World Population Prospects, Penn World Tables 9.1, and own calculations.

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 169–190

Mortality evolution in Algeria: What can we learn about data quality?

Farid Flici[1,*] and Nacer-Eddine Hammouda[2]

Abstract

Mortality in Algeria has declined significantly since the country declared its independence in 1962. This trend has been accompanied by improvements in data quality and changes in estimation methodology, both of which are scarcely documented, and may distort the natural evolution of mortality as reported in official statistics. In this paper, our aim is to detect these methodological and data quality changes by means of the visual inspection of mortality surfaces, which represent the evolution of mortality rates, mortality improvement rates and the male-female mortality ratio over age and time. Data quality problems are clearly visible during the 1977–1982 period. The quality of mortality data has improved after 1983, and even further since the population census of 1998, which coincided with the end of the civil war. Additional inexplicable patterns have also been detected, such as a changing mortality age pattern during the period before 1983, and a changing pattern of excess female mortality at reproductive ages, which suddenly appears in 1983 and disappears in 1992.

Keywords: mortality; Algeria; Lexis map; data quality; methodological change; vital statistics

1 Introduction

Mortality analysis remains a common approach to assessing the efficiency of public health programs (Purdy et al. 2013) and related improvements in the living conditions of a given population (Elo and Preston 1992). The quality and validity of such analyses relies to a large extent on the quality of the underlying data, which is often far from perfect in the context of developing countries. Thus, whenever

[1]Research Center in Applied Economics for Development (CREAD), Algiers, Algeria.

[2]National Committee for Population, Health Ministry, Algiers, Algeria.

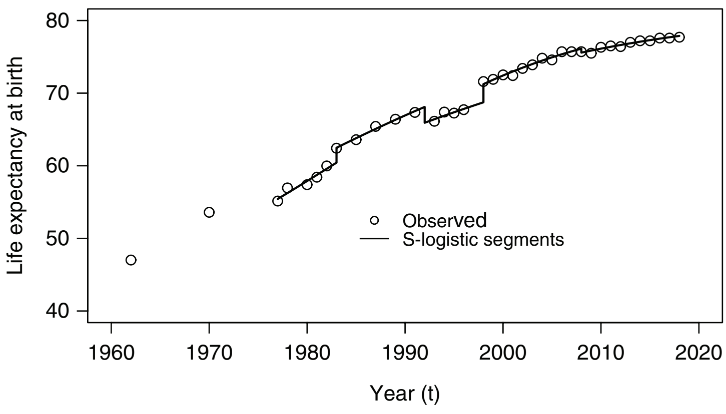
[*]Correspondence to: Farid Flici, farid.flici@cread.dz

DOI: 10.1553/populationyearbook2021.res1.3

possible, it is essential to assess, evaluate and correct these data, as well as estimates based on these data. Assessing and correcting historical mortality estimates is, however, very complex when the crude data used to estimate mortality indicators are not provided, or when the methods used to correct and adjust crude counts are not well known. In such cases, the regularity of the time series of mortality indicators can be used to detect data quality or methodological issues (Rey et al. 2011). In the absence of extreme events like wars, epidemics or geopolitical restructuring, the evolution of different mortality indicators over time, such as life expectancy, age-specific mortality rates, age patterns of mortality improvement and male-female mortality ratios, is expected to happen in a gradual and smooth fashion (Vékás 2020). Sudden changes in these indicators are often due to changes in data processing (Börger et al. 2018), or are related to data quality issues.

Our aim in this paper is to assess historical mortality estimates for Algeria, where life expectancy at birth has increased from 47 years in 1962 to 77.7 years in 2018. While there is a large body of literature on the evolution of mortality in Algeria (Salhi 1984; Daoudi 2001; Hamza-Cherif 2011), most of these studies did not analyse the effects of methodological changes or data quality on the estimated outcomes. One exception is Flici (2020), who performed a change-point analysis on the evolution of life expectancy at birth for the Algerian population from 1977 to 2018, both sexes combined (Figure 1). The aim of the change-point analysis was to detect significant changes in the general life expectancy level, and its improvement rate. The analysis concluded by splitting the series of life expectancy at birth, initially assumed to be of an S-logistic shape, into S-logistic segments separated by four change-points corresponding to 1983, 1992, 1998 and 2008. Except for a

Figure 1:
Evolution of life expectancy at birth (Algeria: 1962–2018) fitted using S-logistic segments



Source: Flici (2020).

steep decline in survival due to the civil war between 1992 and 1998, all of the change-points could be linked to methodological modifications in the estimation of life expectancy. For example, one change-point in 1983 was attributable to an update of the correction factors for death registration incompleteness, and, ultimately, to a change in mortality adjustment methods. Similarly, a change-point in 2008 was due to an update of the population structure based on the census data collected in the same year. Even the observed decline in survival related to the civil war was amplified by changes in the data and methods used for the estimations.

While change-points analyses like that of Flici (2020) can be useful for detecting points in time when general mortality increases (or decreases) significantly due to either mortality shocks or methodological changes, this method is not able to detect all of the data quality problems in the Algerian context, as some significant changes could still be hidden in the mortality age pattern even after a change-point analysis of life expectancy trend has been conducted. In this paper, we seek to fill this gap by investigating the changes that occurred in the mortality age pattern of the Algerian population from 1977 to 2018. Thus, our study is intended to complement the analysis by Flici (2020). In particular, we focus on the break points that have already been defined (i.e., 1983, 1992, 1998 and 2008), and on the inter-break-points periods. We propose the use of data visualisation techniques, not only to study natural variations of mortality, but also, and more importantly, to detect changes that may reflect methodological changes or data quality issues. To this end, we visually inspect a variety of indicators in order to detect any unexpected, irregular or inexplicable patterns over age or over time: namely, the mortality age pattern, the age pattern of mortality variation rates and the male-female mortality ratios.

Shedding light on the unexpected and inexplicable patterns of the different mortality indicators may help to correct the historical evolution of mortality indicators as reported in official statistics, and to better separate the natural evolution of these indicators from the effects of methodological and data quality changes. Our aim is to make evaluating public health policies using mortality indicators more relevant and more pertinent. In addition, we expect that improving the quality and the stability of historical data will improve the robustness of mortality forecasts.

In Section 2 of this paper, we provide a short literature review of analyses of mortality patterns using data visualisation techniques, and we briefly discuss the potential sources of bias when estimating life tables using incomplete data. Next, we describe the Algerian mortality data, and we present the data visualisation methods we use to highlight the methodological changes and the data quality imperfections in the Algerian context. We then present our results in Section 3, and discuss our findings in Section 4.

2 Material and methods

2.1 Analysing mortality patterns using data visualisation techniques

Since the emergence of demography as a science, graphical representations have attracted considerable interest as tools of analysis. This interest further expanded in the second half of the 19th century (Caselli and Enzo 1990), which led to the development of the well-known concept of the 'Lexis diagram' (Lexis 1875), a two-axes plane that crosses time (jointly with the year of birth) and age, and that represents the survivorship and deaths of a group of individuals (Keiding 1990). Arthur and Vaupel (1984) used the term 'Lexis surface' to describe a two-axes plane (age vs. time) representing demographic rates or counts (Vaupel et al. 1987), which are also called a 'demographic surface' or a 'population surface' (Vaupel et al. 1985). The concept of a 'mortality surface' was used by Vaupel et al. (1985) to specify mortality-related measures. Gambill and Vaupel (1985) and Gambill et al. (1985) also employed an improved version of the 'contours map' from Delaporte (1942) to visualise different mortality and demographic measures.

As well as focusing on visualising mortality indicators such as death counts, life expectancy by age, changes in life expectancy by age and cause-specific mortality rates, Vaupel et al. (1987) extended the techniques to visualise other demographic measures, such as population distributions, fertility rates and the ratio of males to females in a population. Likewise, Wilmoth (1985) used Lexis maps to separate age, period and cohort effects on mortality, while Caselli et al. (1985) used the same technique to analyse mortality in Italy. Moreover, Gambill et al. (1986b) used this technique to analyse cause-specific mortality in Japan.

In addition to representing a pertinent tool for mortality data visualisation (see, for example, Aburto and van Raalte 2018; Remund et al. 2018; Rau et al. 2018), Lexis maps have, in recent years, regained their place as a tool for mortality analysis (Minton et al. 2013; Minton 2014; Vanderbloemen et al. 2016; Minton et al. 2017), with some enhancements being introduced mainly through computer software developments. For example, Minton et al. (2017) compared excess mortality in Scotland with that in the UK as a whole and in other Western European countries based on a visual inspection of the Lexis map representing the mortality ratios between the populations for the period from 1950 to 2010. Their aim was to detect through visual inspection unexpected patterns or patterns showing age, time or cohort dependence.

In this paper, we use a similar approach. We start by visualising the Lexis map of mortality rates in logarithm form for the 0, 1–4 and 5–9 age groups, and then by five-age intervals until the 74–79 age group, for the period from 1977 to 2018, for males and females combined. We then analyse the Lexis maps of the mortality variation rates (mortality variation surface) and of the male-female mortality ratios (sex ratio mortality surface). These analyses are complemented with numerical analyses and simple plots when needed. All of the Lexis map type visualisations presented in this paper have been created using the R-package ‘plot3D’.

2.2 Potential sources of bias when estimating life tables based on crude data

In this subsection, we describe the various steps of the process, starting with collecting crude data from the civil registration system to estimating national life tables. We also discuss the shortcomings in this process that have the potential to bias the final estimates. In particular, irregularities can stem from death counts and their corrections, population data and related adjustments, and model life tables (MLTs).

The life tables are calculated based on the death counts and the population numbers distributed by age and sex (United Nations 1961). These death counts are frequently drawn from the crude numbers reported to the civil registration system, and are corrected for incompleteness when death registration is not complete. Such corrections can usually be made using either survey-based estimations (Moriyama 1990) or indirect methods (Brass 1975; Preston and Hill 1980; Preston et al. 1980; Adair and Lopez 2018).

As is the case in many developing countries, in Algeria, the completeness of death registration has improved significantly since the 1970s, but has historically been low. According to estimates of the Multi Rounds Demographic Survey (MRDS-1970), death registration completeness in Algeria was 60.6% in 1970 (Daoudi 2001), had increased to nearly 64% by 1977, and remained at approximately that level in 1978 and 1980 (Daoudi 2001; Hamza-Cherif 2011). The Workforce and Demography Survey of 1982 provided a survey-based estimation of death registration completeness for 1981 of around 81% (Daoudi 2001). If we assume that the estimates for 1981 are more reliable than those for the 1977–1980 period, then the sudden jump observed in 1981 may be attributed to the underestimation of death registration completeness for the previous period (Daoudi 2001). This pattern should, in turn, lead to an underestimation of life expectancy for the same period that should be offset when the new correction factors have been introduced. However, the death registration completeness rates for 1981 were used by the Office of National Statistics (ONS) to correct the crude death counts in response to the death registration incompleteness only from 1983 onwards (Flici 2020). The same correction factors were used continuously, without any updates, until 2002, when the Algerian Family Health Survey was conducted (ONS 2003). The new correction factors have been applied with retroactive effect from 1998 onwards, and are still being used to adjust registered deaths (ONS 2020), without being published.

In contrast to deaths, which are continuously recorded by the civil registration system, population censuses are usually taken every 10 years. In Algeria, five previous population censuses were conducted in 1966, 1977, 1987, 1998 and 2008. A sixth census is currently being performed, and the results of this census are expected to be available by the end of 2021. The population age structure during the intercensal periods is usually estimated by making projections based on past trends (ONS 2005; 2008); based on vital statistics, while assuming zero net migration (ONS 2003; 2020); or simply by assuming that the population age structure remains constant between censuses (ONS 2010). Usually, gaps arise when old estimates are compared to those provided by new censuses, which can lead to some irregularities and jumps in mortality indicators over time (Flici 2020).

In Algeria, as in many other developing countries, another fundamental step involved in constructing national life tables is using regional model life tables (MLTs) to adjust the curve of age-specific mortality when it is irregular, or does not follow a conventional shape (Mouffok 1984). Several MLT systems have been developed for this purpose, such as the United Nations MLTs (United Nations 1982) and the MLTs of Princeton (Coale and Demeny 1966; Coale et al. 1983; Coale and Guo 1989). In addition to the effect on life expectancy estimates of the choice of the MLT (McCaa 1978), a change in the choice of the MLT system – or in the group of MLTs within the same system – used for adjustment can lead to inconsistencies in the life expectancy time series. This kind of effect was observed in Algeria in 1983, when there was a sudden jump that was mainly attributable to a change in the MLT used, as reported by Flici (2020).

2.3 Data

To assess the irregularities related to data quality and estimation methods in mortality estimates for Algeria, we use the national life tables published annually by the ONS from 1977 to 2018. The abridged life tables are published for males, females and the total population at ages zero, 1–4 years and then by five-age intervals until the open age group, which has been varying irregularly from 70 years and older to 85 years and older. From 1998 onwards, these national life tables have been published annually. Before 1998, there were no published life tables for some years; namely, for 1979, 1984, 1986, 1988, 1990, 1992 and 1997.

2.4 Estimating and visualising mortality estimates

For our analysis, a complete mortality surface is required, or the missing calendar years life tables need to be estimated. In addition, based on a common closure age of 80 years for the whole mortality surface, the life tables closed out at younger ages need to be extended up to 80 years. To this end, we propose extending the life tables closed out earlier than at age 80 using the Gompertz model (Gompertz 1825). Consequently, we estimate the mortality rates for the 70–74 and 75–79 age groups for the years 1983, 1985 and 1987, and the mortality rates for the 75–79 age group for the 1993–1996 period. Then, we estimate the missing life tables for 1979, 1984, 1986, 1988, 1990, 1992 and 1997 using linear interpolation. It is important to note that the life table for 1998 was first published in 1999, and was revised in 2002 using the death completeness rates estimated in 2002, combined with the population age structure from the population census of 1998. Thus, there are two versions of the life table corresponding to 1998. In order to estimate the life table for 1997 using linear interpolation between 1996 and 1998, it is advisable to use the old life table for 1998, rather than the revised life table.

While analysing mortality surfaces is supposed to provide a general overview of mortality improvements, especially during the periods when mortality shocks occurred, some effects may be hidden from the reader. Hence, it is necessary to complement the analysis using other indicators. Visualising the Lexis diagram of the mortality variation rates with the age groups on the vertical axis and the time on the horizontal axis is assumed to highlight the ages for which drastic changes were recorded during the studied periods, and the time periods for which unexpected changes were recorded. We consider ${}_nQ_{x,t}$ the mortality rate of the age group $[x, x+n[$ during the year t. The age x, in our case, is equal to $0, 1, 5$, and then to a step of five years. The year t goes from 1977 to 2018. A mortality variation rate is defined by

$$ {}_nV_{x,t} = \frac{{}_nQ_{x,t}}{{}_nQ_{x,t-1}} - 1. $$

We note that the mortality variations rates are estimated for males and females combined.

The colour scale of the resulting mortality variation surface is defined in order to distinguish different degrees of variation: a slight reduction ($-5\% < {}_nV_{x,t} < 0\%$), a moderate reduction ($-10\% < {}_nV_{x,t} \leq -5\%$), a sharp reduction (${}_nV_{x,t} \leq -10\%$), stagnation or a slight increase ($0\% \leq {}_nV_{x,t} < 10\%$), a significant increase ($10\% \leq {}_nV_{x,t} < 20\%$), a sharp increase ($20\% \leq {}_nV_{x,t} < 40\%$) and, finally, a drastic increase (${}_nV_{x,t} \geq 40\%$). In addition, contours are plotted around variation rates of $+10\%$ and -10% to better visualise large variations.

In order to continue with the mortality variation analysis, and considering the change-points already defined by Flici (2020), we visualise the age pattern of mortality variation at the change-points compared to during the inter-change-points periods. Using simple plots, this step aims to detect changes in the age pattern of mortality variation from one period to another, and to detect the ages that benefited the most from mortality reduction over time. In addition, this step seeks to visualise the effects of mortality shocks and methodological changes on the different age groups.

Another way to complement the mortality surface analysis is to visualise the male-female mortality ratio surface for t from 1977 to 2018, and for the ages $0, 1, 5, 10, \ldots, 75$ years. The male-female mortality ratio, noted as ${}_nR_{x,t}$, is calculated by the ratio of the male mortality rate at the age group $[x, x+n[$ during the year t to the corresponding female mortality rate. It can be written as ${}_nR_{x,t} = {}_nQ^m_{x,t} / {}_nQ^f_{x,t}$ with m and f refering to male and female, respectively. The visualisation of the Lexis diagram of ${}_nR_{x,t}$, or the mortality sex ratio surface, as in Rigby and Dorling (2007), is assumed to highlight unexpected patterns caused by the effects of the methodological changes and the data imperfections.

Differences in mortality by sex have been linked to a set of environmental factors (Crimmins et al. 2019), biological differences (Kruger and Nesse 2004; Rogers et al. 2010) and, more importantly, differences between males and females in risk-taking behaviour (Rigby and Dorling 2007). All over the world, females generally have longer life expectancy at birth than men (Population Reference Bureau [PRB] 2020), but mortality can be higher for females than for males at certain ages during specific periods of time (Wisser and Vaupel 2014). The male-female gap in life expectancy at birth ranges from one year (e.g., Algeria, Burkina Faso) to 10 years (e.g., Ukraine, Russia) (PRB 2020). In most industrialised countries, the life expectancy gap between males and females was increasing until the 1980s, and started decreasing thereafter (Meslé 2004).

Building on the historical evolution of the age pattern of male-female mortality and the life expectancy gap between males and females, several different patterns have been observed since the beginning of the 21st century. Thus, analysing sex ratios can shed light on the unexpected patterns that arise due to data quality issues or changes in the vital registration system.

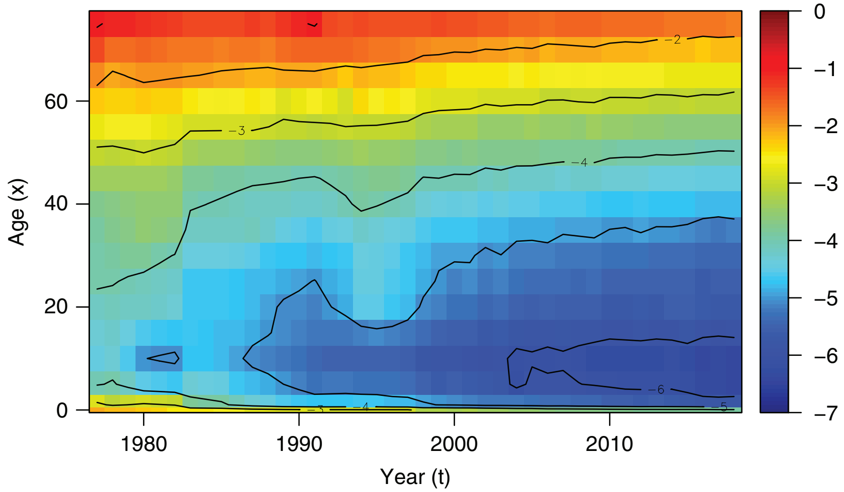
3 Results

3.1 The mortality surface

In an initial visual inspection of Figure 2, two general aspects of the evolution of mortality in Algeria are apparent: i.e., there has been a general improvement in mortality over time, and a continuous increase in mortality with age. Very high mortality rates are shown in red, while the lowest rates are displayed in dark blue. Accordingly, we can see that an individual is less likely to die at ages 5–9 years than s/he is during the rest of his/her life course. The increasing intensity of the blue colour over the years reflects decreasing mortality rates. The rising contour lines indicate an augmentation of the age at which a given level of mortality is reached. When the contour lines are distanced by one in a logarithmic scale, the rates return to levels multiplied by 2.72 in the real scale. This means that the levels defined by −6, −5, −4, −3, −2 and −1 in Figure 2 correspond to mortality rates of 0.0025, 0.0067, 0.0183, 0.0498, 0.1353 and 0.3678, respectively.

Nonetheless, the effect of the civil war during the 1990s is visible between ages 15 and 49. The decline of the contour line corresponding to $\log({}_nQ_x) = -5$ starting in 1992 means that this mortality level, corresponding to ${}_nQ_x = 0.0067$, was crossed at earlier ages than usual: i.e., around age 20, on average, during the 1992–1997 period, compared around age 25 during the preceding years. The line corresponding to −4, which is situated at relatively higher ages, marks a smaller decline, which means that the effects of the civil war were less meaningful at around ages 40 and 49 than they were at younger ages.

Figure 2:
The surface of five age log mortality rates (Algeria: 1977–2018)



Sources: Authors' computations based on data provided by ONS (2012) for 1977–2011; ONS annual publications for 2012–2018. Missing data were estimated by extrapolation and interpolation.

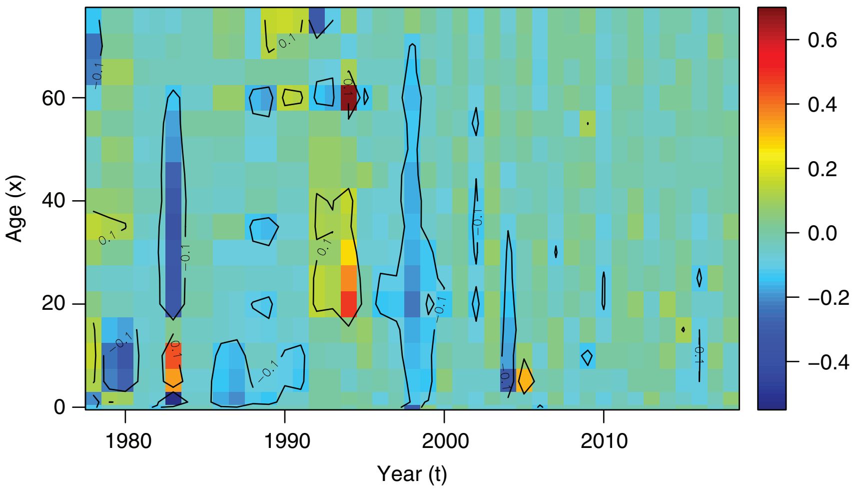
Moreover, in Figure 2, the disconnection of the map shade in 1983 reflects a sharp increase in mortality at ages five to 19, accompanied by a sharp decrease in mortality at ages 20 and older.

3.2 Mortality reduction

Figure 3 shows the annual rates of variation in mortality rates, which range from a minimum of −53% (indicating that mortality decreased, or improved, by one-half) and a maximum of +67% (indicating that mortality increased by two-thirds). The colour scale ranges from dark blue, representing a significant reduction in mortality; to dark red, representing a significant increase in mortality. Contour lines were added at the levels corresponding to +10% and −10%.

Between 1977 and 1978, mortality decreased at ages 0–4, while it increased significantly at ages 5–19. This pattern was reversed in the 1979–1980 period, and mortality returned to decreasing moderately in the 1981–1982 period. Among the 30–44 age group, mortality was increasing noticeably from 1978 to 1980. In 1983, the different age groups reacted very differently to the general evolution of mortality: i.e., mortality suddenly decreased by more than one-half in the 1–4 age group; it increased by more than 30% in the 5–9 age group and by 60% in the 10–14 age group; and it declined by between 10% and nearly 50% in the 15–64 age group.

Figure 3:
Mortality variation surface (Algeria: 1977–2018)



Sources: Authors' computations based on data provided by ONS (2012) for 1977–2011; ONS annual publications for 2012–2018. Missing data were estimated by extrapolation and interpolation.
Note: Mortality variation rates are calculated as the rate of increase/decrease in mortality rates for the different age groups between years (t – 1) and t. The figure shows a variation interval ranging from −53% (blue) to +67% (red). The contours indicate the variations surpassing 10% (in both senses).

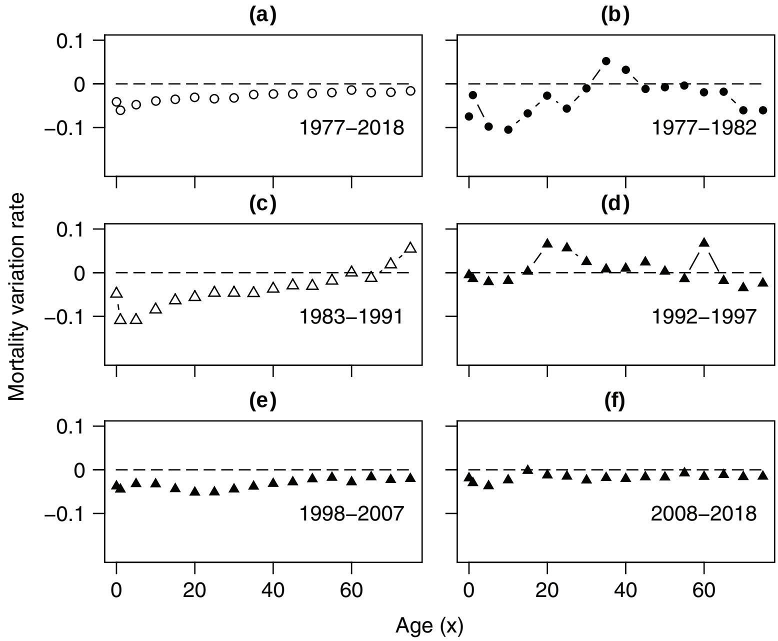
Almost all age groups experienced slight reductions in mortality from 1983 to 1991, and mortality declined rapidly in the 1–14 age group from 1983 to 1991.

After the start of the civil war in 1992, mortality increased in all age groups between zero and 49 years (except at ages 10–14), and especially at ages 20–44. A similar pattern was observed in 1993 and 1994, reflecting increases in the intensity of violence, which reached peak levels in 1994. Mortality was at approximately the same level in 1994 and 1995. In 1996 and 1997, mortality was decreasing steadily at all ages, but most significantly in the 20–29 age group.

In 1998, a sharp decline in mortality was observed at all ages, which means that there was a general decrease in mortality. In 1999, mortality continued to improve significantly for some age groups, and especially from ages one to 34 years. Thereafter, the evolution of mortality was more regular than it had been before 1998. However, sharp reductions in mortality can be observed during certain years in certain age groups: e.g., in the 20–44 age group in 2002, and in the 5–39 age group in 2004.

Figures 4 and 5 display the five-year age average variation of mortality rates by period, and the age pattern of mortality variation at the change-points defined by Flici (2020) based on the S-logistic segmentation. The analysis of the evolution of mortality rates by age revealed that the 1–4 age group benefited the most from the

Figure 4:
Annual variation of mortality rates by sub-periods (Algeria: 1977–2018)

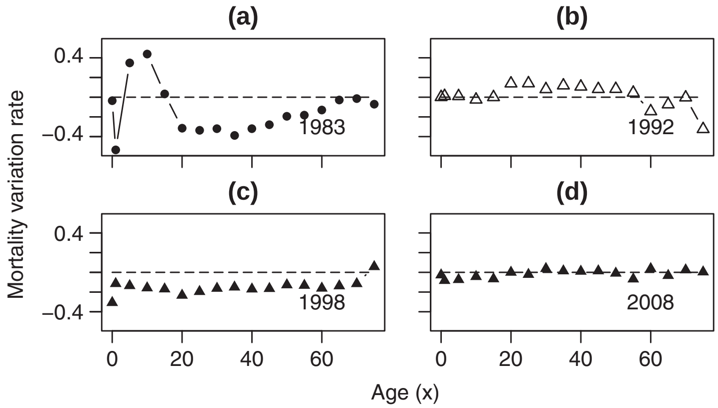


general improvements in mortality from 1977 to 2018, with an average annual rate of mortality reduction of 6% (Figure 4(a)); followed by the 5–9 age group, with a rate of nearly 5%; and by age zero, with a rate of 4.2%. The average annual rates of mortality reduction during this period for the other age groups were as follows: between 3% and 4% for the 10–34 age group, between 2% and 2.5% for the 35–54 age group, and between 1% and 2% for the 55+ age group. Thus, we can conclude that the younger age groups benefited more than the older age groups from the general improvement in mortality.

When we look at the age patterns of the average rates of mortality reduction during the 1983–1991 (Figure 4(c)), 1998–2007 (Figure 4(e)) and 2008–2018 (Figure 4(f)) periods, we see that the pace of mortality reduction was decreasing over time, and that the rate was generally decreasing with age. The average age patterns displayed during the 1977–1982 (Figure 4(b)) and 1992–1997 (Figure 4(d)) periods differed significantly from those in the other periods. This was likely due to the high variability of the age pattern of mortality variation, which was changing substantially in the abovementioned periods. Hence, simply estimating average patterns would have hidden these complex age patterns.

On the other hand, the decomposition of the change-point of 1983 (Figure 1) into age-specific variation (Figure 5(a)) revealed a sharp reduction in mortality at

Figure 5:
Annual variation of mortality rates in 1983, 1992, 1998 and 2008



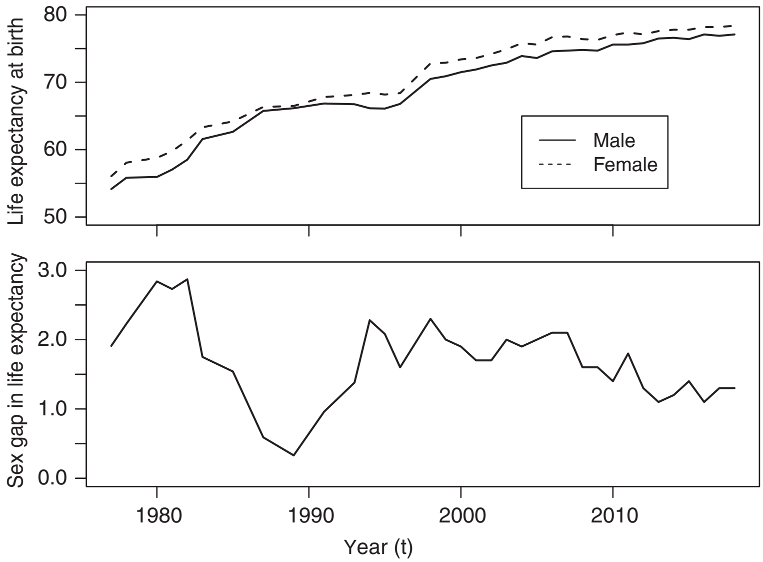
ages 20 and older, accompanied by a significant increase in the 5–14 age group. Figure 5(d) provides some insights that cannot be gleaned by simply looking at the mortality variation surface (Figure 3), even if a change-point for the evolution of life expectancy at birth in that year could be observed (Figure 1). In 2008, the average rate of mortality reduction fell to 1.7%, after it was 3.3%, on average, during the 1999–2007 period. This sudden decline in the general mortality reduction rate was accompanied by a slight distortion of the age pattern, as the rate of mortality reduction for the 1–19 age group suddenly jumped to nearly 7%, from half that level during the previous period. By contrast, the rate of mortality for the 30–49 age group unexpectedly increased by nearly +2%, after decreasing at an average rate of 3.6% during the previous period.

3.3 The evolution of the male-female differential mortality

The gap in life expectancy between males and females and the mortality sex ratio (MSR) represent complementary tools for measuring death registration completeness (Tabutin 1991). When we look at the evolution of male and female life expectancy (Figure 6), we can see that the two curves have followed similar trends, maintaining an average gap of 1.65 years, which is much smaller than the gaps observed in other North African countries (Tabutin and Schoumaker 2005).

Between 1977 and 2018, men gained 23 additional years of life expectancy from 54.2 to 77.1 years; with an average annual gain of 204 days per year. Women gained slightly fewer years of additional life expectancy (22.3 years), from 56 to 78.4 years; with an average annual gain of 198.5 days per year. The gender gap in life expectancy was as large as 2.5 years in the period before 1983. The gap narrowed by 60% in the following period, and had reached 0.3 by 1989. The gap widened

Figure 6:
Male-female differentials in the evolution of life expectancy (Algeria: 1977–2018)



Sources: Authors' computations based on data provided by ONS (2012) for 1977–2011; ONS annual publications for 2012–2018. Missing data were estimated by interpolation.

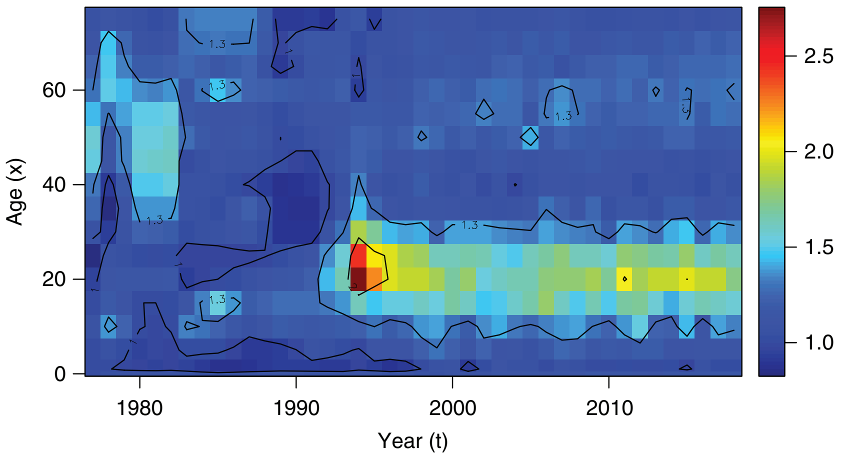
again starting in 1989, reaching 2.3 in 1994, the year when mortality associated with the civil war reached its highest level. After 1994, the gap gradually narrowed, declining to 1.3 by 2018.

Figure 7 shows the male-female mortality ratio from 1977 to 2018 for ages zero, 1–4 and 5–9, and then by five-age groups until ages 75–79. Excess female mortality can be observed in many sub-surfaces of the Lexis map marked with dark blue: i.e., between ages 20 and 29 in 1977 and between ages 30 and 44 in the following year. The 20–29 age group almost returned to having excess female mortality during the 1983–1988 period. For the 30-49 age group, excess female mortality was observed between 1989 and 1991. Moreover, excess female mortality was recorded for the 1–4 age group until 1998, and extended up to age 14 in 1981 and to age nine during the 1986–1990 period.

In addition, until 1982, relatively high excess male mortality was observed: i.e., high excess male mortality was detected in the 45–59 age group in 1977, in the 55–74 age group in 1978 and 1979, and in the 35–64 age group from 1980 to 1982.

The effects of the civil war on the male-female mortality gap were noticeable starting in 1992 with the 20–29 age group, and spread rapidly in the following years to the 15–34 age group. In 1994 and 1995, men aged 20 to 29 were two to three

Figure 7:
Mortality sex ratio surface (Algeria: 1977–2018)



Sources: Authors' computations based on data provided by ONS (2012) for 1977–2011; ONS annual publications for 2012–2018. Missing data were estimated by extrapolation and interpolation.
Note: The mortality sex ratio at age x and time t is calculated by the ratio of the male mortality rate at age x and time t to the female mortality rate at age x and time t. Male mortality is usually slightly higher than female mortality. The plot shows that the MSR ranged from slightly less than one with excess female mortality, to more than 2.5 (dark red) with excess male mortality. The contours lines mark the one, 1.3 and two levels.

times more likely to die than their female counterparts. After the civil war ended, the male-female mortality ratio among the 15–34 age group decreased, but was still larger than it was in the other age groups, and than it was before the start of the civil war.

4 Discussion

The aim of this paper was to assess the Algerian mortality data between 1977 and 2018 based on a visual inspection of Lexis maps representing different mortality indicators over age and time. Thus, we sought to complement the findings by Flici (2020), who detected important level jumps and slope changes attributable to methodological changes and data quality issues. Our analysis contributed to research on this topic by further considering variations in the mortality age patterns, which led to important additional insights into the evolution of life expectancy in Algeria. Overall, mortality rates in Algeria were decreasing over time, but the younger age groups benefited much more from that development than the older age groups, as the pace of the mortality decline decreased substantially with age. Our assessment of the quality of the mortality data represents a fundamental step towards evaluating the efficiency of public health policies. The main assumption we adopted was that under normal conditions – i.e., in the absence of mortality shocks – the evolution of mortality indicators should be gradual and smooth. Thus, we assumed that any irregular and unexpected patterns were likely attributable to the effects of methodological changes or data quality issues.

Our analysis of the mortality surface, the mortality variation surface and the male-female mortality ratios surface revealed that the different mortality measures were more stable after than before 1998. This pattern can be linked to data quality improvements and the relative stability of the methodology used to construct national life tables. Indeed, death registration completeness and the quality of population estimates were much better after 1998 than they were during the civil war (1992–1997) or in the preceding period. Overall, we found clear signs of data quality improvements over time, which means that the vital statistics have also become more reliable. Thus, official statistics no longer have to depend solely on indirect estimation techniques to derive measures of mortality.

The high variability of the mortality age pattern we observed during the 1977–1983 period, which was more evident on the mortality variation surface, might be attributable to changes in the methodology for the adjustment of the crude mortality curves using the MLTs. In addition to the opposite variations in mortality rates at the different ages, we observed extreme reversals of trends over time. Similar changes in patterns were detected on the sex ratio surface. In addition, we found evidence that death registration completeness was underestimated in the period before 1983. Indeed, the completeness rates for 1977–1980, which were not obtained through a specific survey (Hamza-Cherif 2011), were estimated to be merely 64%, whereas the Workforce and Demography Survey provided a completeness rate estimate of 81% for the year 1981 (Daoudi 2001). If death registration completeness was underestimated during the 1977–1982 period, given that the new rates were not used until 1983 onwards (Flici 2020), mortality should have been overestimated during that period. The revision of correction factors for incompleteness resulted in a downward shift in mortality, which is easily noticeable on the mortality surface as a cut of the evolution pattern, and on the mortality variation surface as a vertical line dominated by dark blue. Despite this overall increase in mortality observed in 1983, some age groups were found to have experienced significant decreases in mortality, which led to a distortion of the mortality curve between 1982 and 1983. This distortion was probably due to the changes in the mortality adjustment methods, and in the group of MLTs considered adequate for Algeria.

As the mortality indicators were more stable from 1983 to 1991, the data quality and the methods used for the estimates over this period may have been more stable as well. However, different patterns are displayed after and before 1989 on the mortality sex ratio surface. While the excess female mortality observed at reproductive ages (i.e., 15–49 years) during certain years of the period from 1983 to 1991 can be linked to the effects of maternal mortality, as the theory would suggest (De Forts 1998); the irregularity displayed around 1989 is intriguing. Despite the evidence of the importance of maternal mortality during this period, there are two findings that make it difficult to distinguish natural evolution from the effects of data quality: the observation that the surface delimiting excess female mortality suddenly shifted from the 20–29 age group in 1983–1988 to the 30–49 age group in 1989–1991; and the finding that this effect appeared suddenly in 1983, and then disappeared suddenly in 1992. Indeed, even in a context of high maternal mortality, changes that are driven by actual increases or decreases in mortality are not expected to be sudden, but are instead supposed to be gradual. When we looked at reductions in mortality, we found that the 1–14 age group had much larger reductions than the other age groups from 1983 to 1991.

During the civil war, there were two phases with different age patterns of mortality improvements: a phase of dramatic increases in mortality from 1992 to 1994, and then a return-to-normal phase from 1995 to 1997. Mortality due to violence affected the 15–34 age group more than the other age groups, and males more than females. In addition to the excess mortality due to violence, which has been estimated at 200,000 deaths (Kouaouci and Saadi 2013), there were other factors that may have affected data quality during the civil war in the 1990s. For example, while between 5,000 and 10,000 individuals (Martinez 2003) were reported as 'lost', it is likely that many of these people died, but their deaths were not registered. In addition, the registration of deaths and of other demographic events was made more challenging because civil service functions were disturbed by acts of terrorism. Some administrative offices, including archives, were destroyed in various cities during the civil war. These two factors, along with many others, may have led to a slight underestimation of the mortality of the Algerian population during the civil war. Unfortunately, these potential disruptions to the registration of deaths are not documented in the official publications of the ONS, or in academic papers.

Levels of violence were reduced towards the end of 1997, when a ceasefire agreement was signed between the government and one of the most important terrorist groups (International Crisis Group 2001), and mortality rates fell significantly between 1997 and 1998. However, the decline in mortality levels in 1998 was attributable not only to the end of the civil war, but also to methodological changes, particularly to the second revision of the correction factors for death registration incompleteness (Flici 2020). Using the death registration completeness rates of 1981 to correct crude counts up to 1998 should have generated an overestimation effect of corrected death counts, since coverage was assumed to have improved since then. The update of the correction factors offset the overestimating effect of the old correction factors, and led to a sudden decease in mortality. In addition, the update of the population structure using the data provided by the population census of 1998 contributed to the sudden mortality decrease in 1998. It is, however, intriguing that the sharp decline in 1998 was not visible on the mortality sex ratio surface. This jump was easily distinguishable on the mortality surface, and was much more visible on the mortality variation surface.

The period before 1998 was characterised by excess female mortality in the 1–4 age group, which extended to the 5–9 age group in some years. This is not a new phenomenon, and is less likely to be attributable to data quality issues than to sociological factors, such as better treatment of boys than of girls (Secrétariat d'Etat au Plan 1975), including discrimination against girls in terms of nutrition and medical care, which has sometimes been observed in traditional societies (D'Souza and Chen 1980). Tabutin (1991) found evidence of excess female mortality in the 1–4 age group in Algeria during the 1965–1977 period. Similar patterns have been observed at least until 1990 in many developing countries, including Yemen, Pakistan, Nepal, India, Bangladesh and Iran (Alkema et al. 2014); and at least until the end of the 1990s in the North African countries (Tabutin et al 2007).

After 1998, mortality was evolving at a more regular pace than it was in the period before 1998. This stability, which may reflect not only improvements in data quality, but also stability in the methodology used to construct national life tables, was easily noticeable on the mortality surface, the mortality sex ratios surface and the mortality variation surface. When we looked at the male-female mortality differential, we found that male mortality rates were significantly higher than female mortality rates at ages 15 to 34 until 2018, and that the differential was larger in this age group that it was in the other age groups. The mortality variation surface displayed some intense changes for certain ages, particularly in 2002 and 2004. The change-point observed by Flici (2020) in 2008 did not appear clearly on the mortality variation surface.

To conclude, we have demonstrated in this paper that analysing mortality evolution by age can uncover some variation patterns that cannot be observed by simply using summary mortality indicators, such as life expectancy at birth or crude mortality rates. The use of Lexis maps, or 'shaded contour maps', provides a valuable opportunity to get an informative overview of the evolution of mortality over age and time, and to detect uncommon patterns. However, when using this approach, it is necessary to compare different mortality measures, instead of relying on one measure only. Thus, visualising the mortality sex ratios surface and the mortality variation surface can help us detect certain effects that may be hidden in the mortality surface.

Acknowledgements

We gratefully thank the editors and the anonymous reviewers for their orientations, their remarks and the efforts that they have made throughout the reviewing process, which helped to improve the quality of the manuscript. Also, our special thanks go to Magali Barbieri for her remarks on a previous version of the paper.

References

Aburto, J. M. and A. van Raalte 2018. Lifespan dispersion in times of life expectancy fluctuation: the case of Central and Eastern Europe. *Demography* 55(6): 2071–2096. https://doi.org/10.1007/s13524-018-0729-9

Adair, T. and A. D. Lopez 2018. Estimating the completeness of death registration: an empirical method. *PloS ONE* 13(5): e0197047. https://doi.org/10.1371/journal.pone.0197047

Alkema, L., F. Chao, D. You, J. Pedersen and C. C. Sawyer 2014. National, regional, and global sex ratios of infant, child, and under-5 mortality and identification of countries with outlying ratios: a systematic assessment. *The Lancet Global Health* 2(9): e521–e530. https://doi.org/10.1016/S2214-109X(14)70280-3

Arthur, W. B. and J. W. Vaupel 1984. Some general relationships in population dynamics. *Population Index* 50(2): 214–226. https://doi.org/10.2307/2736755

Börger, M., M. Genz and J. Ruß 2018. Extension, compression, and beyond: A unique classification system for mortality evolution patterns. *Demography* 55(4): 1343–1361. https://doi.org/10.1007/s13524-018-0694-3

Brass, W. 1975. *Methods for estimating fertility and mortality from limited and defective data*. Laboratories for Population Statistic, University of North Carolina.

Caselli, G. and L. Enzo 1990. Graphiques et analyse démographique: quelques éléments d'histoire et d'actualité. *Population* 45(2): 399–414. https://doi.org/10.2307/1533378

Caselli, G., J. W. Vaupel and A. I. Yashin 1985. Mortality in Italy: contours of a century of evolution. *Genus* 41(1–2): 39–55.

Coale, A. J. and P. Demeny 1966. *Regional model life table and stable populations*. Princeton University Press, Princeton, New Jersey.

Coale, A. J. and G. Guo 1989. Revised regional model life tables at very low levels of mortality. *Population Index* 55(4): 613–643.

Coale, A. J., P. Demeny and B. Vaughan 1983. *Regional model life tables and stable populations*. Second Edition, London: Academic Press.

Crimmins, E. M., H. Shim, Y. S. Zhang and J. K. Kim 2019. Differences between men and women in mortality and the health dimensions of the morbidity process. *Clinical chemistry* 65(1): 135–145. https://doi.org/10.1373/clinchem.2018.288332

Daoudi, N. 2001. La mortalité en Algérie depuis l'indépendance. *Genus* 57: 109–121.

De Forts, J. 1998. Indicateurs de la santé maternelle en Algérie. Évolution de 1962 à 1992. *Population* 54(4): 859–873. https://doi.org/10.2307/1534563

Delaporte, P. 1942. Évolution de la mortalité en Europe depuis l'origine des statistiques. *Journal de la société française de statistique* 83: 183–203.

D'Souza, S. and L. C. Chen 1980. Sex differentials in mortality in rural Bangladesh. *Population and Development Review* 6(2): 257–270. https://doi.org/10.2307/1972730

Elo, I. and S. Preston 1992. Effects of early-life conditions on adult mortality: A review. *Population Index* 58(2): 186–212.

Flici, F. 2020. Analyzing the trend of life expectancy evolution in Algeria from 1962 to 2018: The S-logistic segmentation with jumps. *Population Review* 59(1): 56–72. https://doi.org/10.1353/prv.2020.0002

Gambill, B. A. and J. W. Vaupel 1985. The LEXIS program for creating shaded contour maps of demographic surfaces. IIASA Working Paper WP-85-094, Laxenburg, Austria: International Institute of Applied Systems Analysis. http://pure.iiasa.ac.at/id/eprint/2609/1/WP-85-094.pdf

Gambill, B. A., J. W. Vaupel and A. I. Yashin 1986a. The LEXIS computer program for creating shaded contour maps of demographic surfaces. IIASA Working Paper WP-86-037, Laxenburg, Austria: International Institute of Applied Systems Analysis. http://pure.iiasa.ac.at/id/eprint/2822/1/WP-86-037.pdf

Gambill, B. A., A. I. Yashin, J. W. Vaupel, Z. Nanjo and T. Shigematsu 1986b. Cause specific mortality in Japan: Contour maps approach. IIASA Working Paper WP-86-078, Laxenburg, Austria: International Institute of Applied Systems Analysis. http://pure.iiasa.ac.at/id/eprint/2783/1/WP-86-078.pdf

Gompertz, B. 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London* 115: 513–85.

Hamza-Cherif, A. 2011. L'évolution de la mortalité en Algérie de 1970 à 2007. *Journal d'Epidémiologie et de Santé Publique* 7: 21–31.

International Crisis Group 2001. The civil concord: A peace initiative wasted. ICG Africa Report, N° 31, Brussels. https://www.refworld.org/pdfid/3c0cbc604.pdf

Keiding, N. 1990. Statistical inference in the Lexis diagram. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* 332(1627): 487–509. https://doi.org/10.1098/rsta.1990.0128

Kouaouci, A. and R. Saadi 2013. La reconstruction des dynamiques démographiques locales en Algérie (1987–2008) par des techniques d'estimation indirecte. *Cahiers québécois de démographie* 42(1): 101–132. https://doi.org/10.7202/1017099ar

Kruger, D. J. and R. M. Nesse 2004. Sexual selection and the male:female mortality ratio. *Evolutionary Psychology* 2: 66–85. https://doi.org/10.1177/147470490400200112

Lexis, W. 1875. *Einleitung in die Theorie der Bevölkerungsstatistik*. Strassburg: Karl J. Trübner.

McCaa, R. 1978. Chilean social and demographic history: sources, issues, and methods. *Latin American Research Review* 13(2): 104–126.

Martinez, L. 2003. Algérie: les nouveaux défis. Working Paper. SciencesPo. https://www.sciencespo.fr/ceri/sites/sciencespo.fr.ceri/files/artlm.pdf

Meslé, F. 2004. Gender gap in life expectancy: the reasons for a reduction of female advantage. *Revue d'Épidémiologie et de Santé Publique* 52(4): 333–352. https://doi.org/10.1016/s0398-7620(04)99063-3

Minton, J. 2014. Real geographies and virtual landscapes: exploring the influence on place and space on mortality Lexis surfaces using shaded contour maps. *Spatial and spatio-temporal epidemiology* 10: 49–66. https://doi.org/10.1016/j.sste.2014.04.003

Minton, J., L. Vanderbloemen and D. Dorling 2013. Visualizing Europe's demographic scars with coplots and contour plots. *International Journal of Epidemiology* 42(4): 1164–1176. https://doi.org/10.1093/ije/dyt115

Minton, J., R. Shaw, M. A. Green, L. Vanderbloemen, F. Popham and G. McCartney 2017. Visualising and quantifying 'excess deaths' in Scotland compared with the rest of the UK and the rest of Western Europe. *Journal of Epidemiology & Community Health* 71(5): 461–467. https://doi.org/10.1136/jech-2016-207379

Moriyama, I. M. 1990. Measurement of birth and death registration completeness. Technical Report n° 43, Bethesda, Maryland, USA: International Institute for Vital Registration and Statistics. https://unstats.un.org/unsd/demographic-social/crvs/documents/IIVRS_papers/IIVRS_paper43.pdf

Mouffok, A. 1984. L'état civil en Algérie: évolution de l'enregistrement des naissances et des décès infantiles de 1900 à 1981. *Débats et Critique* 9: 57–94.

Office National des Statistiques (ONS) 2003. Démographie algérienne 2002. Données Statistiques n° 375, Alger, Algérie: Direction Technique chargée des statistiques de Population et de l'Emploi. https://www.ons.dz/IMG/pdf/DEMO2002.pdf

Office National des Statistiques (ONS) 2005. Démographie algérienne 2004. Données Statistiques n° 419, Alger, Algérie: Direction Technique chargée des statistiques de Population et de l'Emploi. https://www.ons.dz/IMG/pdf/DEMOGRAPHIE_2004.pdf

Office National des Statistiques (ONS) 2008. Démographie algérienne 2007. Données Statistiques n° 499, Alger, Algérie: Direction Technique chargée des statistiques de Population et de l'Emploi. https://www.ons.dz/IMG/pdf/demographie_algerienne2007final.pdf

Office National des Statistiques (ONS) 2010. Démographie algérienne 2009. Données Statistiques n° 554, Alger, Algérie : Direction Technique chargée des statistiques de Population et de l'Emploi. https://www.ons.dz/IMG/pdf/demographie_algerienne2009f.pdf

Office National des Statistiques (ONS) 2012. Rétrospective 1962-2011 : Démographie. https://www.ons.dz/IMG/pdf/CH1-DEMOGRAPHIE.pdf

Office National des Statistiques (ONS) 2020. Démographie algérienne 2019. Données Statistiques n°890, Alger, Algérie: Direction Technique chargée des statistiques de Population et de l'Emploi. https://www.ons.dz/IMG/pdf/demographie2019.pdf

Population Reference Bureau (PRB) 2020. 2020 World Population Data Sheet, Washington, DC: Population Reference Bureau. https://www.prb.org/wp-content/uploads/2020/07/letter-booklet-2020-world-population.pdf

Preston, S. and Hill, K. 1980. Estimating the completeness of death registration. *Population Studies* 34(2): 349–366. https://doi.org/10.1080/00324728.1980.10410395

Preston, S., A. J. Coale, J. Trussell and M. Weinstein 1980. Estimating the completeness of reporting of adult deaths in populations that are approximately stable. *Population Index* 45(2): 179–202. https://doi.org/10.2307/2736122

Purdy, C., W. Weiss and H. Perry 2013.*The Mortality Assessment for Health Programs (MAP) System: An NGO Field Manual for Registering Vital Events and Assessing Child Survival Outcomes Using the Care Group Model*. 1st Edition. Washington, D.C: CORE Group. https://pdf.usaid.gov/pdf_docs/PA00JMZS.pdf

Rau, R., C. Bohk-Ewald, M. M. Muszyńska and J. W. Vaupel 2018. *Visualizing mortality dynamics in the Lexis diagram*. Cham: Springer. https://doi.org/10.1007/978-3-319-64820-0

Remund, A., C. G. Camarda and T. Riffe 2018. A cause-of-death decomposition of young adult excess mortality. *Demography* 55(3): 957–978. https://doi.org/10.1007/s13524-018-0680-9

Rey, G., A. Aouba, G. Pavillon, R. Hoffmann, I. Plug, R. Westerling, E. Jougla and J. Mackenbach 2011. Cause-specific mortality time series analysis: a general method to detect and correct for abrupt data production changes. *Population Health Metrics* 9(1): 52. https://doi.org/10.1186/1478-7954-9-52

Rigby, J. E. and D. Dorling 2007. Mortality in relation to sex in the affluent world. *Journal of Epidemiology & Community Health* 61(2): 159–164. https://doi.org/10.1136/jech.2006.047381

Rogers, R. G., B. G. Everett, J. M. Saint Onge and P. M. Krueger 2010. Social, behavioral, and biological factors, and sex differences in mortality. *Demography* 47(3): 555–578. https://doi.org/10.1353/dem.0.0119

Salhi, M. 1984. L'évolution récente de la mortalité en Algérie (1965-1981). Working paper N° 117, Louvain-La-Neuve: Département de Démographie, Université Catholique de Louvain.

Secrétariat d'Etat au Plan 1975. *Etude statistique nationale de la population, résultats de l'enquête démographique*. Tome IV : Mortalité. Série 02 : Résultats, volume 7. Alger, Algérie : Secrétariat d'Etat au Plan.

Tabutin, D. 1991. High mortality rates among women in North Africa from 1965 to the present day – Descriptive aspects. *Population* 46(4): 833–854.

Tabutin, D. and B. Schoumaker 2005. The demography of the Arab world and the Middle East from the 1950s to the 2000s: A survey of changes and a statistical assessment. *Population* 60(5): 505–616. https://doi.org/10.2307/4148186

Tabutin, D., C. Gourbin and G. Beninguisse 2007. Surmortalité et santé des petites filles en Afrique. Tendances des années 1970 aux années 1990. In *Genre et sociétés en Afrique. Implications pour le développement*, ed T. Locoh 137–170. Paris: INED.

United Nations 1961. *Demographic Yearbook.* Thirteenth issue, Special Topic: Mortality Statistics. New York: Statistical Office of the United Nations. https://unstats.un.org/unsd/demographic-social/products/dyb/dybsets/1961%20DYB.pdf

United Nations 1982. *Model life tables for developing countries*. Population Studies, No. 77. New York: United Nations, Department of International Economic and Social Affairs.

Vanderbloemen, L., D. Dorling and J. Minton 2016. Visualising variation in mortality rates across the life course and by sex, USA and comparator states, 1933–2010. *Jorunal of Epidemiology & Community Health* 70(8): 826–831. http://doi.org/10.1136/jech-2014-205226

Vaupel, J. W., B. A. Gambill and A. I. Yashin 1985. Contour maps of population surfaces. IIASA Research Report. IIASA, Laxenburg, Austria: WP-85-47. http://pure.iiasa.ac.at/id/eprint/2653/1/WP-85-047.pdf

Vaupel, J. W., B. A. Gambill and A. I. Yashin 1987. Thousands of Data at a Glance: Shaded Contour Maps of Demographic Surfaces. IIASA Research Report RR-87-016, Laxenburg, Austria: International Institute of Applied Systems Analysis. http://pure.iiasa.ac.at/id/eprint/2905/1/RR-87-016.pdf

Vékás, P. 2020. Rotation of the age pattern of mortality improvements in the European Union. *Central European Journal of Operations Research* 28(3): 1031–1048. https://doi.org/10.1007/s10100-019-00617-0

Wilmoth, J. R. 1985. Identifiable Age, Period, and Cohort Effects: An Exploratory Approach Applied to Italian Female Mortality. IIASA Working Paper WP-85-069. Laxenburg, Austria: International Institute of Applied Systems Analysis. http://pure.iiasa.ac.at/id/eprint/2632/1/WP-85-069.pdf

Wisser, O. and J. W. Vaupel 2014. The sex differential in mortality: a historical comparison of the adult-age pattern of the ratio and the difference. MPIDR Working Paper WP 2014-005. Rostock: Max Planck Institute for Demographic Research. https://www.demogr.mpg.de/papers/working/wp-2014-005.pdf

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 191–213

Gender disparities in health at older ages and their consequences for well-being in Latin America and the Caribbean

Marília R. Nepomuceno[1,*], *Vanessa di Lego*[2], *and Cássio M. Turra*[3]

Abstract

Women live longer but can expect to spend more years in poorer health compared to men. In the context of population aging and declining gender ratios at older ages, there are increasing concerns about how this disadvantage in female health will affect well-being and sustainability, particularly in developing regions that are rapidly aging. Our study compares differences in health expectancies at older ages for men and women in order to assess gender disparities in health. We use data from the Survey on Health, Well-Being, and Aging in Latin America and the Caribbean to decompose the gender gap into total and age-specific mortality and disability effects in seven cities in the region. Our results show that at older ages, higher disability rates among women reduced the gender gap in healthy life expectancy by offsetting women's mortality advantage. In addition, we find that women's mortality advantage decreased almost systematically with age, which reduced the contribution of the mortality effect to the gender gap at older ages. Although the gender gap in health followed a similar pattern across the region, its decomposition into mortality and disability effects reveals that there was substantial variation among cities. Thus, across the region, the implications of the gender gap in health for well-being vary, and the policies aimed at reducing this gap should also differ.

Keywords: gender gap; healthy life expectancy; disability; older ages; Latin America and the Caribbean; decomposition

[1]Max Planck Institute for Demographic Research, Rostock, Germany
[2]Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria
[3]Demography Departament – Cedeplar, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[*]Correspondence to: Marília R. Nepomuceno, nepomuceno@demogr.mpg.de

DOI: 10.1553/populationyearbook2021.res2.1

1 Introduction

Promoting health and well-being at older ages is one of the three priority areas established in the 2002 Madrid International Plan of Action on Ageing (MIPAA) to ensure that population aging is sustainable and equitable (World Health Organization 2004). It is also one of the 17 Sustainable Development Goals (SDGs), which extended the target to all ages (United Nations Development Programme 2015). However, countries face a number of challenges in meeting and monitoring progress toward those goals, including unique preexisting gender- and age-specific disparities in health and mortality that directly affect well-being. Scholars have examined these gaps across several countries in different parts of the world (Jagger and Matthews 2002; Oksuzyan et al. 2008; Nusselder et al. 2010). It appears, however, that the consequences of these gaps are more dramatic in countries with rapidly aging populations where inequality levels are high and long-term care is provided informally by families, as is the case in many Latin American countries (Palloni and McEniry 2007; United Nations 2017). Moreover, due to the high levels of socioeconomic inequality in the region, early-life conditions in these countries are often poor. As these conditions are predictive of disability at older ages, they may increase the burden of health for older people, and negatively affect their well-being (Monteverde et al. 2009). Thus, in Latin America and the Caribbean, the speed of population aging has outstripped the ability of societies to solve old and new social, economic and health issues (Turra and Fernandes 2020).

Even though gender disparities in health have been extensively studied worldwide (Case and Paxson 2005; Crimmins and Kim 2010; Andrade et al. 2011; di Lego et al. 2020; Grundy 2006; Jacobsen et al. 2008; Mathers et al. 2001; Oksuzyan et al. 2014; Verbrugge 1989; Yong et al. 2010), there has been no conclusive explanation for why, despite living longer than men, women perform worse in terms of their disability, chronic morbidity and self-rated health outcomes (Crimmins et al. 2002; Luy and Minagawa 2014; Robine et al. 2001; Spiers et al. 2003). Compared to men, women have higher morbidity from acute conditions, nonfatal chronic diseases and short-term disabilities (Green and Pope 1999; Verbrugge 1985). The rate of decline in physical functioning is also higher among women, with women being outperformed by men across all ages on tests of hand grip strength (Leong et al. 2015; Sanderson and Scherbov 2014), walking speed and standing balance (Keevil et al. 2013). Moreover, compared to men, women are less likely to recover once disabled (Beckett et al. 1996), and they are more likely to use health care services (Green and Pope 1999; Redondo-Sendino et al. 2006) and prescription and nonprescription drugs (Roe et al. 2002). Studies have shown that these gender differences are consistent across countries and at older ages; with the exception of the pattern of gender differences in depressive symptoms, which appears to be more country-specific (Oksuzyan et al. 2010). Because men have higher mortality than women at all ages and for many leading causes of death, it is puzzling that the proportion of life spent in good health is higher for men than for women (Crimmins and Saito 2001; Nusselder et al. 2010; Rieker and Bird 2005; Van Oyen et al. 2013). These contradictions in the mortality-morbidity differences between the genders have led to numerous studies describing this phenomenon as the "gender and health paradox" (Rieker and Bird 2005), the "morbidity paradox" (Gorman and Read 2006), the "morbidity-mortality paradox" (Kulminski et al. 2008) or the "male–female health-survival paradox" (Oksuzyan et al. 2008).

Studies conducted in Latin American and the Caribbean (LAC) countries have systematically described women as suffering from excess morbidity, despite having higher longevity (Andrade et al. 2014; Camargos et al. 2007; Palloni and McEniry 2007; Zunzunegui et al. 2009). In LAC, women are more likely than men to report poor self-rated health, and the prevalence of disability is around 50% higher among women than among men (Barbosa et al. 2005; Duarte et al. 2005). In terms of the gender gap in healthy life expectancy, the pattern in LAC is similar to the pattern that has been documented in the wealthiest regions: i.e., women live longer than men, but the proportion of time they spend living with a disability is greater (Andrade et al. 2011; Camargos et al. 2008). Although these previous studies have shown that there are health disparities between men and women at older ages, the specific contributions of mortality and disability to these gender disparities in health expectancy remain unclear (Angel et al. 2017; Campos et al. 2015; Andrade et al. 2011; Wong et al. 2005).

There are several ways to measure gender disparities in health, including health expectancy. This summary measure is widely used to evaluate the average level of population health (Murray et al. 2002). In addition, it is a policy planning tool that is generally used to compare the level of health across populations (Nusselder et al. 2010; Robine et al. 2009; Van Oyen et al. 2010; Yokota et al. 2019). Decomposition analyses of the contributions of mortality and disability to the gender gap in health expectancy in high-income countries have suggested that considerable gender differences in mortality and disability can be masked when only the total gap is analyzed (Mairey et al. 2014; Nusselder et al. 2010; Nusselder and Looman 2004; Van Oyen et al. 2013). For example, decomposition analyses conducted in the Netherlands found there are substantial effects of mortality (3.1 years) and disability (−3.5 years) by gender; but also that because the two effects work in opposite directions, they result in a small net gap (−0.4 years). Therefore, when examining the gender gap in health expectancy, it is important to focus not only on the total gender gap, but also on the contributions of its components. This approach is crucial for targeting policy priorities, particularly those that seek to reduce health differentials by subgroups.

Our study aims to compare differences in the health expectancies of men and women across large cities in seven countries in Latin America and the Caribbean: Argentina, Brazil, Chile, Mexico, Uruguay, Cuba and Barbados. We use information from the Survey on Health, Well-Being, and Aging in Latin America and the Caribbean (SABE) to examine whether the gender gap in health expectancy follows a similar pattern across those countries, and to estimate the contributions of disability and mortality by age. We then discuss the consequences of these findings for monitoring overall health and well-being in the region. By 2050, 198 million people aged 60 years or older will be living in Latin America and the Caribbean, with a ratio of 84 men per 100 women (United Nations 2017). Therefore, understanding the mechanisms that contribute to morbidity and mortality differentials among older individuals in such contexts is critical for improving the health and well-being of these individuals, and for meeting the Sustainable Development Goals.

2 Gender differences in health and mortality across LAC countries

The overall pattern of gender differences in healthy life expectancy across countries in LAC is similar to the patterns observed in higher-income countries: i.e., women live longer than men, but the proportion of time they spend living with a disability is greater (Andrade et al. 2011; Camargos et al. 2008). However, the specific contributions of mortality and health to these differentials remain unclear. In addition, the LAC region, and specifically the SABE countries analyzed here, have mortality and health characteristics that make it particularly challenging to determine what drives these gender gaps, and their magnitudes (Angel et al. 2017; Campos et al. 2015; Andrade et al. 2011; Wong et al. 2005).

The health and mortality conditions that underlie these observed gender disparities vary considerably across the region. For example, it has been shown that in Chile, musculoskeletal disorders and fall-related fractures play an important role in health differences by gender, as the prevalence of osteoarthritis is twice as high among older women as among older men, independent of their socioeconomic position (Moreno et al. 2018; Murtagh and Hubert 2004). Similarly, the prevalence of osteoporosis is more than five times higher among women than among men, with women reporting more frequent falls (Albala et al. 2011). In Mexico, by contrast, obesity and diabetes appear to be the most critical factors, especially at ages 70–75, as in this age group, the share of women with diabetes is high (29%), and the percentage of women with obesity is significantly higher than that of men (26.7% vs. 22.3%) (Wong et al. 2015). Both diabetes and obesity are risk factors for a series of chronic conditions, and are associated with the incidence of multiple comorbidities, including type II diabetes, cancer, and cardiovascular diseases (Guh et al. 2009; Kearns et al. 2014). On the other hand, Cuba, despite being the country with the lowest levels of diabetes and obesity in the Latin American and Caribbean region (Palloni and McEniry 2007), has high prevalence levels of acute myocardial infarction, with men having a higher risk of dying from this cause than women of the same age (Armas et al. 2012). Moreover, the incidence of lung cancer and the mortality risk associated with it has increased among older men in LAC (Galán et al. 2009). Bridgetown (Barbados) has struggled over the past decade with a very high burden of deaths from liver cirrhosis attributed to alcohol abuse and heavy smoking among men (Moonie and Quashie 2011). In Brazil, gender differences in disability severity and in levels of life-threatening conditions seem to contribute more to the gender gap in healthy life expectancy. It has, for example, been reported that compared to their male counterparts, Brazilian women are more likely to be disabled, and to have more severe disabilities (Camargos et al. 2007; Camargos et al. 2008; Andrade et al. 2011). It has also been shown that in Brazil, unhealthy life expectancy is higher among women than among men at all ages, and that a woman at age 60 can expect to live two years longer with severe limitations than a man at the same age. These findings imply that Brazilian women live proportionally more years with severe limitations than men (Belon et al. 2014; Nepomuceno and Turra 2012; Parahyba et al. 2005). Evidence for higher-income countries indicates that for men, having severe limitations is associated with a higher mortality risk. Thus, gender differences in mortality in these countries may be explained by gender differences in the effects these conditions have on mortality (Case and Paxson 2005). This might also be the case for Brazil.

There are significant differences in the socioeconomic, demographic and health conditions of the countries in which the SABE was conducted. For example, despite not being the wealthiest country in the LAC region, Cuba has been prioritizing health for almost half a century. The country has an extensive primary health care network, a long history of controlling infectious diseases, relatively low infant mortality, and below replacement level fertility since 1978. Thus, the aging process in Cuba is very advanced (Franco et al. 2007a). The trajectory of Bridgetown resembles that of Cuba in terms of the stage of the demographic transition and the efficiency of the public health care system (Hennis et al. 2005). In Bridgetown, life expectancy at birth in 2001 was 76.9 years, and was thus higher than the average of 70.3 years for the entire region (Knight et al. 2004). In addition, of the SABE cities, Bridgetown has the highest proportion of women aged 75+ (Palloni and McEniry 2007). Interestingly, Uruguay has the lowest poverty and urban economic inequality levels in Latin America (urban Gini = 0.44, 10/40 ratio = 8.8), spends the most per capita on medical care, and had the lowest rate of poverty among the elderly population in the year 2000 (Wallace and Gutiérrez 2005). Nonetheless, in the early 2000s, Uruguay had one of the lowest remaining life expectancies at age 60 for men, and one of the highest for women, in the LAC region (United Nations 2019). It has also been shown that older Mexicans account for more than 50% of families living in extreme poverty; i.e., households living on one dollar a day or less. As women live longer than men, higher proportions of women than of men are living in poverty, which may contribute to the higher levels of disability among women (Angel et al. 2017; Huenchuan 2013).

Social inequality in health care services, expected outpatient medical care, and hospital admissions also vary across SABE cities. Inequalities in health care access are less marked in Cuba, Argentina and Uruguay, but are more pronounced in Brazil and Mexico. Previous research has indicated that these inequalities have resulted in somewhat puzzling health care patterns across the region (Noronha and Viegas 2005). For instance, given the relatively high level of income inequality in Brazil and the lower level in Uruguay, equity in access to care for older people was found to be better than expected in Brazil and worse than expected in Uruguay.

Not surprisingly, inequity in access to care was attributed mainly to low levels of health insurance coverage in Mexico and to differences in socioeconomic status in Chile (Wallace and Gutiérrez 2005). Finally, excess mortality from external causes (homicide, motor vehicle accidents and suicide) among men aged 15–29 is much higher in these countries than it is in developed countries. LAC countries have the highest homicide rates for young men in the world. Thus, external causes of death have contributed to significant changes in life expectancy in the LAC region, and have hindered improvements in male longevity (Canudas-Romo and Aburto 2019; Acosta et al. 2018). All of these factors make it particularly challenging to understand how mortality and health are contributing to gender disparities in healthy life expectancy in LAC countries, and what is driving those observed differences.

3 Data and methods

The data come from the Survey on Health, Well-Being, and Aging in Latin America and the Caribbean (SABE) (Pelaez et al. 2006). The SABE interviewed non-institutionalized representative samples of individuals aged 60 and older in seven cities in Latin America and the Caribbean in 1999–2000. Table 1 presents the number of respondents by age group and city in 1999–2000. For some cities, such as São Paulo, the survey has a longitudinal design, with other data collection waves taking place in 2006 and 2010 (Lebrão et al. 2018). Unfortunately, the longitudinal study was not carried out in all cities, which prevents us from performing trend analyses. Therefore, we can only perform cross-city comparisons for the year 2000.

The SABE sample is limited to urban areas, which restricts our ability to generalize our results to the total population. However, as previous research has shown that the demographic profiles of the samples are close to the national averages, our results should not diverge too dramatically from those for the general population (Palloni and McEniry 2007; Wong et al. 2005).

In this study, we compare estimates for the seven cities included in the 2000 survey: Buenos Aires (Argentina), São Paulo (Brazil), Santiago (Chile), Mexico City (Mexico), Montevideo (Uruguay), Havana (Cuba) and Bridgetown (Barbados). The most significant advantages of the SABE are the consistency and the comparability of the data across countries, with the surveys being translated into three different languages (Spanish, Portuguese and English); and the high rates of response. Havana had the highest response rate (95%), followed by Bridgetown (85%), São Paulo (85%), Mexico City (85%) and Chile (84%). The response rates in Montevideo and Buenos Aires were somewhat lower, at 66% and 60%, respectively (Palloni and McEniry 2007).

To measure health status, we use the question that evaluates functional limitations based on the Katz Index of Independence in Activities of Daily Living: i.e., limitations in bathing, dressing, toileting, walking across a room, eating and getting in and out of bed. Thus, the question covers the fundamental skills typically required to manage an individual's basic physical needs, and is comparable to the questions

Table 1:
SABE sample size by age and city, 2000

Age Group	Bridgetown (Barbados)	Buenos Aires (Argentina)	Havana (Cuba)	Mexico City (Mexico)	Montevideo (Uruguay)	Santiago (Chile)	São Paulo (Brazil)
	Sample size						
60–64	283	241	483	383	317	291	426
65–69	340	252	402	310	366	327	379
70–74	300	252	349	230	330	240	336
75–79	264	160	257	156	242	209	472
80–84	177	84	207	98	124	128	307
85+	144	54	207	70	71	106	223
Total	1,508	1,043	1,905	1,247	1,450	1,301	2,143

Source: SABE.

used in other surveys of older people, particularly the Health and Retirement Survey (HRS) (2002). We define as disabled those individuals who answered "yes" to at least one of the following questions: "Do you have difficulty bathing?"; "Do you have difficulty dressing?"; "Do you have difficulty using the toilet?"; "Do you have difficulty walking across a room?; "Do you have difficulty eating?"; and "Do you have difficulty getting in and out of a bed?" This definition is in line with the recommendations in previous studies that used ADLs from the SABE (Albala et al. 2005; Palloni et al. 2002; Palloni and McEniry 2007).

Information on the prevalence of disability for each city drawn from the SABE is combined with city-specific life tables by sex in 2000. The exception is Bridgetown, for which we use a national life table. All of the mortality data were generously provided by Flavia Andrade, who estimated them for an earlier article (Andrade 2009). Since SABE is a sample survey, we calculate the proportion of disabled individuals in the sample, and compute a 95% binomial confidence interval for the disabled proportion. The confidence intervals rely on approximating the binomial distribution with a normal distribution. This approximation is based on the central limit theorem, and, since our samples are large, it is reliable. Using this approach, we are able to present a range of values for measures derived from the SABE data.

To examine gender disparities in health expectancy, we estimate the disability-free life expectancy (*DFLE*) using the Sullivan Method (Sullivan 1971), a methodological approach that has been used before in similar analyses (e.g., Crimmins et al. 2016). For each age group, we estimate the prevalence of disability from SABE data and combine it with the total number of person-years lived obtained from life tables, in order to disentangle the number of years lived with and without disability. The number of person-years lived free of disability is calculated as,

$$ {}_nL_x^i = {}_nL_x(1 - {}_n\pi_x), \tag{1} $$

where ${}_nL_x^i$ is the number of person-years lived without disability between ages x and $x + n$, ${}_nL_x$ is the total number of person-years lived in the age group x and $x + n$, and ${}_n\pi_x$ is the proportion of disabled individuals in the age group x and $x + n$.

Then, life expectancy free of disability (*DFLE*) is calculated as,

$$DFLE_x = \frac{\sum_{k=x}^{w}({}_nL_k^i)}{\ell_x} \tag{2}$$

where $DFLE_x$ is the number of years lived without disability at age x, w is the starting age of the open age interval, and ℓ_x is the number of survivors at age x.

We then calculate gender gap in *DFLE* as,

$$\Delta_x = DFLE_x^{Women} - DFLE_x^{Men}, \tag{3}$$

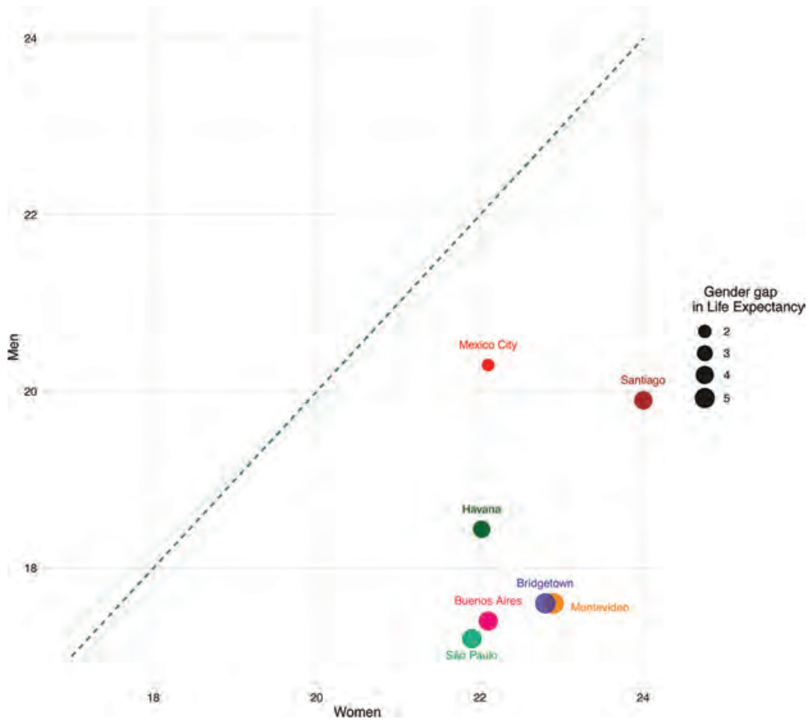
and split the gender differences in *DFLE* at age x into mortality and disability effects by five-year age groups. To decompose the gap, we apply the continuous change decomposition method that was developed by Horiuchi et al. (2008) and implemented in R by Riffe (2018). The continuous change decomposition method assumes that covariates (e.g., age-specific mortality rates and age-specific prevalence of disability) change continuously along an actual or hypothetical dimension, such as between two periods or between two populations, thereby modifying aggregate measures such as life expectancy and healthy life expectancy. Each of these tiny changes in the aggregate indices can be approximated by a linear combination of n partial derivatives of the function with respect to the covariates (Horiuchi et al. 2008). Then, numerical integration is used to obtain the total contribution of the covariates for the variation of the aggregate measure. This method is very flexible, and can be used for decomposing gaps in different aggregate measures, including healthy life expectancy, as presented by van Raalte and Nepomuceno (2020).

4 Results

Figure 1 compares life expectancy at age 60 for women and men. Not surprisingly, the results indicate that women were living longer than men in all cities. Among women, life expectancy at age 60 varied from 21.9 in São Paulo (Brazil) to 24 years in Santiago (Chile). Among men, the estimates show more variation across cities, ranging from 17.2 years (São Paulo) to 20.3 years (Mexico City). The gender gap (women-men) in life expectancy was larger for the four cities (São Paulo, Buenos Aires, Bridgetown and Montevideo) where male life expectancy at age 60 was below 18 years.

Among women, the mortality advantage contrasted with a disability disadvantage. Figure 2 shows the differences in the prevalence of disability between women and men aged 60 and older in each city. In all cities, the prevalence rates were higher for women than for men. Buenos Aires and Santiago were the cities with the largest

Figure 1:
Total life expectancy (LE) at age 60 and the size of the gender gap in life expectancy (women–men), by city, 2000

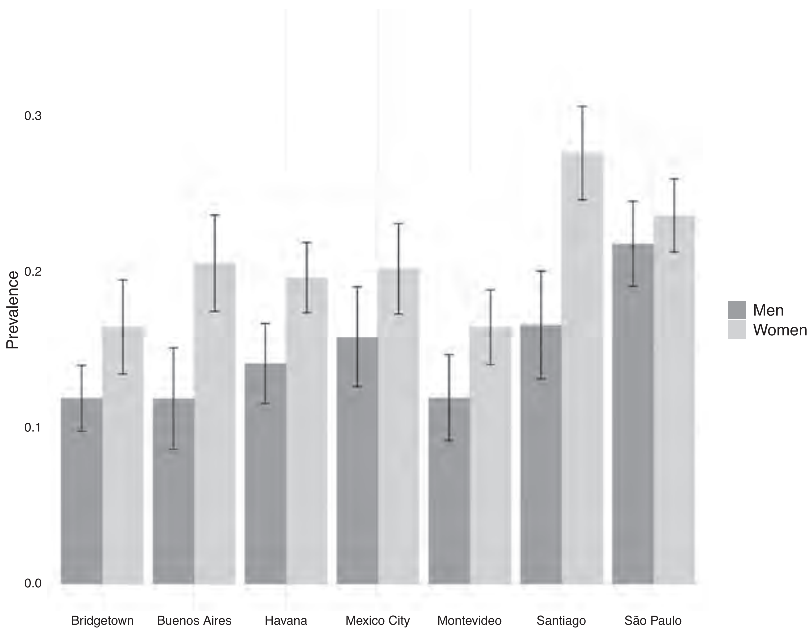


Note: The dashed line corresponds to equal life expectancies at age 60 for men and women.
Source: Own elaboration using data from Andrade (2009).

differences by gender. Meanwhile, São Paulo had the lowest variation in disability by gender.

The gender differences in mortality and disability can be examined simultaneously by estimating and decomposing the gender gap in life expectancy. Table 2 shows total life expectancy (LE), disability-free life expectancy (DFLE), and life expectancy with disability (LEWD) at age 60 for the seven LAC cities in the year 2000. As expected, we can see that in all cities, the gender differences in LE favored women. The largest gap in LE at age 60 was in Montevideo (5.30 years), followed by in Bridgetown (5.10 years), São Paulo (4.70 years), Buenos Aires (4.60 years), Santiago (4.10 years), Havana (3.58 years) and Mexico City (1.80 years). Women were living longer than men in all cities, but in Buenos Aires, Mexico City and Santiago, most of the women's additional years of life were lived with disabilities. For example, in Mexico City, a full 90% of the life expectancy

Figure 2:
Prevalence of disability at ages 60 and older, by gender and city, 2000



Note: Black segments correspond to a 95% confidence interval.
Source: Own elaboration using data from the SABE.

advantage of women was lived with disabilities. A similar pattern is observed in Santiago, although the number of years women were living with disabilities was much higher in Santiago (3.71) than in Mexico City (1.62). The comparison of the results in the two cities shows why focusing on only one gender gap health indicator can mask significant differences between cities. In Montevideo, Bridgetown and Havana, about 39% of women's additional years of life were spent with disabilities. In comparison, in São Paulo, 45% of women's remaining years of life were spent with disabilities.

In all cities, both the disability and the mortality effects increased the gender gap in LEWD. In other words, women were living more years with disabilities because they were living longer and had higher rates of disability. The disability effects were larger than the mortality effects in Buenos Aires, Mexico City and Santiago; whereas the mortality effects were larger than the disability effects in Bridgetown, Havana, Montevideo and São Paulo.

The gender gap in disability-free life expectancy (DFLE) indicates that in absolute terms, women also enjoyed more healthy years of life than men in all cities. However, for Buenos Aires, Mexico City and Santiago, DFLE represented a smaller

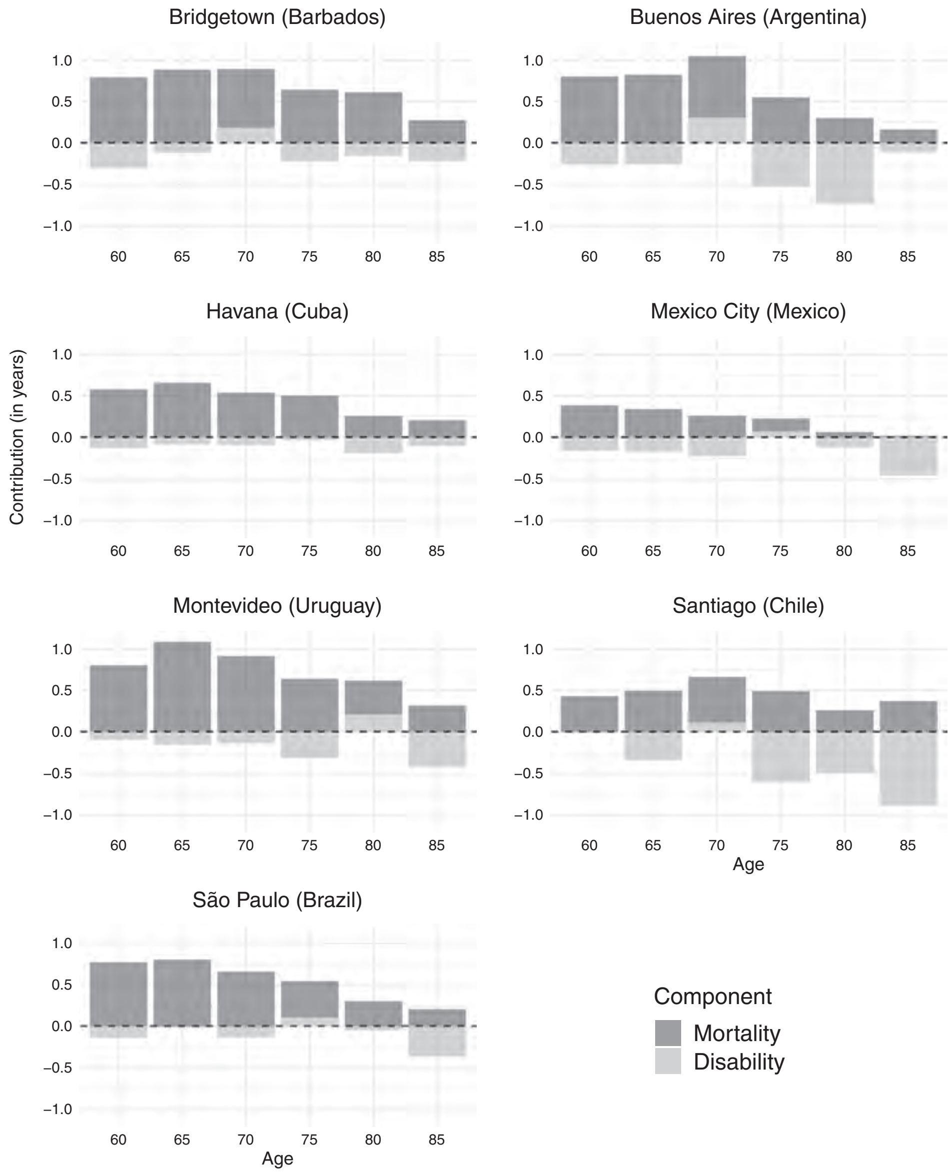
Table 2:
Decomposition of the gender gap (women-men) in total life expectancy (LE), disability-free life expectancy (DFLE) and life expectancy with disability (LEWD) into mortality and disability effects at age 60 (confidence intervals in brackets), by city, 2000

	LE	DFLE	LEWD
Bridgetown (Barbados)			
Gender gap	5.10	3.11 (2.84; 3.37)	1.99 (1.73; 2.26)
Mortality effect	5.10	3.91 (3.46; 4.34)	1.20 (0.76; 1.64)
Disability effect		−0.80 (−0.97; −0.62)	0.80 (0.62; 0.97)
Buenos Aires (Argentina)			
Gender gap	4.60	1.82 (1.57; 2.07)	2.78 (2.53; 3.03)
Mortality effect	4.60	3.38 (2.81; 3.94)	1.23 (0.66; 1.79)
Disability effect		−1.55 (−1.87; −1.24)	1.55 (1.24; 1.87)
Havana (Cuba)			
Gender gap	3.58	2.11 (2.10; 2.13)	1.47 (1.45; 1.48)
Mortality effect	3.58	2.73 (2.42; 3.04)	0.85 (0.54; 1.16)
Disability effect		−0.61 (−0.91; −0.32)	0.61 (0.32; 0.91)
Mexico City (Mexico)			
Gender gap	1.80	0.18 (0.09; 0.26)	1.62 (1.54; 1.71)
Mortality effect	1.80	1.22 (1.02; 1.43)	0.58 (0.37; 0.78)
Disability effect		−1.04 (−1.32; −0.76)	1.04 (0.76; 1.32)
Montevideo (Uruguay)			
Gender gap	5.30	3.28 (3.05; 3.50)	2.02 (1.80; 2.25)
Mortality effect	5.30	4.17 (3.64; 4.71)	1.13 (0.59; 1.66)
Disability effect		−0.90 (−1.21; −0.59)	0.90 (0.59; 1.21)
Santiago (Chile)			
Gender gap	4.10	0.39 (0.32; 0.46)	3.71 (3.64; 3.78)
Mortality effect	4.10	2.61 (2.15; 3.06)	1.50 (1.04; 1.95)
Disability effect		−2.21 (−1.69; −2.74)	2.21 (2.74; 1.69)
São Paulo (Brazil)			
Gender gap	4.70	2.59 (2.34; 2.75)	2.11 (1.95; 2.36)
Mortality effect	4.70	3.18 (2.85; 3.50)	1.53 (1.20; 1.85)
Disability effect		−0.58 (−0.75; −0.41)	0.58 (0.41; 0.75)

Source: Own elaboration using data from the SABE and Andrade (2009).

fraction of the gender gap in LE than LEWD. The largest gender gaps in DFLE occurred in cities where women's mortality advantage was substantial and surpassed women's disability disadvantage, which was the case in Bridgetown, Buenos Aires, Havana, Montevideo and São Paulo. In Mexico City, the combination of a high prevalence of disability among women (−1.04 years) and a small women's mortality

Figure 3:
Decomposition of the gender gap in disability-free life expectancy (DFLE) into mortality and disability effects by five-year age groups and city, 2000



Source: Own elaboration using data from the SABE and Andrade (2009).

advantage (1.22 years) resulted in men and women having nearly the same number of healthy years of life after age 60. In Santiago, the gender gap in DFLE was also close to zero (0.39 years). Figure 3 details the age decomposition of the gap in DFLE. As expected, the mortality effects – meaning more years of life – were higher among women in all age groups and across all cities. However, the mortality effects decreased almost systematically with age, which suggests that the contribution of gender differences in mortality to the gap in DFLE was lower at older ages. For instance, the mortality effects were more than 70% lower between ages 60 and 85+ in Buenos Aires, Mexico City and São Paulo. In contrast, Figure 3 does not show a clear age pattern for the disability effects across the seven cities, although they tended to affect women's health measures negatively at most ages. There were, however, exceptions: in Buenos Aires and Santiago, the disability effects were positive in the 70–74 age group, which means that the prevalence rates of ADLs were higher among men than women. Still, the effects were negative at all other ages. Havana was the only city where women experienced higher levels of disability at all ages, but they were lower in magnitude than in the other LAC cities.

5 Discussion

Our results have confirmed that in Latin American and Caribbean cities in 2000, women were living longer than men, but were spending more years in poor health. These results are consistent with those of other studies that focused on gender differences in mortality and morbidity, not only in the LAC region, but worldwide (Camargos et al. 2007; Camargos et al. 2008; Andrade et al. 2011; Jagger and Matthews 2002; Oksuzyan et al. 2008; Palloni and McEniry 2007; Van Oyen et al. 2010; Wong et al. 2005; Yokota et al. 2019; Zunzunegui et al. 2009).

Although our findings indicate that the gender gap in health in terms of both disability-free life expectancy (DFLE) and life expectancy with disability (LEWD) followed a similar pattern within the region, its decomposition into the contribution of mortality and disability effects showed that there was substantial variation across cities, which suggests that the underlying mechanisms of the gender gap differed considerably. Our results highlight the importance of investigating not only the contributions of disability and mortality to the gender gap in DFLE, but also the years lived with disability. DFLE was higher for women than for men because women were living longer despite having a higher prevalence of disabilities. The women's disability disadvantage was not able to cancel out the women's mortality advantage in any of the cities in our sample, although it came close in Mexico City (Mexico) and Santiago (Chile). On the other hand, women were living more years with disability (LEWD) than men because of the combination of two effects: women were living longer and had higher rates of disability. In three of the cities – Buenos Aires (Argentina), Mexico City (Mexico) and Santiago (Chile) – the gender gap in LEWD represented a larger fraction of the total gender gap in life expectancy than DFLE. In São Paulo, DFLE and LEWD were comparable in size. In Bridgetown, Havana and Montevideo, the gender gap in DFLE represented a larger fraction of the gender gap in LE than LEWD.

There are many potential explanations for the patterns we observed across LAC countries. As was previously mentioned, the differences in Chile could be attributed to evidence suggesting that the prevalence of osteoporosis is five times higher among older women than men (Albala et al. 2011). In Mexico, the combination of a high prevalence of disability and lower mortality among women may be connected at least in part to the high shares of women who have diabetes, especially at ages 70–75 (29%). Similarly, previous studies have shown that the mean BMI was higher for women than for men, and that the percentage of women with obesity was significantly higher than that of men (26.7% vs. 22.3%) (Wong et al. 2015). Taken together, these factors could explain why in these cities, the gender gap in LEWD contributed more than DFLE to the total gender gap in life expectancy.

On the other hand, Brazil, Cuba, Uruguay and Barbados were found to be the countries where the women's mortality advantage contributed the most to the gender gap in the years lived with disabilities. As we previously noted, in Brazil, this may be related to the gender gap in mortality by disability severity, or to the levels of life-threatening gender-specific conditions. Compared to their male peers, Brazilian women live longer lives, but are more likely to be disabled, and to have severe disabilities (Camargos et al. 2008; Andrade et al. 2011). Thus, in Brazil, women at age 60 can expect to live two years longer with severe limitations than men at the same age (Belon et al. 2014; Nepomuceno and Turra 2012; Parahyba et al. 2005). Similarly, the gender gap in unhealthy life-years in Cuba may be due to the high prevalence levels of acute myocardial infarction among men, and to men's increased risk of dying from this cause relative to that of women of the same age (Armas et al. 2012), together with the higher lung cancer incidence and associated mortality risk among older men (Galán et al. 2009). Surprisingly, the finding that in Uruguay the women's mortality advantage contributed to the gender gap in life years lived with disability may be due to a stronger disadvantage in mortality among men than to a mortality advantage among women. As estimates in Figure 1 showed, the remaining life expectancy at age 60 for women in Montevideo (Uruguay) was the second-highest among all of the cities in the sample (approximately 23 years), but the remaining life expectancy at age 60 for men was the third-lowest for all of the cities (approximately 18 years). Thus, further investigation is needed to understand the low remaining LE at age 60 among men, given the high remaining LE among women of the same age. It is also noteworthy that Uruguay had one of the lowest rates of response in the SABE (66%), while the rates for the other countries varied between 85% and 95%. It is unclear whether this lower response rate could have biased some of the results. However, the fact that the socioeconomic indicators were better for Uruguay than for the other countries suggests that the lower disability effects found in Uruguay were not mainly attributable to response rates.

Finally, we should note that reports on the SABE have mentioned that Uruguay is the only country included in the SABE where the proportion of older people living in institutions is higher than it is in the other countries. Since the SABE samples are household-based and exclude those in institutions, there may be a selection among those who remain independent instead of becoming institutionalized that affects the health measures (Palloni and McEniry 2007). However, since the main source of long-term care services in LAC is unpaid care that family members, primarily women, provide to care-dependent older people, it is unlikely that this issue played an important role in our results. Indeed, in all seven cities in our sample, the shares of older adults living alone were lower than the shares in North America and Western Europe (Palloni et al. 2002; DeVos and Palloni 2002).

In Bridgetown (Barbados), like in Montevideo (Uruguay), our findings that there was a large gender gap in life expectancy, and that a large proportion of the gap was spent without disability (61%), may be attributable to a larger male disadvantage, since it has been reported that the region has struggled in the past decade with a very large burden of deaths from liver cirrhosis attributed to alcohol abuse and heavy smoking among men (Moonie and Quashie 2011).

Explanations for these contradictory findings regarding health and mortality among women and men – or the so-called health-survival paradox – have been mainly developed from the perspective of excess mortality among males (due to biological and acquired risks) or excess disability among females (due to types and severity of illnesses and disability, illness and prevention orientation and health reporting behavior). The evidence for these explanations is still inconclusive and conflicting, with studies showing either that male excess mortality is more important or that female excess morbidity is the more prominent factor (Austad 2006; Austad 2011; Grundy 2006).

Our work adds to this body of literature by showing that such an ambiguous relationship was not only present in the Latin American and the Caribbean context, but also that the relative contributions of women's morbidity and men's mortality varied substantially throughout the region. The results suggest that policy-makers seeking to address health inequalities should consider not only the total gender gap, but also whether those differences are related to health or mortality factors.

Acknowledgements

This study was supported by the European Research Council within the EU Framework Programme for Research and Innovation Horizon 2020, ERC Grant Agreement No. 716323; the European Research Council within the EU Framework Programme for Research and Innovation Horizon 2020, ERC Grant Agreement No. 725187 (LETHE); the Brazilian Graduate Studies Coordinating Board (Capes, Code 001), which funds the Demography Program at the Federal University of Minas Gerais; and the Brazilian National Research Council (CNPq).

References

Albala, C., M. L. Lebrao, E. M. Léon Díaz, R. Ham-Chande, A. J. Hennis, A. Palloni, M. Peláez and O. Pratts 2005. The health, well-being, and aging ("SABE") survey: Methodology applied and profile of the study population. *Revista Panamericana de Salud Publica/Pan American Journal of Public Health* 17: 307–322. https://doi.org/10.1590/s1020-49892005000500003

Albala, C., H. Sánchez, L. Lera, B. Angel and X. Cea 2011. Efecto sobre la salud de las desigualdades socioeconómicas en el adulto mayor. Resultados basales del estudio expectativa de vida saludable y discapacidad relacionada con la obesidad (Alexandros). *Revista Medica de Chile* 139: 1276–1285. https://doi.org/10.4067/S0034-98872011001000005

Acosta, E., C. Torres, R. Silva-Ramírez and R. Bourbeau 2018. Violence and mortality from external causes in Colombia: Analysis of demographic costs during the period 1979–2016. Paper presented at the VIII Congresso de La Asociación Latinoamericana de Poblacíon, Mexico - Puebla 2018. http://www.alapop.org/Congreso2018/PDF/0021a.pdf

Andrade, F. C. D., P. E. Guevara, M. L. Lebrão, Y. A. de Oliveira Duarte and J. L. F. Santos 2011. Gender differences in life expectancy and disability-free life expectancy among older adults in São Paulo, Brazil. *Women's Health Issues* 21: 64–70. https://doi.org/10.1016/j.whi.2010.08.007

Andrade, F. 2009. Estimating diabetes and diabetes-free life expectancy in Mexico and seven major cities in Latin America and the Caribbean. *Revista Panamericana de Salud Publica/Pan American Journal of Public Health* 26(1): 9–16. https://doi.org/10.1590/s1020-49892009000700002

Andrade, F. C. D., L. P. Corona, M. L. Lebrão and Y. A. O. de Duarte 2014. Life expectancy with and without cognitive impairment among Brazilian older adults. *Archives of Gerontology and Geriatrics* 58(2): 219–225. https://doi.org/10.1016/j.archger.2013.10.007

Angel, J. L., W. Vega, M. López-Ortega and R. Pruchno 2017. Aging in Mexico: Population trends and emerging issues. *Gerontologist* 57(2): 153–162. https://doi.org/10.1093/geront/gnw136

Armas, N. B., Y. Y. Ortega, R. de la Noval, R. Suárez, L. Llerena and A. F. Dueñas 2012. Acute myocardial infarction mortality in Cuba, 1999-2008. *MEDICC Review* 14: 19–25. https://doi.org/10.37757/MR2012V14.N4.6

Austad, S. N. 2011. Sex differences in longevity and aging. In: *Handbook of the Biology of Aging*, eds E. J. Masoro and S. N. Austad, 479–495. Academic Press. https://doi.org/10.1016/B978-0-12-378638-8.00023-3

Austad, S. N. 2006. Why women live longer than men: Sex differences in longevity. *Gender Medicine* 3(2): 79–92. https://doi.org/10.1016/S1550-8579(06)80198-1

Barbosa, A. R., J. M. P. Souza, M. L. Lebrão, R. Laurenti and M. de Fátima Nunes Marucci 2005. Functional limitations of Brazilian elderly by age and gender differences: data from SABE Survey. *Cadernos de Saúde Pública/Reports in Public Health* 21: 1177–1185. https://doi.org/10.1590/S0102-311X2005000400020

Barragán-Martínez, C., J. Amaya-Amaya, R. Pineda-Tamayo, R. D. Mantilla, J. Castellanos-De La Hoz, S. Bernal-Macías, A. Rojas-Villarraga and J. M. Anaya 2012. Gender differences in latin-american patients with rheumatoid arthritis. *Gender Medicine* 9(6): 490–510.e5. https://doi.org/10.1016/j.genm.2012.10.005

Beckett, L. A., D. B. Brock, J. H. Lemke, C. F. Mendes de Leon, J. M. Guralnik, G. G. Fillenbaum, L. G. Branch, T. T. Wetle and D. Evans 1996. Analysis of change in self-reported physical function among older persons in four population studies. *American Journal of Epidemiology* 143(8): 766–78. https://doi.org/10.1093/oxfordjournals.aje.a008814

Belon, A. P., M. G. Lima and M. B. A. Barros 2014. Gender differences in healthy life expectancy among Brazilian elderly. *Health Quality Life Outcomes* 12(88). https://doi.org/10.1186/1477-7525-12-88

Camargos, M. C. S., C. J. Machado and R. N. Rodrigues 2007. Disability life expectancy for the elderly, city of São Paulo, Brazil, 2000: Gender and educational differences. *Journal of Biosocial Science* 39(3): 455–463. https://doi.org/10.1017/S0021932006001428

Camargos, M. C. S., C. J. Machado and R. N. Rodrigues 2008. Life expectancy among elderly Brazilians in 2003 according to different levels of functional disability. *Cadernos de Saúde Pública/Reports in Public Health* 24: 845–852. https://doi.org/10.1590/S0102-311X2008000400015

Campos, A. C. V., C. Albala, L. Lera, H. Sánchez, A. M. D. Vargas and E. F. E. Ferreira 2015. Gender differences in predictors of self-rated health among older adults in Brazil and Chile. *BMC Public Health* 15(365): 1–11. https://doi.org/10.1186/s12889-015-1666-9

Canudas-Romo, V. and J. M. Aburto 2019. Youth lost to homicides: Disparities in survival in Latin america and the caribbean. *BMJ Global Health* 4(2). https://doi.org/10.1136/bmjgh-2018-001275

Case, A. and C. Paxson 2005. Sex differences in morbidity and mortality. *Demography* 42: 189–214. https://doi.org/10.1353/dem.2005.0011

Costa Filho, A. M., J. Vaz de Melo Mambrini, D. C. Malta, M. F. Lima-Costa and S. V. Peixoto 2018. Contribution of chronic diseases to the prevalence of disability in basic and instrumental activities of daily living in elderly Brazilians: The National Health Survey 2013. *Cadernos de Saúde Pública/Reports in Public Health* 34: 1–12. https://doi.org/10.1590/0102-311x00204016

Crimmins, E. M., M. D. Hayward and Y. Saito 1996. Differentials in active life expectancy in the older population of the United States. *The Journal of Gerontology: Series B* 51B(3): S111–S120. https://doi.org/10.1093/geronb/51B.3.S111

Crimmins, E. M. and J. K. Kim 2010. Gender differences in health: Results from SHARE, ELSA and HRS. *European Journal of Public Health* 21: 81–91. https://doi.org/10.1093/eurpub/ckq022

Crimmins, E. M., J. K. Kim and A. Hagedorn 2002. Life with and without disease: Women experience more of both. *Journal of Women & Aging* 14(1–2): 47–59. https://doi.org/10.1300/J074v14n01_04

Crimmins, E. M. and Y. Saito 2001. Trends in healthy life expectancy in the United States, 1970–1990: Gender, racial, and educational differences. *Social Science and Medicine* 52: 1629–41. https://doi.org/10.1016/s0277-9536(00)00273-2

Crimmins, E. M., Y. Zhang and Y. Saito 2016. Trends over 4 decades in disability-free life expectancy in the United States. *American Journal of Public Health* 106(7): 1287–1293. https://doi.org/10.2105/AJPH.2016.303120

DeVos, S. and A. Palloni 2002. *Living arrangements of elderly people around the world*. Madison: University of Wisconsin-Madison, Center for Demography & Ecology.

di Lego, V., P. Di Giulio and M. Luy 2020. Gender differences in healthy and unhealthy life expectancy. *International Handbook of Health Expectancies. International Handbooks of Population*, eds C. Jagger, E. M. Crimmins, Y. Saito, R. T. C. Yokota, H. Van Oyen and J.-M. Robine, 151–172. Cham: Springer. https://doi.org/10.1007/978-3-030-37668-0_11

Duarte, Y. A. O. D, M. L. Lebrão, F. D. Lima 2005. Contribuição dos arranjos domiciliares para o suprimento de demandas assistenciais dos idosos com comprometimento functional em São Paulo, Brasil. *Panamericana de Salud Publica/Pan American Journal of Public Health*. 17(5/6): 370–378.

Franco, M., J. F. Kennelly, R. S. Cooper and P. Ordúñez-García 2007a. Health in Cuba and the millennium development goals. *Revista Panamericana de Salud Publica/Pan American Journal of Public Health* 21: 239–250. https://doi.org/10.1590/s1020-49892007000300007

Franco, M., P. Orduñez, B. Caballero, J. A. Tapia Granados, M. Lazo, J. L. Bernal, E. Guallar and R. S. Cooper 2007b. Impact of energy intake, physical activity, and population-wide weight loss on cardiovascular disease and diabetes mortality in Cuba, 1980–2005. *American Journal of Epidemiology* 166(12): 1374–1380. https://doi.org/10.1093/aje/kwm226

Galán, Y., L. Fernández, P. Torres and M. García 2009. Trends in Cuba's cancer incidence (1990 to 2003) and mortality (1990 to 2007). *MEDICC Review* 11(3): 19–26. https://doi.org/10.37757/MR2009V11.N3.6

Gorman, B.K. and J. G. Read 2006. Gender disparities in adult health: An examination of three measures of morbidity. *Journal of Health and Social Behavior* 47(2): 95–110. https://doi.org/10.1177/002214650604700201

Green, C. A. and C. R. Pope 1999. Gender, psychosocial factors and the use of medical services: A longitudinal analysis. *Social Science and Medicine* 48(10): 1363–72. https://doi.org/10.1016/S0277-9536(98)00440-7

Grundy, E. 2006. Gender and healthy aging. In: *Longer Life and Healthy Aging*, eds Y. Zeng, E. M. Crimmins, Y. Carriere and J.-M. Robine, 173–199. Dordrecht: Springer. https://doi.org/10.1007/1-4020-4032-6_12

Guh, D. P., W. Zhang, N. Bansback, Z. Amarsi, C. L. Birmingham and A. H. Anis 2009. The incidence of co-morbidities related to obesity and overweight: A systematic review and meta-analysis. *BMC Public Health* 9: 88. https://doi.org/10.1186/1471-2458-9-88

Hennis, A., I. Hambleton, H. Broome, S. Crichlow and H. Fraser 2005. *Health, welfare and aging in Bridgetown, Barbados: SABE 2000*. Pan American Health Organization.

Horiuchi, S., J. R. Wilmoth, and S. D. Pletcher 2008. A decomposition method based on a model of continuous change. *Demography* 45: 785–801. https://doi.org/10.1353/dem.0.0033

Huenchuan, S. 2013. *Ageing, solidarity and social protection in Latin America and the Caribbean: Time for progress towards equality*. Libros de la CEPAL, 117. https://repositorio.cepal.org/handle/11362/2620

International Health Conference. 2002. Constitution of the World Health Organization. 1946. *Bulletin of the World Health Organization* 80(12): 983–984. World Health Organization. https://apps.who.int/iris/handle/10665/268688

Jacobsen, R., A. Oksuzyan, H. Engberg, B. Jeune, J. W. Vaupel and K. Christensen 2008. Sex differential in mortality trends of old-aged Danes: A nation wide study of age, period and cohort effects. *European Journal of Epidemiology* 23(11): 723–730. https://doi.org/10.1007/s10654-008-9288-5

Jagger, C. and F. Matthews 2002. Gender differences in life expectancy free of impairment at older ages. *Journal of Women & Aging* 14: 85–97. https://doi.org/10.1300/J074v14n01_06

Kearns, K., A. Dee, A. P. Fitzgerald, E. Doherty and I. J. Perry 2014. Chronic disease burden associated with overweight and obesity in Ireland: The effects of a small BMI reduction at population level. *BMC Public Health* 14(143). https://doi.org/10.1186/1471-2458-14-143

Keevil, V. L., S. Hayat, N. Dalzell, S. Moore, A. Bhaniani, R. Luben, N. J. Wareham and K.-T. Khaw 2013. The physical capability of community-based men and women from a British cohort: The European Prospective Investigation into Cancer (EPIC)-Norfolk study. *BMC Geriatrics* 13(93). https://doi.org/10.1186/1471-2318-13-93

Knight, F. W., J. Sued Badillo, K. O. Laurence, J. Ibarra, B. Brereton and B. W. Higman 2004. *General history of the Caribbean*. Volume 5, Unesco.

Kulminski, A. M., I. V. Culminskaya, S. V. Ukraintseva, K. G. Arbeev, K. C. Land and A. I. Yashin 2008. Sex-specific health deterioration and mortality: The morbidity-mortality paradox over age and time. *Experimental Gerontology* 43(12): 1052–1057. https://doi.org/10.1016/j.exger.2008.09.007

Leong, D. P., K. K. Teo, S. Rangarajan, P. Lopez-Jaramillo, A. Avezum, A. Orlandini, P. Seron, S. H. Ahmed, A. Rosengren, R. Kelishadi, O. Rahman, S. Swaminathan, R. Iqbal, R. Gupta, S. A. Lear, A. Oguz, K. Yusoff, K. Zatonska, J. Chifamba, E. Igumbor, V. Mohan, R. M. Anjana, H. Gu, W. Li and S. Yusuf 2015. Prognostic value of grip strength: Findings from the Prospective Urban Rural Epidemiology (PURE) study. *Lancet* 386(9990): 266–273. https://doi.org/10.1016/S0140-6736(14)62000-6

Lebrão, M. L., Y. A. O. Duarte, J. L. F. S. Santos and N. N. Silva 2018. 10 years of SABE Study: Background, methodology and organization of the study. *Revista Brasileira de Epidemiologia* 21(2). https://doi.org/10.1590/1980-549720180002.supl.2

Luy, M. and Y. Minagawa 2014. Gender gaps–Life expectancy and proportion of life in poor health. *Health Reports* 25(12): 12–9.

Mairey, I., P. Bjerregaard and H. Brønnum Hansen 2014. Gender difference in health expectancy trends in Greenland. *Scandinavian Journal of Public Health* 42(8): 751–758. https://doi.org/10.1177/1403494814550174

Mathers, C., J. Salomon and A. Tandon 2001. Estimates of healthy life expectancy for 191 countries in the year 2000: Methods and results. World Health Organization 2001.

Monteverde, M., K. Noronha and A. Palloni 2009. Effect of early conditions on disability among the elderly in Latin America and the Caribbean. *Population Studies* 63(1): 21–35. https://doi.org/10.1080/00324720802621583

Moonie, S. and N. Quashie 2011. Social health protection for the elderly in the English-speaking Caribbean. Series Studies and Perspectives de la CEPAL. https://repositorio.cepal.org/bitstream/handle/11362/5052/1/S2012131_en.pdf

Moreno, X., C. Albala, L. Lera, B. Leyton, B. Angel and H. Sánchez 2018. Gender, nutritional status and disability-free life expectancy among older people in Santiago, Chile. *PLoS ONE* 13(3). https://doi.org/10.1371/journal.pone.0194074

Murray, C. J. L., J. A. Salomon, C. D. Mathers, P. J. Van Der Maas, A. D. Lopez and J. Richardson 2002. *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*. Geneva: World Health Organization. https://apps.who.int/iris/handle/10665/42439

Murtagh, K. N. and H. B. Hubert 2004. Gender differences in physical disability among an elderly cohort. *American Journal of Public Health* 94(8): 1406–1411. https://doi.org/10.2105/AJPH.94.8.1406

Nepomuceno, M. R. and C. M. Turra 2012. Expectativa de vida saudável no Brasil com base no método intercensitário. In: *Anais do V Congresso de La Asociación Latinoamericana de Poblacíon*, Montevideo, Uruguay, 23–26 October 2012 http://www.alapop.org/Congreso2012/DOCSFINAIS_PDF/ALAP_2012_FINAL479.pdf

Noronha, K. V. M. and M. A. Viegas 2005. Social inequality in health and the utilization of health services among the elderly in Latin America. *Revista Panamericana de Salud Publica/Pan American Journal of Public Health* 17: 410–418. https://doi.org/10.1590/s1020-49892005000500013

Nusselder, W. J. and C. W. N. Looman 2004. Decomposition of differences in health expectancy by cause. *Demography* 41: 315–334. https://doi.org/10.1353/dem.2004.0017

Nusselder, W. J., C. W. N. Looman, H. van Oyen, J. M. Robine and C. Jagger 2010. Gender differences in health of EU10 and EU15 populations: The double burden of EU10 men. *European Journal of Ageing* 7: 219–227. https://doi.org/10.1007/s10433-010-0169-x

Oksuzyan, A., H. Brønnum-Hansen and B. Jeune 2010. Gender gap in health expectancy. *European Journal of Ageing* 7(4): 213–218. https://doi.org/10.1007/s10433-010-0170-4

Oksuzyan, A., K. Juel, J. W. Vaupel and K. Christensen 2008. Men: Good health and high mortality. Sex differences in health and aging. *Aging Clinical and Experimental Research* 20(2): 91–102. https://doi.org/10.1007/BF03324754

Oksuzyan, A., M. Shkolnikova, J. W. Vaupel, K. Christensen and V. M. Shkolnikov 2014. Sex differences in health and mortality in Moscow and Denmark. *European Journal of Epidemiology* 29: 243–252. https://doi.org/10.1007/s10654-014-9893-4

Palloni, A. and M. McEniry 2007. Aging and health status of elderly in Latin America and the Caribbean: Preliminary findings. *Journal of Cross-Cultural Gerontology* 22: 263–285. https://doi.org/10.1007/s10823-006-9001-7

Palloni, A., G. Pinto-Aguirre and M. Pelaez 2002. Demographic and health conditions of ageing in Latin America and the Caribbean. *International Journal of Epidemiology* 31: 762–771. https://doi.org/10.1093/ije/31.4.762

Parahyba, M. I., R. Veras and D. Melzer 2005. Disability among elderly women in Brazil. *Revista de Saude Publica* 39: 383–391. https://doi.org/10.1590/s0034-89102005000300008

Pelaez, M., A. Palloni, C. Albala, J. C. Alfonso, R. Ham-Chande, A. Hennis and O. Prats 2006. SABE—Survey on Health, Well-Being, and Aging in Latin America and the Caribbean. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 17 February 2006. https://doi.org/10.3886/ICPSR03546.v1

Redondo-Sendino, Á., P. Guallar-Castillón, J. Banegas and F. Rodríguez-Artalejo 2006. Gender differences in the utilization of health-care services among the older adult population of Spain. *BMC Public Health* 6: 155. https://doi.org/10.1186/1471-2458-6-155

Rieker, P. P. and C. E. Bird 2005. Rethinking gender differences in health: Why we need to integrate social and biological perspectives. *Journals Gerontology Serie B* 60B: 40–47. https://doi.org/10.1093/geronb/60.Special_Issue_2.S40

Riffe, T. 2018. DemoDecomp: Decompose Demographic Functions. R Package version 101. https://cran.r-project.org/web/packages/DemoDecomp/DemoDecomp.pdf

Robine, J.-M., C. Jagger and I. Romieu 2001. Disability-free life expectancies in the European Union countries: calculation and comparisons. *Genus* 57: 89–101. https://www.jstor.org/stable/29788693

Robine, J.-M. and J. W. Vaupel 2001. Supercentenarians: Slower ageing individuals or senile elderly? *Experimental Gerontology* 36(4–6): 915–930. https://doi.org/10.1016/S0531-5565(00)00250-3

Robine, J., Y. Saito and C. Jagger 2009. The relationship between longevity and healthy life expectancy. *Quality in Ageing and Older Adults* 10(2): 5–14. https://doi.org/10.1108/14717794200900012

Roe, C. M., A. M. McNamara and B. R. Motheral 2002. Gender- and age-related prescription drug use patterns. *Annals of Pharmacotherapy* 36(1): 30–39. https://doi.org/10.1345/aph.1A113

Sanderson, W. C. and S. Scherbov 2014. Measuring the speed of aging across population subgroups. *PLoS ONE* 9: 3–6. https://doi.org/10.1371/journal.pone.0096289

Spiers, N., C. Jagger, M. Clarke and A. Arthur 2003. Are gender differences in the relationship between self-rated health and mortality enduring? Results from three birth cohorts in Melton Mowbray, United Kingdom. *Gerontologist* 43(3): 406–411. https://doi.org/10.1093/geront/43.3.406

Sullivan, D. 1971. A single index of mortality and morbidity. *HSMHA Health Reports* 86: 347–54.

Turra, C. M. and F. Fernandes 2020. Demographic transition: Opportunities and challenges to achieve the Sustainable Development Goals in Latin America and the Caribbean, Project Documents, (LC/TS.2020/105), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

UNECE/European Commission 2015. Active Ageing Index 2014: Analytical Report.

United Nations, Department of Economic and Social Affairs, Population Division 2019. World Population Prospects 2019. https://population.un.org/wpp/

United Nations, Department of Economic and Social Affairs, Population Division 2017. World Population Ageing 2017. https://www.un.org/en/development/desa/population/theme/ageing/WPA2017.asp

United Nations Development Programme 2015. Transforming our world: The 2030 Agenda for Sustainable Development. https://sdgs.un.org/2030agenda

Van Oyen, H., B. Cox, C. Jagger, E. Cambois, W. Nusselder, C. Gilles and J.-M. Robine 2010. Gender gaps in life expectancy and expected years with activity limitations at age 50 in the European Union: Associations with macro-level structural indicators. *European Journal of Ageing* 7: 229–237. https://doi.org/10.1007/s10433-010-0172-2

Van Oyen, H., W. Nusselder, C. Jagger, P. Kolip, E. Cambois and J.-M. Robine 2013. Gender differences in healthy life years within the E.U.: An exploration of the "health-survival" paradox. *International Journal of Public Health* 58: 143–155. https://doi.org/10.1007/s00038-012-0361-1

van Raalte, A. A. and M. R. Nepomuceno 2020. Decomposing gaps in healthy life expectancy. In: *International handbook of health expectancies. International handbooks of population*, eds C. Jagger, E. M. Crimmins, Y. Saito, R. T. C. Yokota, H. Van Oyen and J.-M. Robine, 107–122. Cham: Springer. https://doi.org/10.1007/978-3-030-37668-0_7

Verbrugge, L. M. 1989. The twain meet: Empirical explanations of sex differences in health and mortality. *Journal of Health and Social Behavior* 30(3): 282–304. https://psycnet.apa.org/doi/10.2307/2136961

Verbrugge, L. M. 1985. Gender and health: An update on hypotheses and evidence. *Journal of Health and Social Behavior* 26(3): 156–182. https://doi.org/10.2307/2136750

Wallace, S. P. and V. F. Gutiérrez 2005. Equity of access to health care for older adults in four major Latin American cities. *Revista Panamericana de Salud Publica/Pan American Journal of Public Health* 17: 394–409. https://doi.org/10.1590/S1020-49892005000500012

Wang, Z., Y. Zeng, B. Jeune and J. W. Vaupel 1998. Age validation of Han Chinese centenarians. *Genus* 54: 123–41. https://www.jstor.org/stable/29788544

Wong, R., A. Michaels-Obregon, A. Palloni, L. M. Gutierrez-Robledo, C. Gonzalez-Gonzalez, M. Lopez-Ortega and M. M. Tellez-Rojo 2015. Progression of aging in Mexico: The Mexican Health and Aging Study (MHAS) 2012. *Salud Publica de Mexico* 57: S79–S89. https://doi.org/10.21149/spm.v57s1.7593

Wong, R., M. Peláez and A. Palloni 2005. Self-reported general health in older adults in Latin America and the Caribbean: Usefulness of the indicator. *Revista Panamericana de Salud Publica/Pan American Journal of Public Health* 17: 323–332. https://doi.org/10.1590/s1020-49892005000500004

World Health Organization. 2004. International plan of action on ageing: Report on implementation report by the secretariat. https://apps.who.int/gb/archive/pdf_files/EB115/B115_29-en.pdf

Yokota, R. T. C., W. J. Nusselder, J.-M. Robine, J. Tafforeau, F. Renard, P. Deboosere and H. Van Oyen 2019. Contribution of chronic conditions to gender disparities in health expectancies in Belgium, 2001, 2004 and 2008. *European Journal of Public Health* 29: 82–87. https://doi.org/10.1093/eurpub/cky105

Yong, V., Y. Saito and A. Chan 2010. Changes in the prevalence of mobility limitations and mobile life expectancy of older adults in Singapore, 1995–2005. *Journal of Aging and Health* 22(1): 120–140. https://doi.org/10.1177/0898264309351932

Zunzunegui, M. V., B. E. Alvarado, F. Béland and B. Vissandjee 2009. Explaining health differences between men and women in later life: A cross-city comparison in Latin America and the Caribbean. *Social Science and Medicine* 68: 235–242. https://doi.org/10.1016/j.socscimed.2008.10.031

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 215–254

Gender and socioeconomic inequalities in health and wellbeing across age in France and Switzerland

***Anna Barbuscia*[1,2,*] *and Chiara Comolli*[2]**

Abstract

There is increasing evidence that wellbeing is unequally distributed across sociodemographic groups in contemporary societies. However, less is known about the divergence across social groups of trajectories of wellbeing across age groups. This issue is of great relevance in contexts characterised by changing population structures and growing imbalances across and within generations, and in which ensuring that everyone has the opportunity to have a happy and healthy life course is a primary welfare goal. In this study, we investigate wellbeing trends in France and Switzerland across age, gender, and socioeconomic status groups. We use two household surveys (the Santé et Itinéraires Professionnels and the Swiss Household Panel) to compare the unfolding inequalities in health and wellbeing across age groups in two rich countries. We view wellbeing as multidimensional, following the literature highlighting the importance of considering different dimensions and measures of wellbeing. Thus, we investigate a number of outcomes, including different measures of physical and mental health, as well as of relational wellbeing, using a linear regression model and a linear probability model. Our findings show interesting country and dimension-specific heterogeneities in the development of health and wellbeing over age. While our results indicate that there are gender and educational inequalities in both Switzerland and France, and that gender inequalities in mental health accumulate with age in both countries, we also find that educational inequalities in health and wellbeing remain rather stable across age groups.

Keywords: multidimensional wellbeing; sociodemographic inequalities; age development; cross-country comparison

[1]INED, Paris, France

[2]University of Lausanne, Switzerland

[*]Correspondence to: Anna Barbuscia, anna.barbuscia@ined.fr

DOI: 10.1553/populationyearbook2021.res2.2

1 Introduction

In contexts characterised by changing population structures and growing imbalances across and within generations and social groups, ensuring a high quality of life and a healthy life course development for everyone becomes a paramount welfare objective. Despite the abundant literature on health and wellbeing, how these indicators vary with age remains the subject of theoretical and empirical debates. While there is consistent evidence that physical health tends to worsen with age, patterns of mental health and subjective wellbeing are less clear, with a majority of studies showing that both might actually be lower among mid-age groups (Lang et al. 2011; Blanchflower and Oswald 2016). However, other studies have reported that wellbeing increases with age (Walker 2005; Frijters and Beatton 2012), even though physical health worsens, or that levels of wellbeing do not change at all across age groups (De Neve et al. 2012). The few studies that have considered domain-specific satisfaction have tended to show that age trajectories of satisfaction diverge considerably across domains (McAdams et al. 2012; Easterlin 2006).

In parallel, a growing number of studies in the fields of economics, psychology, sociology and gerontology have focused on how wellbeing is (unequally) distributed across sociodemographic groups in contemporary societies, and have shown that inequalities are increasingly being observed in a number of countries (Mackenbach 2012; Townsend and Davidson 1982; Elo 2009). A rich literature on health inequalities using different measures of both self-reported and objective health has reported that individuals with higher incomes and educational levels tend to experience better health over their life course (e.g., Mackenbach et al. 1997; Marmot 2005), albeit with important differences across countries (Kunst et al. 1995). While there are indications that gender differences in overall health might be smaller than was previously thought (Arber and Cooper 1999; Oksuzyan et al. 2019), women generally report having lower levels of physical and mental health than men (Crimmins and Saito 2001; Dahlin and Härkönen 2013; Lee et al. 2016). It also appears that subjective wellbeing is heterogeneously distributed across social groups. There is, for example, extensive evidence that higher education tends to be associated with greater happiness, as well as with better health. However, findings on gender differences in life satisfaction have been more contradictory, with some studies showing that women tend to be happier than men (Blanchflower and Oswald 2004; Easterlin 2001), and others reaching the opposite conclusion (Pinquart and Sörensen 2001).

Moreover, even less is known about the divergence of age trajectories of health and wellbeing across those social groups. According to the cumulative disadvantage theory, early inequalities may be exacerbated by the ageing process, causing well-being trajectories to diverge over the life course (Dannefer 2003). Well-educated individuals are less likely to face risky situations and are less vulnerable when they experience adverse life course events such as job loss, illness, or financial strain. Educational achievement equips individuals with positive attitudes that are useful for the maintenance of health across the life course. Furthermore, being employed favours social integration, and can increase an individual's social support and opportunities (Loscocco and Spitze 1990; Cerci and Dumludag 2019). By contrast, the age-as-leveller hypothesis argues that the ageing process could smooth initial differences across groups if the resources that generated the inequality in the first place become less useful over the life course (Lynch 2003). In addition, the survival selection process means that the happier and healthier individuals in each group live longer, and thus participate in surveys longer (Ulloa et al. 2013; Kratz and Patzina 2020). This implies that the group of survivors at older ages are more homogenous. The empirical findings reflect this theoretical duality: some studies have found that gender and educational gaps widen with age (Pinquart and Sörensen 2001), while others have reported constant or shrinking inequalities (Yang 2008).

Finally, a number of studies have highlighted the importance of broadly interpreting the concept of wellbeing as being composed of a number of factors (Pollard and Lee 2003; Cronin de Chavez et al. 2005; Huppert 2013; Infurna and Luthar 2017). A satisfactory ageing process entails not just having high levels of subjective, psychological, relational, and financial wellbeing; but also remaining in good health. A number of investigations have compared multidimensional measures with standard satisfaction questions, and most found only small or moderate correlations (e.g., Huppert and So 2013; Ryff and Keyes 1995). Thus, measurements of wellbeing may not be reducible to a simple, unidimensional notion such as life satisfaction without losing a great deal of potentially valuable information (Keyes 2007). However, most of the work on wellbeing has been undertaken within single disciplines, and has tended to focus on one aspect of the concept, rather than on drawing together the physical, psychological and socioeconomic aspects of wellbeing (Cronin de Chavez et al. 2005).

In this study, we contribute to the literature on unequal age trends in health and wellbeing across gender and socioeconomic status by illustrating how these trends unfold in two different contexts using data from France and Switzerland. While both of these European countries are considered conservative welfare states (Esping-Andersen 1990), they differ in a number of aspects that shape inequalities in sociodemographic health and wellbeing. First, compared to the French welfare state, the Swiss welfare state shares many more features with the liberal welfare system, especially in terms of its privatised health care system. Second, gender norms and the related family and labour market institutions tend to be much more traditional and oriented towards the male breadwinner model (and towards greater financial insecurity among women) in Switzerland than in France. Interesting differences in the health and wellbeing profiles of the two countries can be observed, which provide us with important insights into the cultural and institutional differences in their determinants, and again highlight the importance of considering the multiple dimensions of quality of life (Huppert et al. 2009). Because the questions on health and wellbeing in the two surveys used here were formulated in different ways, we cannot make direct cross-country comparisons. However, exploring the unfolding of health and wellbeing inequalities across age groups in two different contexts may enable us to identify the circumstances that are associated with growing or shrinking inequalities within countries (Kunst et al. 1995).

We use data from the 2006 waves of the Santé et intineraire professionnel (SIP) for France and the Swiss Household Panel (SHP) for Switzerland to study individuals of working ages. The two surveys provide a variety of indicators of health and wellbeing, as well as demographic and socioeconomic variables. We focus on four dimensions of physical and mental health and two domain-specific wellbeing measures: relational and professional wellbeing. Our study is descriptive in nature. Our main aim is to explore the gender and educational inequalities in the age trends in two different contexts, while focusing on less investigated indicators of health and wellbeing. We contribute to – and hope to partially reconcile – the streams of literature on the age-wellbeing nexus, the unfolding of wellbeing inequalities over the life course, and the contextual determinants of those age trends and heterogeneities.

2 Background

2.1 The social stratification of health and wellbeing across age

Sociology and social stratification studies have documented that wellbeing is unequally distributed across demographic and socioeconomic groups (Mackenbach 2012; Townsend and Davidson 1982; Elo 2009). In particular, health and wellbeing appear to differ between women and men. Although women are consistently found to have higher life expectancy than men, net of age, men seem to fare better than women with respect to objective and subjective measures of health. On average, women report lower self-rated health and a higher prevalence of mental health issues (e.g., depression and sleep disorders) than men (Crimmins and Saito 2001; Dahlin and Härkönen 2013; Lee et al. 2016; Troxel et al. 2010). Interestingly, men and women tend to diverge in terms of the types of mental and physical illnesses they have (Needham and Hill 2010). Moreover, the results on gender differences in subjective wellbeing are not consistent. Some studies have shown that women are (slightly) happier than men (Arrosa and Gandelman 2016), while others have found that men are happier and more satisfied with life at all ages (Pinquart and Sörensen 2001).

The positive association between higher education and both subjective wellbeing and physical and mental health is well established (Lleras-Muney 2005; Blanchflower and Oswald 2004; Easterlin 2001; Subramanian et al. 2005). Because having formal education develops people's competences on many levels, it increases their ability and motivation to control and shape their lives (Mirowsky and Ross 1998, 2007; Wheaton 1980). Having higher levels of education, human capital, self-esteem and self-efficacy directly increases people's emotional wellbeing by enabling them to develop emotional resilience and the ability to cope with adversity and stress. A person who has a strong sense of control over her life tends to have lower levels of physical and psychological distress, although her level of dissatisfaction with life may not be lower[1] (Ross and Van Willigen 1997; Edgerton et al. 2012; Hale 2005). Other factors linking formal education to health are having a greater command of the communication and social-psychological skills needed for building social contacts and constructing stable social relationships (McPherson et al. 2006) and lifestyle practices (Pampel et al. 2010). Finally, although education precedes and influences people's employment, earnings and income, its beneficial effects on people's health and wellbeing are mediated in part by the lower financial distress and the better and more secure employment conditions (Ross and Wu 1995) that having higher education tends to provide. Indeed, most studies have reported that there is a positive association between education and happiness even when income is controlled for (Blanchflower and Oswald 2004), and that highly educated individuals tend to choose types of activities – physical, social, cultural – that positively affect life satisfaction (Fernández-Ballesteros et al. 2001).

The evidence on the question of how wellbeing inequalities unfold across age groups has been more mixed (Yang 2008). Theoretically, the development of wellbeing over the life course across groups is related to the following: first, structural differences in the starting points of groups; second, the existence of processes of accumulation of advantages/disadvantages; and, third, group differentials in exposure to life events that are also strongly correlated with age, such as marriage and childbearing, or the evolution of an individual's health status, labour market attachment and income. There is evidence that the most critical life events tend to affect women more negatively than men. For example, compared to men, women tend to face more challenges in reconciling work and family, and they are more likely to be widowed. In addition, critical events known to affect wellbeing (such as unemployment or divorce) tend to produce more negative outcomes for women than for men (Keizer et al. 2010; Madero-Cabib and Fasang 2016; Melchior et al. 2007; Troxel et al. 2010). If gender roles and behaviour shift during the life course, the wellbeing gap between men and women linked to those events also varies. Indeed, longitudinal studies have suggested that gender differences in happiness vary across age groups. It has, for example, been shown that prior to reaching middle age, women are happier than men; but that this pattern reverses as people grow older and experience particular events or transitions (Yang 2008). Thus, it may be the case that, on average, men and women start with an even distribution of wellbeing, but then experience gender-specific transitions differently and with different consequences, which causes their trajectories of wellbeing to diverge at older ages.

More comprehensive social stratification theories, such as the cumulative (dis)advantage (CAD) theory (Dannefer 2003; DiPrete and Eirich 2006), predict that disparities in wellbeing increase over the life course because early disadvantages

---

[1] Higher educational attainment might also enhance a person's expectations for achievement, and increase her standards when evaluating her life satisfaction.

accumulate with age. The CAD theory argues that the benefits associated with a person's structural position early in the life course tend to gradually accumulate over time, through a path-dependent process that links trajectories earlier in life to outcomes later in life. This implies that the social differences between groups tend to widen as they age (DiPrete and Eirich 2006; McDonough et al. 2015). Initial individual differences in race, gender or socioeconomic origin generate structures of opportunity and pathways that further differentiate individuals over time. Early life course heterogeneities can therefore become greater over the life course. By contrast, the "age-as-leveller" theory (Lynch 2003) posits that with age, the resources that generated and reinforced the disparities earlier in life matter less and less for life satisfaction, as people's attachment to the labour market and their social relationships weaken.

In their meta-analysis, Pinquart and Sörensen (2001) showed that the gender gap in subjective wellbeing and self-concept, although smaller than expected, increases with age. This gap has been attributed to the disadvantages women suffer in terms of their everyday competences, their health and socioeconomic conditions, and their greater likelihood of being widowed in old age. It has generally been shown that the self-rated and physical health trajectories of men and women differ substantially over the life course (McDonough et al. 2015; Oncini and Guetto 2018). There is also evidence suggesting that education-related health disparities grow across the life span, or at least until people reach their mid-sixties, when the divergence attenuates (Mirowsky and Ross 2007; Ross and Wu 1995; Prus 2004; Lee et al. 2016). Mirowsky and Ross (2005) argued that education put people onto a track that permeates all aspects of life. The cumulative beneficial effects of higher education on health over time are evident when we look at a range of factors: socioeconomic status, behaviours, psychological health, anatomical health and "perhaps intracellular" characteristics (2005: pp. 27). Moreover, these factors interact over the life course to amplify the beneficial effects of education (Mirowsky and Ross 2007). In contradiction to the body of literature that has supported the notion that disadvantages accumulate over the life course (Mayer 2009), and in line with the age-as-leveller theory, Yang (2008) showed that in the US, gender and educational (but not racial) inequalities in happiness decline with age. Yang (2008) argued that these social differences are attenuated by the differential exposures of these groups to correlates of happiness, such as retirement, widowhood and eligibility for social benefits. The loss of social support and integration tends to reduce these differences at older ages because it erodes the advantages of some groups relative to others (Yang 2008: pp. 221). Thus, it appears that the gender and educational gaps in subjective wellbeing decrease with rising age.

2.2 The French and Swiss contexts

Existing studies have documented that gender and educational health and wellbeing inequalities vary between countries (Huisman et al. 2003, Von dem Knesebeck etal.

Table 1:
Main demographic and labour market indicators. France and Switzerland (2018)

	France	**Switzerland**
Crude marriage rate (per 1000)	3.5	4.8
Total fertility rate	1.88	1.52
Out-of-wedlock births	60.4%	25.7%
Male unemployment	8.7%	4.3%
Female unemployment	8.8%	5.1%
Female labour force participation rate	74.4%	82.4%
Female part-time employment share	28.7%	63%
Government health expenditure (% GDP)	8.1%	2.2%

Source: Elaboration of the authors based on Eurostat 2020.

2006; Bambra et al. 2009). While inequalities in conservative welfare states – such as Switzerland and France – tend to be smaller than they are in other regimes, they are still present (Eikemo et al. 2008). Moreover, despite belonging to the same welfare regime classification, the two countries differ with respect to fundamental aspects related to inequality.

First, Switzerland and France differ in terms of their gendered family and labour market institutions. Table 1 displays the main demographic and labour market indicators for the two countries (Eurostat).[2] Marriage rates are higher in Switzerland (4.8 per 1000) than in France (3.5 per 1000), but while France's fertility rates are among the highest in Europe (1.88 children per woman in 2018), Switzerland's fertility rates are much lower, close to those of Eastern and Southern European countries (1.52 children per woman in 2018). However, in relation to family norms, the most striking difference between the two countries lies in the share of births that take place outside of wedlock (mostly in cohabitation): 60.4% in France versus 25.7% in Switzerland. When we look at the countries' labour markets, we see that overall unemployment rates are much lower in Switzerland than in France. However, whereas in France the male and female unemployment rates are equal (8.7% and 8.8%, respectively), in Switzerland, the female unemployment rate is 25% higher than the male unemployment rate (4.3% and 5.1%). Moreover, even though female labour force participation is quite high in both countries (82.4% in Switzerland and 74.4% in France), the share of women in part-time work in the two countries differs tremendously, at 63% in Switzerland and 28.7% in France.

[2] We report data for the most recent available year at the time of writing, 2018; but the differences between France and Switzerland were similar in 2006. For instance, the marriage rates were 4.3 (FR) and 5.3 (CH); the TFRs were 1.98 (FR) and 1.46 (CH); the male and the female unemployment rates were 8.2% (FR) and 4.5% (CH) and 9.5% (FR) and 5.2% (CH), respectively; and the government health expenditures were 7.6% (FR) and 1.7% (CH).

In Switzerland, a traditional gender model tends to be much more dominant, which implies that family responsibilities are borne almost entirely by women. Thus, compared to men, women in Switzerland end up having a weaker and irregular labour market attachment over their life course, which exposes them to greater financial insecurity (Widmer et al. 2003). In France, by contrast, the dual earner model is widespread, including among parents (Fagnani 2010). In Switzerland, the incentives for following a traditional male breadwinner–female caretaker model include gender-segregated labour markets, large gender wage gaps and high marginal tax rates that penalise second earners (Cooke and Baxter 2010; Peters 2014). Moreover, in Switzerland, public child care is limited and expensive, which creates significant trade-offs between employment and care time for mothers (Wall and Escobedo 2013). By contrast, the French welfare state has long been oriented towards the promotion of women's full employment through a number of family policies that allow families to pursue diverse employment-family reconciliation strategies, such as the provision of broad access to child care facilities (Fagnani 2010), as well as family benefits that encourage couples to have a large number of children (Pailhé et al. 2008).

Second, Switzerland has the most privatised health system in Europe. Based on universal private insurance, the Swiss system has one of the highest shares of out-of-pocket health expenditures in the OECD. France, by contrast, has a compulsory public insurance programme that covers the cost of medical treatment by private doctors, while limiting doctors' fees. Indeed, as Table 1 shows, government spending on the health care system is relatively high in France (8.1% of gross domestic product, GDP in 2018), and is relatively low in Switzerland (2.2% of GDP in 2018).

3 Data and method

In our analysis, we use data from the Santé et Itinéraire Professionnel (SIP) and the Swiss Household Panel (SHP). The SIP survey was conducted among individuals aged 20 to 74 in 2006 living in ordinary households in mainland France. Two waves (2006 and 2010) are available. The Swiss Household Panel is a random sample of private households in Switzerland in which all members of the family aged 14+ are interviewed annually. The SHP has 20 waves (1999–2018), but to ensure comparability, we used the 2006 wave for both surveys. We decided to use the 2006 wave instead of the 2010 wave for two main reasons. First, compared to the data from the 2010 wave, the French and Swiss panel data from the 2006 wave either are not or are less affected by attrition, and the samples are larger.[3] Second, while there

[3] After the initial sample of 1999 respondents, a refreshment sample was added to the SHP in 2004 to partially compensate for attrition, which was just two years before our wave of 2006. The SIP is also a panel, but given that we use its first wave, we do not have problems of attrition for the French data.

is no reason to expect to observe significant differences between the two waves since they are only few years apart, cross-sectional findings based on the 2010 wave could be affected by the financial crisis of 2008–2009. The recession had heterogenous effects on health and wellbeing across contexts and social groups that are not the focus of this study (Burgard and Kalousova 2015). We considered individuals of working ages (aged 25–69) in order to further limit the impact of attrition in the Swiss data, and because one of the dimensions of wellbeing we are interested in is professional satisfaction (N = 2136 for men and N = 2753 for women for Switzerland and N = 4367 for men and N = 5300 for women for France).

We focused on four outcomes of physical and mental health: self-reported health; daily activity limitations; depressive symptoms and sleep disorders; and two relational wellbeing indicators: satisfaction with social relationships (reliability of the social network in the SIP) and satisfaction with professional life. These measures were chosen because they provide information about similar health and wellbeing domains across the two surveys. However, as we mentioned, some important differences in terms of how the specific questions were phrased and the responses were coded persist, which limit their direct cross-country comparability. The first difference between the two surveys concerns the time frame the questions refer to. Table 2 reports the specific questionnaire formulations in the two surveys. In the French SIP, most of the health questions refer to a specific time frame, albeit one that varies across items: six weeks for depressive symptoms, six months for daily activity limitations and 12 months for sleeping problems. The Swiss SHP, by contrast, only uses a specific time frame (of four weeks) for sleeping disorders, while all of the other questions are posed in more general terms.

The second difference between the two surveys relates to the operationalisation of the responses. Table 3 reports the summary statistics of the main variables in the two surveys. For health status, the responses vary in the two surveys on a scale from one (very well) to five (not well at all), and are recoded in the opposite direction for both countries so that higher values mean better health. For work satisfaction, the responses vary on a scale of 0-10. All of the other indicators in the SIP data are binary (0/1). However, while sleep disorders were originally measured on a scale of 1–4, they are also used as a binary variable indicating whether the individual has had sleeping problems for at least several days a week. By contrast, in the SHP, the variables for satisfaction with social relationships, feelings of depression, and daily activity limitations are measured on a scale of 0–10, while sleep disorders are measured on a scale of one (not at all) to three (very much).

These phrasing and measurement differences limit the comparability of the levels of the indicators both between the two countries (e.g.: asking whether French women of different ages suffer from more or fewer sleeping problems than their Swiss counterparts) and within the countries. However, our primary research interest lies not in performing cross-country comparisons, but rather in illustrating the age trends and providing within-country comparisons across gender and educational groups. Therefore, we do not expect the different formulations of the questions to invalidate our analyses.

Table 2:
Questionnaire formulation in SIP (France) and SHP (Switzerland)

	SIP	SHP
Self-reported health	How is your health status overall? 1 means "very well" and 5 "very bad"	How do you feel right now? 1 means "very well" and 5 means "not well at all"
Daily activity limitations	Have you been limited for at least 6 months because of health problems in your daily activities? 1 means "yes" and 2 means "no"	Please tell me to what extent, generally, your health is an impediment in your everyday activities (in your housework, your work or leisure activities); 0 means "not at all" and 10 means "a great deal".
Sleep disorders	Have you had sleeping problems (difficulties in falling asleep, waking up during the night...) in the past 12 months? From 1 "never or rarely" to 4 "almost every day"	During the last 4 weeks, have you suffered from any of the following disorders or health problems? "not at all", "somewhat", "very much"? Difficulty in sleeping, or insomnia
Depressive symptoms	During the last 6 weeks, have you felt particularly sad, depressed, most of the time of the day, most days? 1 means "yes" and 2 means "no"	Do you often have negative feelings such as having the blues, being desperate, suffering from anxiety or depression, if 0 means "never" and 10 means "always"?
Satisfaction with social network/ relationship	Do you have someone you can rely on to discuss personal matters or make a difficult decision? 1 means "yes" and 2 means "no"	How satisfied are you with your personal relationships, if 0 means "not at all satisfied" and 10 means "completely satisfied"?
Satisfaction with work	Overall, are you satisfied with your professional career? 0 means "I don't agree at all" and 10 means "I totally agree"	On a scale from 0 "not at all satisfied" to 10 "completely satisfied", can you indicate your degree of satisfaction for each of the following points? Your job in general

Source: Elaboration of the authors based on SIP and SHP 2006 questionnaires.

We studied separately for the two countries the association between age and the multiple measures of health and wellbeing by gender and socioeconomic status. Ages from 25 to 69 were grouped into five-year categories to allow some degree of flexibility in the association between age and health and wellbeing, but without

Table 3:
Summary statistics SIP (France) and SHP (Switzerland)

Variables	**Units**	**Obs**	**Mean/ Proportion**	**Std.Dev.**	**Min**	**Max**
SIP, France						
Satisfaction with career	Mean	9369	7.394	2.227	0	10
Health status	Mean	9480	3.919	0.838	1	5
Someone to rely on	Proportion of yes	9480	0.907	0.290	0	1
Sleep disorders	Proportion of yes	9480	0.262	0.440	0	1
Activity limitations	Proportion of yes	9480	0.162	0.368	0	1
Depressive symptoms	Proportion of yes	9480	0.262	0.439	0	1
Gender						
Men		4291	45.26			
Women		5189	54.74			
Education						
Primary or lower secondary		5499	58.01			
Upper secondary		1488	15.70			
Tertiary		2493	26.30			
Age						
25–29		634	6.56			
30–34		869	8.99			
35–39		1141	11.80			
40–44		1212	12.54			
45–49		1255	12.98			
50–54		1397	14.45			
55–59		1395	14.43			
60–64		985	10.19			
65–69		779	8.06			
SHP, Switzerland						
Satisfaction with the job	Mean	3863	8.077	1.463	0	10
Health status	Mean	4889	4.028	.639	1	5
Satisfaction with social relationship	Mean	4889	8.141	1.515	0	10
Sleep disorders	Mean	4889	1.429	.652	1	3
Activity limitations	Mean	4889	1.75	2.481	0	10
Depressive symptoms	Mean	4889	2.124	2.083	0	10
Gender						
Men		2136	43.69			
Women		2753	56.31			

Continued

Table 3: Continued

Variables	Units	Obs	Mean/ Proportion	Std.Dev.	Min	Max
Education						
Primary or lower secondary		292	5.97			
Upper secondary		2814	57.56			
Tertiary		1783	36.47			
Age						
25–29		340	6.95			
30–34		426	8.71			
35–39		617	12.62			
40–44		791	16.18			
45–49		708	14.48			
50–54		599	12.25			
55–59		529	10.82			
60–64		472	9.65			
65–69		407	8.32			

Source: Elaboration of the authors based on SIP and SHP 2006 questionnaires.

over-specifying the model with year-dummies. We focused on respondents aged 25–69 to ensure that the respondents' education had been completed, and to exclude the oldest old, whose participation in the survey would have been even more selected based on health reasons than was the case for the other age groups. Education is a widely applied measure of socioeconomic position that reflects people's material and non-material resources, and that precedes their labour market and income prospects. We recoded education into three categories: primary or lower secondary, upper secondary and tertiary education. However, for the sake of clarity, we present in the figures only the findings for the low and the high educated.

The statistical analyses conducted in this study are based on the Ordinary Least Squares (OLS) method. We use linear regression models for continuous dependent variables and linear probability models for the binary health indicators in the SIP. Logistic regressions were also performed and provided similar results. However, we decided to show the results of the linear probability models because they are easier to interpret. We use cross-sectional weights corresponding to the countries' populations. We do not control for socioeconomic variables, such as marital or employment status, because we are interested in the total association between age and health and wellbeing, and part of this association is indirectly convened by the expedience of such life course events at given ages (Easterlin 2006). We report our findings here by graphically showing predicted health and wellbeing by age groups in the two countries, while the complete tables are included the appendix (Tables A.2–A.5).

4 Results

Results from models with age-gender (Figures 1–2) and age-education (Figures 3–4) interactions (although many interaction terms do not reach statistical significance) show that in both France and in Switzerland and in all groups, people's overall health status and physical health clearly deteriorate at older ages, while the trends in mental health vary more depending on gender and education. In France, the two measures of relational wellbeing display opposite age trends: i.e., the reliability of people's social networks steadily drop with age, while people's satisfaction with their career increases with tenure. In Switzerland, levels of relationship satisfaction are quite constant across age groups, while job satisfaction increases significantly at very old ages only.

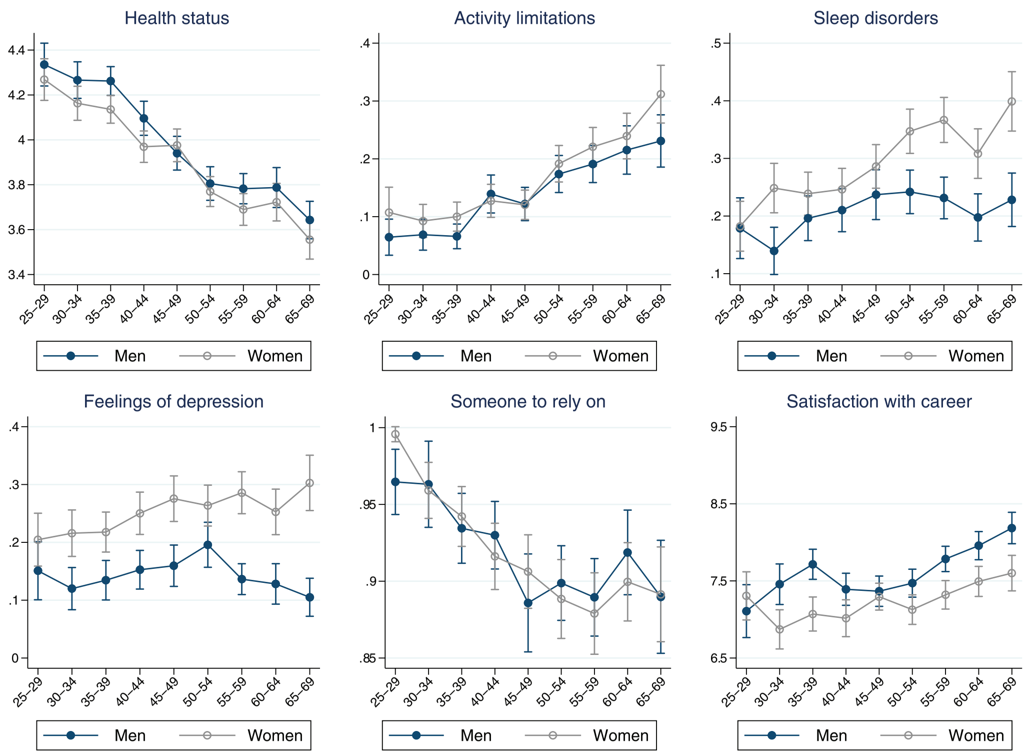
4.1 Gender inequalities in general health

All in all, the gender differences seem small. For example, in Switzerland, women have a significantly worse perceived health status than men early in their life course (age 25–29), as well as in their fifties. In France, levels of self-rated health are rather similar among men and women, with the latter reporting significantly lower levels of health only at ages 35–39. The overall decline with age is quite pronounced for both French men and women. Our estimates indicate that from early adulthood to retirement age, self-perceived health in France declines from around 4.3 (between well and very well) to around 3.6 (between well and average). In contrast, while Swiss men's self-rated health status does not change significantly after age 30, remaining above four (between well and very well), Swiss women's self-rated health status is lower in the age groups after age 50 (around 3.8, between well and average) than it is in their thirties (above four, in line with men's status). An increase in physical limitations with age can be observed among both men and women in France: by age 60, both have a probability of reporting limitations of between 20% and 30%, compared with a probability below 10% as young adults. Only Swiss women report a significant increase in limitations, such that by age 65, their impediment levels are almost twice as high (around three on a 0–10 scale) as those of younger women (1.5).

4.2 Gender inequalities in mental health

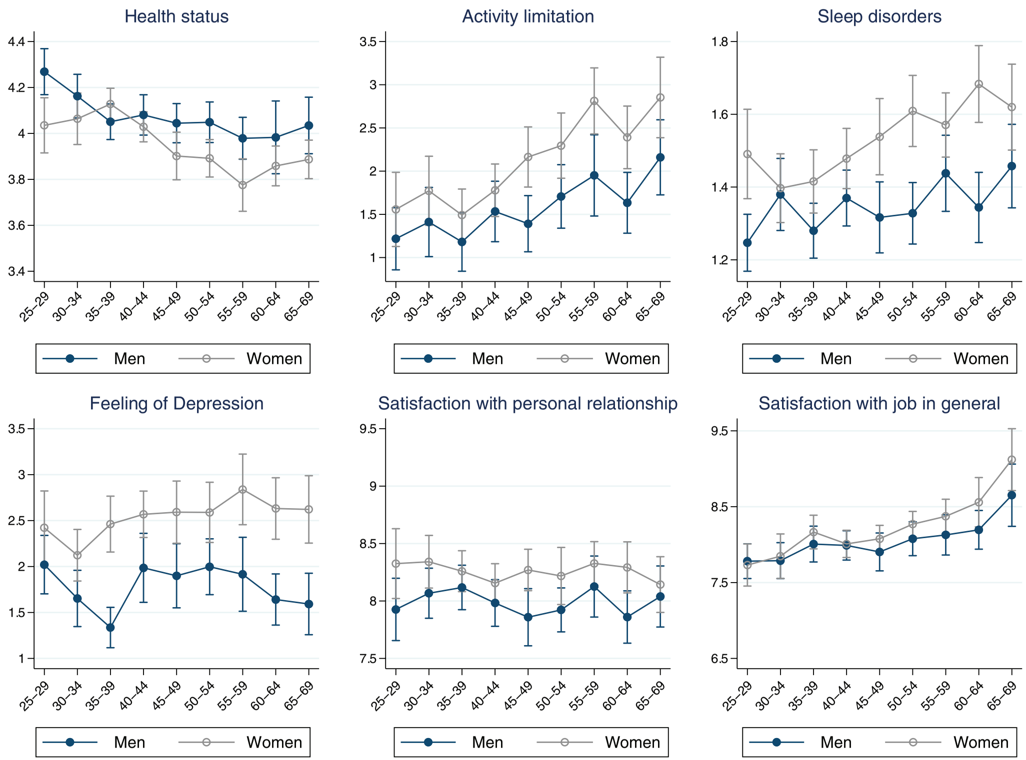
Greater gender inequalities emerge when we look at indicators of mental health. Women report a higher frequency of both sleeping disorders and depressive

Figure 1:
Health and wellbeing by age and gender in France



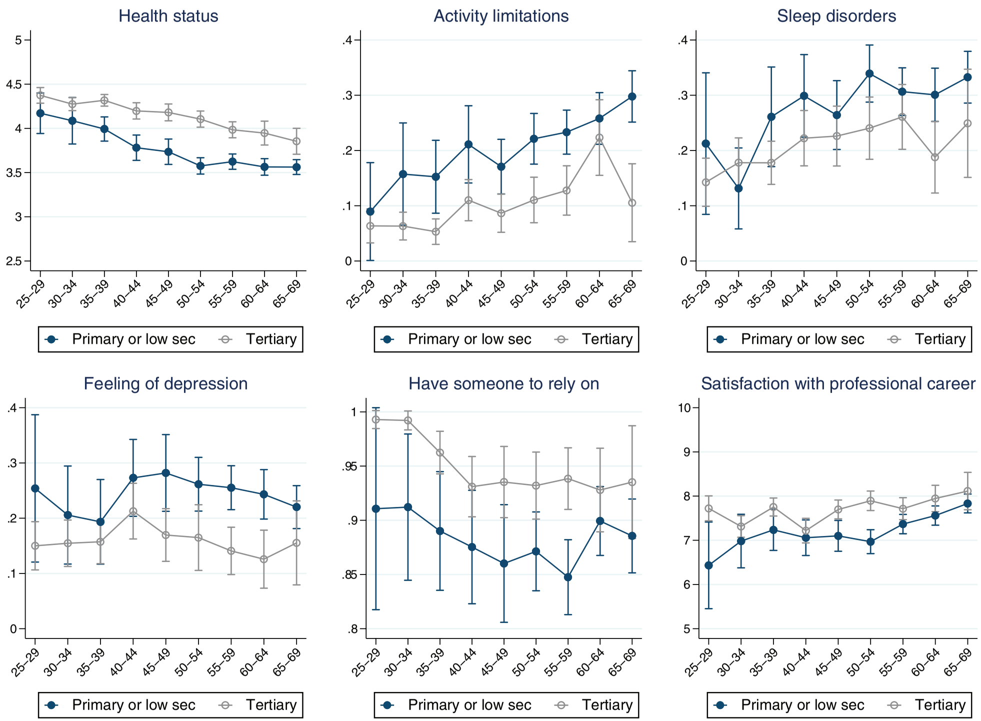
Source: Elaboration of the authors based on SIP 2006. Note: Vertical axes report (clockwise): the predicted health status on a scale of 1–5; the probability of frequently experiencing activity limitations, sleep disorders and feelings of depression and of reporting having someone to rely on; satisfaction with career on a scale of 1–10.

Figure 2:
Health and wellbeing by age and gender in Switzerland



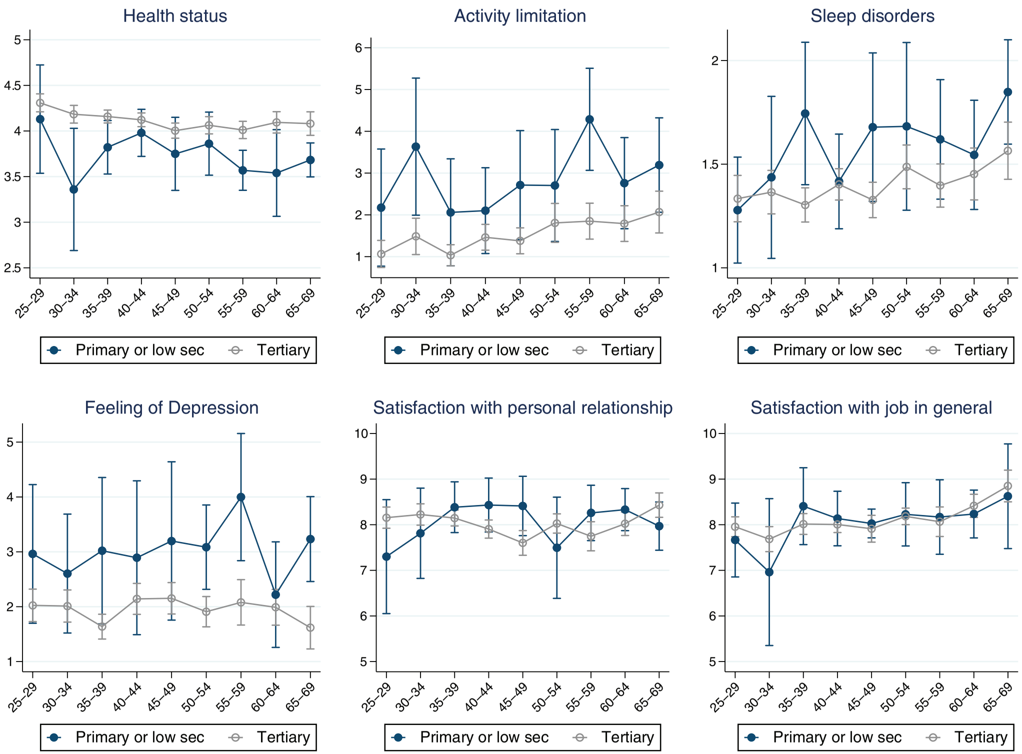
Source: Elaboration of the authors based on SHP 2006. Note: Vertical axes report (clockwise): predicted health status on a scale of 1–5; predicted frequency of activity limitations on a scale of 0–10, sleep disorders on a scale of 1–3 and feelings of depression and the predicted satisfaction with personal relationships and job on a scale of 0–10.

Figure 3:
Health and wellbeing by age and education in France



Source: Elaboration of the authors based on SIP 2006. Note: Vertical axes report (clockwise): the predicted health status on a scale of 1–5; the probability of frequently experiencing activity limitations, sleep disorders and feelings of depression and of reporting having someone to rely on; satisfaction with career on a scale of 1–10. Only the upper and lower categories of educational level are shown in order to make the graphs easier to read.

Figure 4:
Health and wellbeing by age and education in Switzerland



Source: Elaboration of the authors based on SHP 2006. Note: Vertical axes report (clockwise): predicted health status on a scale of 1–5, predicted frequency of activity limitations on a scale of 0–10, sleep disorders on a scale of 1–3 and feelings of depression and predicted satisfaction with personal relationships and job on a scale of 0–10.

symptoms, and the gap seems to widen among older age groups in both countries. In particular, women report greater sleeping problems than men at younger ages (at ages 25–29 in Switzerland and at ages 30–34 in France), and quite consistently after age 45 in Switzerland and age 50 in France. The trends are clearly diverging in France, where the probability of suffering from frequent sleep disorders (dichotomous scale) reaches 40% among women and only slightly more than 20% among men. Despite the large confidence intervals and the rather low overall incidence (always below two, somewhat), in Switzerland, the frequency of sleep disorders (1–3 scale) increases more with age for women than for men. In France, feelings of depression increase similarly for men and women up to their early fifties, but start diverging at older ages. Compared to their male counterparts, French women in the 65–69 age group have three times the probability of reporting feelings of depression (dichotomous scale). In Switzerland, the gender gap in depressive symptoms (0–10 scale) increases first among the 25–35 age groups, mostly due to declining symptoms among men. Thereafter, the gender gap again becomes pronounced after the age of 55.

4.3 Gender inequalities in relational wellbeing

In France, the probability of having someone to rely on (dichotomous scale) is significantly lower in the age groups over age 45 than it is for young adults, but the magnitude of the decline is quite small, and the trends do not differ significantly among men and women. In Switzerland, relational wellbeing does not seem to change much across age groups, remaining in the range of 8–8.5 (out of 10). Swiss women tend to report higher levels of satisfaction with social relationships than Swiss men, but the difference is statistically significant only in two age groups (50–54 and 60–64), and it is very small in magnitude (Table A.3). The differences in the age trends in relational wellbeing between the two countries might be related to the wording of the question in the two surveys: i.e., as having someone to rely on is more strongly linked to isolation and lack of independence among the elderly, we can expect this item to be more negatively associated with age than the overall satisfaction with social relationships.

Job satisfaction tends to rise with age in both countries (by around one point on a 0–10 scale), although the age differences are statistically significant only in the older age groups who are close to or older than the retirement age (Tables A.2 and A.3). An interpretation of this finding might be that the men and women who keep working after retirement age are those who enjoy it the most. More surprisingly, we observe no substantial gender differences in job satisfaction in Switzerland. Meanwhile, in France, women consistently report having lower levels of career satisfaction than men in most age groups, except in their forties and early fifties, due to a significant decline in men's levels of career satisfaction in those age groups. The differences found between countries might again be due to the different wording of the questions. Feelings of satisfaction with one's professional career more clearly reflect an evaluation of the professional opportunities of the individual, which might explain the pronounced gender differences found in France. Satisfaction with one's job in general might additionally reflect the type and the quality of the job, as well as working conditions, flexibility and economic conditions, which might have diluted the gender heterogeneities found in Switzerland.

4.4 Educational inequalities in general health

In both countries, important socioeconomic inequalities emerge when we consider measures of general and physical health, with lower educated individuals displaying worse health than higher educated individuals across most age groups.[4] Primary educated individuals have a significantly lower self-reported health status than tertiary educated individuals in all age groups except for the youngest one in France, and the difference is especially pronounced in mid-life, at ages 40–54, where the difference is around one point on a 1–5 scale. In Switzerland, the overall health of the tertiary educated individuals remains stable across age groups, and is constantly above that of the primary educated individuals. The confidence intervals are large[5] and the differences are small in magnitude, but a significant educational gap can nonetheless be observed for the 30–39 and 55+ age groups. In both countries, individuals with primary education also report greater activity limitations at ages 30-39, and especially after ages 50–54. With the exception of the 60-64 age group, a divergent trend is observed in France, whereas the confidence intervals for the primary educated in Switzerland are too large to allow us to draw conclusions.

4.5 Educational inequalities in mental health

Socioeconomic inequalities also emerge when we look at measures of mental health. Lower educated individuals display significantly higher levels of sleep disorders only at ages 35–39 in Switzerland and at ages 60-64 in France, and we see a slight trend towards increasing sleep problems for both socioeconomic groups and in both countries. The estimates show that depressive symptoms tend to be higher among the primary educated, especially after age 50. Due to the large confidence intervals, we cannot conclude that the gap in mental health across socioeconomic groups is widening at older ages, but we can definitely confirm that it does not shrink.

[4] To make the graphs more easily readable, Figures 3–4 omit the (intermediate) trends for the upper secondary educated respondents.

[5] Table A.1 in the appendix shows that there are few respondents with primary education in Switzerland, and the cell size gets very small once they are divided by age groups.

4.6 Educational inequalities in relational wellbeing

Individuals' levels satisfaction with their social relationships and with their jobs do not change between age groups in Switzerland, and the trends are basically identical across educational levels. For France, the confidence intervals are quite large for the primary educated, so definite conclusions are difficult to draw. However, the tertiary educated in France seem to report higher relational reliability, with statistically significant differences only at ages 30–39 and 55–59; and higher levels of career satisfaction, albeit only in the 50–54 age group (Tables A.4 and A.5).

Only the upper and lower categories of educational level are shown in order to make the graphs easier to read.

5 Discussion

Ensuring a satisfying quality of life throughout the life course means not only remaining in good physical and mental health and having sufficient financial means, but also maintaining high levels of overall and relational satisfaction. All of these dimensions contribute to the overall construct of *wellbeing*. There is increasing evidence that there are growing inequalities in wellbeing in contemporary societies – albeit with large contextual differences – which strongly suggests that researchers and policy-makers should focus on reducing demographic and socioeconomic gaps in quality of life.

The main contribution of our study was to show how unequally wellbeing is distributed across social groups, and how such inequalities develop over age in two different contexts. We were particularly interested in examining how different dimensions of physical and mental health and relational wellbeing heterogeneously develop with age. We exploited two rich household surveys, SIP and SHP, to examine the development of health and wellbeing across age by gender and education in France and Switzerland. Our assumption was that analysing these trends in different contexts could provide insights into the social and institutional differences in the determinants of health and wellbeing trajectories. Despite both being rich European societies with conservative welfare regimes, France and Switzerland differ in two crucial domains: namely, in their prevailing gendered family and labour market institutions and in the nature and accessibility of their health care systems (Cullati 2015).

Before examining our findings, it is important to consider that our study had a few limitations. First, the questions on health and wellbeing were phrased differently in the questionnaires of the two surveys. Nevertheless, we believe that these measures provided reliable information for examining age trends in health and wellbeing across gender and education within countries. Interestingly, some wording differences between the two surveys even allowed us to explore various aspects of the same indicator. For instance, questions about work satisfaction in the SIP asked specifically about career, while the SHP asked about overall job satisfaction. Our finding of a large gender gap in France suggests (among other plausible explanations outlined below) that it is the specific aspect of career progression that causes women to feel much less satisfied than men, even in a context like France, which ranks relatively high in terms of gender equality.

Another important difference in how the questions on health and wellbeing were phrased in the two surveys was in the time frames they referred to. While most of the questions in the French survey referred to a specific time frame, most of the questions in the Swiss survey were phrased in general terms. This difference might have had some implications for the observed inequalities. Overall, the differences across age groups might have been smaller and the estimates might have been more prone to measurement error in Switzerland, where the questions were more general. It is also important to keep in mind that all of our measures of health and wellbeing were self-reported. This means that differences in reporting behaviour by age, gender or education might have played a role in the observed differences. While recent studies found no systematic gender or educational differences in reporting either good or bad health, they found a declining concordance between self-reported and actual health with age among older respondents (Spitzer and Weber 2019; Oksuzyan et al. 2019), although much less so among younger adults (Miilunpalo et al. 1997; Pursey et al. 2014).

The second limitation of our study was its cross-sectional nature. This means that our findings reflect differences between individuals of different ages in 2006, and not the effect of the ageing process within the individual. Healthier and happier individuals tend to live longer and participate longer in surveys; therefore, our older respondents were selected in terms of higher wellbeing. This might have led us to underestimate the age differences, especially in the Swiss data (2006 is the first wave of the SIP), as well as the education gap, given that the health selection mechanism tended to be stronger among the lowest socioeconomic strata of the population. Relatedly, because of the cross-sectional character of our analysis and the well-known age-period-cohort problem, we could not distinguish between the age and the cohort effects on health and wellbeing (Bell and Jones 2014).

The final limitation of our analyses was the low number of observations – again, especially in the SHP, and especially in the group of primary educated respondents – which prevented us from providing statistically sound estimates for these groups, and from further exploring the gender and educational inequalities.

Despite these limitations, we can draw important conclusions from our analysis. We found crucial heterogeneities across indicators, and depending on the types of inequalities we looked at. Our results indicated that in both countries, overall and physical health deteriorated across age groups. We found gender inequalities in physical health only in Switzerland, but a rather large socioeconomic divide in both countries. The educational gap was shown to be largest in mid-life, which might be explained by the different degrees of physical effort required in the types of work the lower and the highly educated individuals did. The low skilled workers tended to have more physically demanding jobs, and this might have affected their health, even though no cumulative effects were observed (although, as we noted above, there was a selection process). The age trends in mental health were less clear, and varied more among the social groups. Large gender differences that widened with age were observed in both France and Switzerland. This finding suggests that there were consistent gender inequalities in both contexts, and that those inequalities tended to accumulate over time. In contrast, no large educational differences in mental health emerged.

The results further showed that the reliability of social networks decreased over age in France, while relationship satisfaction in general was more stable across age groups in Switzerland. Gender and socioeconomic status differences in relational wellbeing were observed in France, but not in Switzerland. The strong reference to the support received from significant others in the SIP questionnaire may explain the decline in relational wellbeing found for older age groups in France. Significant gender differences that increased with age were observed in France for measures of career satisfaction, but not for measures of general job satisfaction in Switzerland. This finding suggests that women tended to have less satisfying professional career opportunities than men, but may not have been less satisfied than men with other aspects of their work. Our failure to find a gender gap in job satisfaction in Switzerland may have also been driven by the selected group of Swiss women who were in full-time work (especially during their childrearing years), as they may have been more career-oriented. Additionally, the low levels of satisfaction observed among French women might be due to a mismatch between the women's professional careers and their expectations, which might have been higher in a country where gender norms are more egalitarian.

To conclude, our results suggest that both gender and socioeconomic inequalities in health and wellbeing are present in France and Switzerland. In particular, we observed significant gender inequalities on a number of wellbeing dimensions in Switzerland, as well as pronounced socioeconomic inequalities in France. Our results on gender in the Swiss context are in line with the traditionalist character of Swiss institutions, which often cause women to be dependent on men. In contrast, the finding of pronounced education inequalities in health in France might be surprising in light of the inclusive public health French system. The crucial factor here might be the much lower prevalence of primary education in Switzerland (6%, Table A.1) than in France (26%, Table A.1). On the one hand, due to the low number of observations in the lower educated group, our results might be underestimating the true inequalities in Switzerland because of the large confidence intervals of the estimates. On the other hand, the low prevalence of low education in Switzerland, which is a very rich society, shows that socioeconomic inequalities (at least those based on education) are relatively low in this context. However, this observation also implies that the very few individuals who are disadvantaged in Switzerland are much more deprived than they would be in other contexts, which makes it all the more crucial that studies are conducted to identify the sources and the outcomes of inequalities.

Finally, our findings show that while gender inequalities, when present, tended to accumulate with age, the educational gaps did not seem to diverge over time, but were largest in mid-life. This observation suggests that gender inequalities tend to be based on structural disadvantages that grow and accumulate with life events and the ageing process. For instance, the difficulties women face in reconciling family and work and managing career interruptions might have cumulative negative effects that make them more vulnerable than men at older ages. In contrast, socioeconomic inequalities, at least in the two contexts and given the indicators analysed here, seem to be linked to characteristics that matter mostly in mid-life, while they are less relevant for reducing inequalities at older ages.

Acknowledgments

This research is supported by the Swiss National Science Foundation (SNF).

References

Arber, S. and H. Cooper 1999. Gender differences in health in later life: the new paradox? *Social Science & Medicine* 48(1): 61–76. https://doi.org/10.1016/S0277-9536(98)00289-5

Arrosa, M. L. and N. Gandelman 2016. Happiness decomposition: Female optimism. *Journal of Happiness Studies* 17(2): 731–756. https://doi.org/10.1007/s10902-015-9618-8

Bambra, C., D. Pope, V. Swami, D. Stanistreet, A. Roskam, A. Kunst and A. Scott-Samuel 2009. Gender, health inequalities and welfare state regimes: A cross-national study of 13 European countries. *Journal of Epidemiology and Community Health* 63(1): 38–44. https://doi.org/10.1136/jech.2007.070292

Bell, A. and K. Jones 2014. Don't birth cohorts matter? A commentary and simulation exercise on Reither, Hauser, and Yang's (2009) age-period-cohort study of obesity. *Social Science & Medicine* 101: 176–180. https://doi.org/10.1016/j.socscimed.2013.09.004

Blanchflower, D. G. and A. J. Oswald 2004. Money, sex and happiness: An empirical study. *Scandinavian Journal of Economics* 106(3): 393–415. https://doi.org/10.1111/j.0347-0520.2004.00369.x

Blanchflower, D. G. and A. J. Oswald 2016. Antidepressants and age: A new form of evidence for U-shaped well-being through life. *Journal of Economic Behavior and Organization* 127: 46–58. https://doi.org/10.1016/j.jebo.2016.04.010

Breton, D., M. Barbieri, N. Belliot, H. d'Albis and M. Mazuy 2019. L'évolution démographique récente de la France: une singularité en Europe? *Population* 74(4): 409–497. https://doi.org/10.3917/popu.1904.0409

Burgard, S. A. and L. Kalousova 2015. Effects of the great recession: Health and wellbeing. *Annual Review of Sociology* 41: 181–201. https://doi.org/10.1146/annurev-soc-073014-112204

Cerci, P. A. and D. Dumludag 2019. Life satisfaction and job satisfaction among university faculty: the impact of working conditions, academic performance and relative income. *Social Indicators Research* 144(2): 785–806. https://doi.org/10.1007/s11205-018-02059-8

Cooke, L. P. and J. Baxter 2010. "Families" in international context: comparing institutional effects across western societies. *Journal of Marriage and Family* 72(3): 516–536. https://doi.org/10.1111/j.1741-3737.2010.00716.x

Crimmins, E. M. and Y. Saito 2001. Trends in healthy life expectancy in the United States, 1970–1990: gender, racial, and educational differences. *Social Science & Medicine* 52(11): 1629–1641. https://doi.org/10.1016/S0277-9536(00)00273-2

Cronin de Chavez, A., K. Backett-Milburn, O. Parry and S. Platt 2005. Understanding and researching wellbeing: Its usage in different disciplines and potential for health research and health promotion. *Health education Journal* 64(1): 70–87. https://doi.org/10.1177/001789690506400108

Cullati, S. 2015. The influence of work-family conflict trajectories on self-rated health trajectories in Switzerland: A life course approach. *Social Science & Medicine* 113: 23–33. https://doi.org/10.1016/j.socscimed.2014.04.030

Dahlin, J. and J. Härkönen 2013. Cross-national differences in the gender gap in subjective health in Europe: Does country-level gender equality matter? *Social Science & Medicine* 98: 24–28. https://doi.org/10.1016/j.socscimed.2013.08.028

Dannefer, D. 2003. Cumulative advantage/disadvantage and the life course: Cross-fertilizing age and social science theory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 58(6): S327–S337. https://doi.org/10.1093/geronb/58.6.S327

De Neve, J. E., N. A. Christakis, J. H. Fowler and B. S. Frey 2012. Genes, economics, and happiness. *Journal of Neuroscience, Psychology, and Economics* 5(4): 193–211. https://doi.org/10.1037/a0030292

DiPrete, T. A. and G. M. Eirich 2006. Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology* 32: 271–297. https://doi.org/10.1146/annurev.soc.32.061604.123127

Easterlin, R. A. 2001. Life cycle welfare: trends and differences. *Journal of Happiness Studies* 2(1): 1–12. https://doi.org/10.1023/A:1011504817292

Easterlin, R. A. 2006. Life cycle happiness and its sources: Intersections of psychology, economics, and demography. *Journal of Economic Psychology* 27(4): 463–482. https://doi.org/10.1016/j.joep.2006.05.002

Edgerton, J. D., L. W. Roberts and S. von Below 2012. Education and quality of life. In *Handbook of social indicators and quality of life research*, eds K. Land, A. Michalos and M. Sirgy, 265–296. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-2421-1_12

Eikemo, T. A., M. Huisman, C. Bambra and A. E. Kunst 2008. Health inequalities according to educational level in different welfare regimes: A comparison of 23 European countries. *Sociology of Health and Illness* 30(4): 565–582. https://doi.org/10.1111/j.1467-9566.2007.01073.x

Elo, I. 2009. Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual Review of Sociology* 35: 553–572. https://doi.org/10.1146/annurev-soc-070308-115929

Esping-Andersen, G. 1990. *The three worlds of welfare capitalism.* Princeton University Press.

Fagnani, J. 2010. Childcare policies in France: The influence of organizational changes in the workplace. In *From child welfare to child well-being*, eds S. B. Kamerman, S. Phipps, A. Ben-Arieh, 385–402. Springer, Dordrecht. https://doi.org/10.1007/978-90-481-3377-2_21

Fernández-Ballesteros, R., M. D. Zamarrón and M. Á. Ruiz 2001. The contribution of socio-demographic and psychosocial factors to life satisfaction. *Ageing and Society* 21(1): 25–43. https://doi.org/10.1017/S0144686X01008078

Frijters, P. and T. Beatton 2012. The mystery of the U-shaped relationship between happiness and age. *Journal of Economic Behavior and Organization* 82(2–3): 525–542. https://doi.org/10.1016/j.jebo.2012.03.008

Hale, L. 2005. Who has time to sleep? *Journal of Public Health* 27(2): 205–211. https://doi.org/10.1093/pubmed/fdi004

Huisman, M., A. E. Kunst and J. P. Mackenbach 2003. Socioeconomic inequalities in morbidity among the elderly; a European overview. *Social Science & Medicine* 57(5): 861–873. https://doi.org/10.1016/S0277-9536(02)00454-9

Huppert, F. A. 2013. The state of Wellbeing Science: Concepts, Measures, Interventions, and Policies. In *Wellbeing*, ed C. L. Cooper. https://doi.org/10.1002/9781118539415.wbwell036

Huppert, F. A. and T. T. C. So 2013. Flourishing across Europe: Application of a new conceptual framework for defining wellbeing. *Social Indicators Research* 110(3): 837–861. https://doi.org/10.1007/s11205-011-9966-7

Huppert, F. A., N. Marks, A. Clark, J. Siegrist, A. Stutzer, J. Vittersø and M. Wahrendorf 2009. Measuring wellbeing across Europe: Description of the ESS wellbeing module and preliminary findings. *Social Indicators Research* 91(3): 301–315. https://doi.org/10.1007/s11205-008-9346-0

Infurna F. J. and S. S. Luthar 2017. The multidimensional nature of resilience to spousal loss. *Journal of Personality and Social Psychology* 112(6): 926–947. https://psycnet.apa.org/doi/10.1037/pspp0000095

Keizer, R., P. A. Dykstra and A. R. Poortman 2010. The transition to parenthood and well-being: the impact of partner status and work hour transitions. *Journal of Family Psychology* 24(4): 429. https://psycnet.apa.org/doi/10.1037/a0020414

Keyes, Corey L. M. 2007. Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. *American Psychologist* 62(2): 95–108. https://psycnet.apa.org/doi/10.1037/0003-066X.62.2.95

Kratz, F. and A. Patzina 2020. Endogenous selection bias and cumulative inequality over the life course: Evidence from educational inequality in subjective well-being. *European Sociological Review* 36(3): 333–350. https://doi.org/10.1093/esr/jcaa003

Kunst, A. E., J. J. Geurts and J. van den Berg 1995. International variation in socioeconomic inequalities in self reported health. *Journal of Epidemiology and Community Health* 49(2): 117–123. http://doi.org/10.1136/jech.49.2.117

Lang, I. A., D. J. Llewellyn, R. E. Hubbard, K. M. Langa and D. Melzer 2011. Income and the midlife peak in common mental disorder prevalence. *Psychological Medicine* 41(7): 1365–1372. http://doi.org/10.1017/S0033291710002060

Lee, C. T., Y. C. Chiang, J. Y. Huang, D. M. Tantoh, O. N. Nfor, J. F. Lee and Y. P. Liaw 2016. Incidence of major depressive disorder: Variation by age and sex in low-income individuals: A population-based 10-year follow-up study. *Medicine* 95(15). https://doi.org/10.1097%2FMD.0000000000003110

Lleras-Muney, A. 2005. The relationship between education and adult mortality in the United States. *The Review of Economic Studies* 72(1): 189–221. https://doi.org/10.1111/0034-6527.00329

Loscocco, K. A. and G. Spitze 1990. Working conditions, social support, and the wellbeing of female and male factory workers. *Journal of Health and Social Behavior* 31(4): 313–327. https://doi.org/10.2307/2136816

Lynch, S. M. 2003. Cohort and life-course patterns in the relationship between education and health: A hierarchical approach. *Demography* 40(2): 309–331. https://doi.org/10.1353/dem.2003.0016

Mackenbach, J. P. 2012. The persistence of health inequalities in modern welfare states: The explanation of a paradox. *Social Science & Medicine* 75(4): 761–769. https://doi.org/10.1016/j.socscimed.2012.02.031

Mackenbach, J. P., A. E. Kunst, A. E. Cavelaars, F. Groenhof, J. J. Geurts and EU Working Group on Socioeconomic Inequalities in Health. 1997. Socioeconomic inequalities in morbidity and mortality in western Europe. *The Lancet* 349(9066): 1655–1659. https://doi.org/10.1016/S0140-6736(96)07226-1

Madero-Cabib, I. and A. E. Fasang 2016. Gendered work–family life courses and financial well-being in retirement. *Advances in Life Course Research* 27: 43–60. https://doi.org/10.1016/j.alcr.2015.11.003

Marmot, M. 2005. Social determinants of health inequalities. *The Lancet* 365(9464): 1099–1104. https://doi.org/10.1016/S0140-6736(05)71146-6

Mayer, K. U. 2009. New directions in life course research. *Annual Review of Sociology* 35: 413–433. https://doi.org/10.1146/annurev.soc.34.040507.134619

McAdams, K. K., R. E. Lucas and M. B. Donnellan 2012. The role of domain satisfaction in explaining the paradoxical association between life satisfaction and age. *Social Indicators Research* 109(2): 295–303. https://doi.org/10.1007/s11205-011-9903-9

McDonough, P., D. Worts, C. Booker, A. McMunn and A. Sacker 2015. Cumulative disadvantage, employment–marriage, and health inequalities among American and British mothers. *Advances in Life Course Research* 25: 49–66. https://doi.org/10.1016/j.alcr.2015.05.004

McPherson, M., L. Smith-Lovin and M. E. Brashears 2006. Social isolation in America: Changes in core discussion networks over two decades. *American Sociological Review* 71(3): 353–375. https://doi.org/10.1177%2F000312240607100301

Melchior, M., A. Caspi, B. Milne, A. Danese, R. Poulton and T. Moffitt 2007. Work stress precipitates depression and anxiety in young, working women and men. *Psychological Medicine* 37: 1119–1129. https://doi.org/10.1017%2FS0033291707000414

Miilunpalo, S., I. Vuori, P. Oja, M. Pasanen and H. Urponen 1997. Self-rated health status as a health measure: The predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *Journal of Clinical Epidemiology* 50(5): 517–528. https://doi.org/10.1016/S0895-4356(97)00045-0

Mirowsky, J. and C. E. Ross 1998. Education, personal control, lifestyle and health: A human capital hypothesis. *Research on Aging* 20(4): 415–449. https://doi.org/10.1177%2F0164027598204003

Mirowsky, J. and C. E. Ross 2005. Education, cumulative advantage, and health. *Ageing International* 30(1): 27–62. https://doi.org/10.1007/BF02681006

Mirowsky, J. and C. E. Ross 2007. Life course trajectories of perceived control and their relationship to education. *American Journal of Sociology* 112(5): 1339–1382. https://doi.org/10.1086/511800

Needham, B. and T. D. Hill 2010. Do gender differences in mental health contribute to gender differences in physical health? *Social Science and Medicine* 71(8): 1472–1479. https://doi.org/10.1016/j.socscimed.2010.07.016

Oksuzyan, A., Dańko, M. J., J. Caputo, D. Jasilionis and V. M. Shkolnikov 2019. Is the story about sensitive women and stoical men true? Gender differences in health after adjustment for reporting behavior. *Social Science & Medicine* 228: 41–50. https://doi.org/10.1016/j.socscimed.2019.03.002

Oncini, F. and R. Guetto 2018. Cultural capital and gender differences in health behaviours: A study on eating, smoking and drinking patterns. *Health Sociology Review* 27(1): 15–30. https://doi.org/10.1080/14461242.2017.1321493

Pailhé, A., C. Rossier and L. Toulemon 2008. French family policy: long tradition and diversified measures. *Vienna Yearbook of Population Research 2008*: 149–164. https://doi.org/10.1553/populationyearbook2008s149

Pampel, F. C., P. M. Krueger and J. T. Denney 2010. Socioeconomic disparities in health behaviors. *Annual Review of Sociology* 36: 349–370. https://doi.org/10.1146/annurev.soc.012809.102529

Peters, R. 2014. La discrimination fiscal entre les couples mariés et non mariés dans les cantons et à la Confédération, Département Fédéral des Finances, Suisse.

Pinquart, M. and S. Sörensen 2001. Gender differences in self-concept and psychological well-being in old age: A meta-analysis. *The Journals of Gerontology Series B: Psychological sciences and social sciences* 56(4): 195–213. https://doi.org/10.1093/geronb/56.4.P195

Pollard, E. L. and P. D. Lee 2003. Child wellbeing: A systematic review of the literature. *Social Indicators Research* 61(1): 59–78. https://doi.org/10.1023/A:1021284215801

Prus, S. G. 2004. A life course perspective on the relationship between socio-economic status and health: Testing the divergence hypothesis. *Canadian Journal on Aging/La Revue canadienne du vieillissement* 23(5): 145–153. http://doi.org/10.1353/cja.2005.0040

Pursey, K., T. L. Burrows, P. Stanwell and C. E. Collins 2014. How accurate is web-based self-reported height, weight, and body mass index in young adults? *Journal of medical Internet research* 16(1): e4. https://doi.org/10.2196/jmir.2909

Ross, C. E. and M. Van Willigen 1997. Education and the subjective quality of life. *Journal of Health and Social Behavior* 38(3): 275–297. https://doi.org/10.2307/2955371

Ross, C. E. and C. Wu 1995. The links between education and health. *American Sociological Review* 60(5): 719–745. https://doi.org/10.2307/2096319

Ryff, C. D. and C. L. M. Keyes 1995. The structure of psychological wellbeing revisited. *Journal of Personality and Social Psychology* 69(4): 719–727. https://psycnet.apa.org/doi/10.1037/0022-3514.69.4.719

Spitzer, S. and D. Weber 2019. Reporting biases in self-assessed physical and cognitive health status of older Europeans. *PLoS ONE* 14(10): e0223526. https://doi.org/10.1371/journal.pone.0223526

Subramanian, S. V., D. Kim and I. Kawachi 2005. Covariation in the socioeconomic determinants of self-rated health and happiness: a multivariate multilevel analysis of individuals and communities in the USA. *Journal of Epidemiology and Community Health* 59(8): 613. http://doi.org/10.1136/jech.2004.025742

Townsend, P. and N. Davidson (eds). 1982. *Inequalities in Health: The Black Report*. Harmondsworth: Penguin.

Troxel, W. M., D. J. Buysse, K. A. Matthews, H. M. Kravitz, J. T. Bromberger, M. Sowers and M. H. Hall 2010. Marital/cohabitation status and history in relation to sleep in midlife women. *Sleep* 33(7): 973–981. https://doi.org/10.1093/sleep/33.7.973

Ulloa, B. F. L., V. Møller, and A. Sousa-Poza 2013. How does subjective well-being evolve with age? A literature review. *Journal of Population Ageing* 6(3): 227–246. https://doi.org/10.1007/s12062-013-9085-0

Von dem Knesebeck, O., P. E. Verde and N. Dragano 2006. Education and health in 22 European countries. *Social Science & Medicine* 63(5): 1344–1351. https://doi.org/10.1016/j.socscimed.2006.03.043

Walker, A. 2005. A European perspective on quality of life in old age. *European Journal of Ageing* 2: 2–12. https://doi.org/10.1007/s10433-005-0500-0

Wall, K. and A. Escobedo 2013. Parental leave policies, gender equity and family well-being in Europe: A comparative perspective. In *Family Well-Being*, ed A. Moreno Minguez, 103–129. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-4354-0_6

Wheaton, B. 1980. The sociogenesis of psychological disorder: An attributional theory. *Journal of Health and Social Behavior* 21: 100–124. https://doi.org/10.2307/2136730

Widmer, E., R. Levy, A. Pollien, R. Hammer and J. A. Gauthier 2003. Entre standardisation, individualisation et sexuation: une analyse des trajectoires personnelles en Suisse. *Revue Suisse de Sociologie* 29: 35–67.

Yang, Y. 2008. Social inequalities in happiness in the United States, 1972 to 2004: An age-period-cohort analysis. *American Sociological Review* 73(2): 204–226. https://doi.org/10.1177%2F000312240807300202

Appendix

Table A.1:
Age distribution by educational level in SIP and SHP

		SIP				SHP			
		Primary or Low Sec.	Upper Secondary	Tertiary	Total	Primary or Low Sec.	Upper Secondary	Tertiary	Total
25–29	*N*	45	260	302	607	14	188	138	340
	%	7.41	42.83	49.75		4.12	55.29	40.59	
30–34	*N*	83	390	383	856	14	234	178	426
	%	9.70	45.56	44.74		3.29	54.93	41.78	
35–39	*N*	138	576	408	1122	26	329	262	617
	%	12.30	51.34	36.36		4.21	53.32	42.46	
40–44	*N*	176	697	324	1197	42	426	323	791
	%	14.70	58.23	27.07		5.31	53.86	40.83	
45–49	*N*	246	720	270	1236	35	419	254	708
	%	19.90	58.25	21.84		4.94	59.18	35.88	
50–54	*N*	394	704	273	1371	27	362	210	599
	%	28.74	51.35	19.91		4.51	60.43	35.06	
55–59	*N*	503	580	287	1370	42	327	160	529
	%	36.72	42.34	20.95		7.94	61.81	30.25	
60–64	*N*	419	373	167	959	40	289	143	472
	%	43.69	38.89	17.41		8.47	61.23	30.30	
65–69	*N*	442	241	79	762	52	240	115	407
	%	58.01	31.63	10.37		12.78	58.97	28.26	
Total	*N*	2446	4541	2493	9480	292	2814	1783	4889
	%	25.80	47.90	26.30		5.97	57.56	36.47	100.00

Source: Elaboration of the authors based on SIP and SHP 2006.

Table A.2:
Health and wellbeing outcomes by age groups and gender. France

	Model (1) Health status	Model (2) Activity limitations	Model (3) Sleep disorders	Model (4) Depression	Model (5) Personal relations satisfaction	Model (6) Job satisfaction
Age groups (25–29 Ref)						
Age 30–34	−0.085	0.018	−0.040	−0.012	0.021	0.090
	(−0.212 – 0.042)	(−0.039 – 0.075)	(−0.109 – 0.028)	(−0.076 – 0.051)	(−0.024 – 0.066)	(−0.257 – 0.436)
Age 35-39	−0.111*	0.017	0.024	−0.005	−0.022	0.450***
	(−0.230 – 0.008)	(−0.037 – 0.070)	(−0.040 – 0.088)	(−0.064 – 0.055)	(−0.064 – 0.021)	(0.126 – 0.775)
Age 40–44	−0.282***	0.078***	0.046	0.024	−0.031	0.163
	(−0.400 – −0.164)	(0.025 – 0.131)	(−0.018 – 0.109)	(−0.035 – 0.083)	(−0.073 – 0.011)	(−0.158 – 0.485)
Age 45–49	−0.419***	0.073***	0.062*	0.028	−0.067***	0.088
	(−0.537 – −0.300)	(0.020 – 0.126)	(−0.001 – 0.126)	(−0.031 – 0.087)	(−0.109 – −0.025)	(−0.235 – 0.410)
Age 50–54	−0.555***	0.126***	0.074**	0.042	−0.057***	0.201
	(−0.669 – −0.440)	(0.074 – 0.177)	(0.012 – 0.136)	(−0.015 – 0.100)	(−0.098 – −0.016)	(−0.112 – 0.515)
Age 55–59	−0.576***	0.136***	0.051	0.013	−0.062***	0.510***
	(−0.689 – −0.462)	(0.085 – 0.187)	(−0.010 – 0.112)	(−0.044 – 0.070)	(−0.103 – −0.022)	(0.199 – 0.820)
Age 60–64	−0.594***	0.151***	0.028	−0.004	−0.040*	0.714***
	(−0.717 – −0.472)	(0.096 – 0.206)	(−0.038 – 0.094)	(−0.065 – 0.057)	(−0.084 – 0.003)	(0.380 – 1.047)
Age 65–69	−0.710***	0.175***	0.050	−0.033	−0.058**	0.992***
	(−0.838 – −0.583)	(0.118 – 0.232)	(−0.018 – 0.119)	(−0.097 – 0.031)	(−0.103 – −0.012)	(0.645 – 1.339)
Woman	−0.069	0.037	0.034	0.094***	0.042*	0.008
	(−0.198 – 0.060)	(−0.022 – 0.095)	(−0.036 – 0.103)	(0.030 – 0.159)	(−0.004 – 0.088)	(−0.345 – 0.361)

Continued

Table A.2:
Continued

	Model (1) Health status	Model (2) Activity limitations	Model (3) Sleep disorders	Model (4) Depression	Model (5) Personal relations satisfaction	Model (6) Job satisfaction
Age 30–34 ∗ Woman	−0.043	−0.026	0.076	−0.001	−0.054*	−0.384
	(−0.213 – 0.126)	(−0.102 – 0.051)	(−0.015 – 0.167)	(−0.085 – 0.084)	(−0.115 – 0.006)	(−0.847 – 0.080)
Age 35–39 ∗ Woman	−0.041	−0.012	0.005	0.005	−0.032	−0.558**
	(−0.201 – 0.119)	(−0.084 – 0.060)	(−0.081 – 0.092)	(−0.075 – 0.085)	(−0.090 – 0.025)	(−0.997 – −0.120)
Age 40–44 ∗ Woman	−0.008	−0.054	−0.014	−0.003	−0.056*	−0.313
	(−0.167 – 0.151)	(−0.125 – 0.018)	(−0.100 – 0.071)	(−0.082 – 0.076)	(−0.113 – 0.000)	(−0.747 – 0.120)
Age 45–49 ∗ Woman	0.108	−0.042	0.029	−0.003	−0.031	−0.086
	(−0.050 – 0.266)	(−0.113 – 0.029)	(−0.056 – 0.114)	(−0.082 – 0.076)	(−0.087 – 0.025)	(−0.519 – 0.346)
Age 50–54 ∗ Woman	0.072	−0.027	0.077*	−0.011	−0.044	−0.308
	(−0.083 – 0.227)	(−0.096 – 0.043)	(−0.007 – 0.161)	(−0.088 – 0.067)	(−0.100 – 0.011)	(−0.732 – 0.117)
Age 55–59 ∗ Woman	0.009	−0.012	0.108**	0.048	−0.055*	−0.432**
	(−0.146 – 0.164)	(−0.082 – 0.057)	(0.024 – 0.191)	(−0.030 – 0.125)	(−0.110 – 0.001)	(−0.856 – −0.008)
Age 60–64 ∗ Woman	0.007	−0.003	0.078*	0.036	−0.067**	−0.464**
	(−0.158 – 0.172)	(−0.077 – 0.071)	(−0.011 – 0.167)	(−0.046 – 0.119)	(−0.126 – −0.008)	(−0.915 – −0.012)
Age 65–69 ∗ Woman	−0.006	0.028	0.148***	0.124***	−0.055*	−0.704***
	(−0.179 – 0.167)	(−0.050 – 0.106)	(0.055 – 0.242)	(0.037 – 0.210)	(−0.116 – 0.007)	(−1.178 – −0.229)
Constant	4.344***	0.062***	0.176***	0.139***	0.949***	7.243***
	(4.248 – 4.440)	(0.019 – 0.105)	(0.124 – 0.227)	(0.091 – 0.187)	(0.915 – 0.983)	(6.981 – 7.504)
Observations	9,480	9,480	9,480	9,480	9,480	9,369
R-squared	0.067	0.029	0.023	0.023	0.011	0.017

Source: Elaboration of the authors based on SIP 2006. Note: Confidence intervals in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.3:
Health and wellbeing outcomes by age groups and gender. Switzerland

	Model (1) Health status	Model (2) Sleep disorders	Model (3) Activity limitations	Model (4) Depression	Model (5) Personal relations Satisfaction	Model (6) Job satisfaction
Age groups (25–29 Ref)						
Age 30–34	−0.145**	0.117*	0.168	−0.265	0.014	−0.040
	(−0.277 – −0.013)	(−0.016 – 0.251)	(−0.340 – 0.676)	(−0.692 – 0.163)	(−0.299 – 0.328)	(−0.357 – 0.277)
Age 35–39	−0.182***	0.040	0.051	−0.519**	0.019	0.190
	(−0.305 – −0.060)	(−0.084 – 0.164)	(−0.421 – 0.523)	(−0.916 – −0.122)	(−0.272 – 0.310)	(−0.105 – 0.485)
Age 40–44	−0.173***	0.115*	0.249	−0.098	−0.087	0.214
	(−0.289 – −0.057)	(−0.003 – 0.233)	(−0.199 – 0.697)	(−0.475 – 0.279)	(−0.364 – 0.189)	(−0.067 – 0.495)
Age 45–49	−0.232***	0.054	0.166	−0.100	−0.239*	0.263*
	(−0.351 – −0.114)	(−0.067 – 0.174)	(−0.291 – 0.623)	(−0.485 – 0.284)	(−0.521 – 0.043)	(−0.022 – 0.549)
Age 50–54	−0.244***	0.111*	0.559**	−0.027	−0.135	0.354**
	(−0.366 – −0.122)	(−0.013 – 0.235)	(0.088 – 1.030)	(−0.424 – 0.369)	(−0.425 – 0.156)	(0.059 – 0.649)
Age 55–59	−0.239***	0.171***	0.599**	−0.124	−0.037	0.359**
	(−0.365 – −0.113)	(0.043 – 0.298)	(0.113 – 1.085)	(−0.533 – 0.285)	(−0.337 – 0.263)	(0.054 – 0.664)
Age 60–64	−0.270***	0.116*	0.491*	−0.111	−0.058	0.465***
	(−0.399 – −0.141)	(−0.015 – 0.247)	(−0.006 – 0.988)	(−0.529 – 0.307)	(−0.365 – 0.249)	(0.133 – 0.798)
Age 65–69	−0.231***	0.222***	1.047***	−0.309	0.086	0.843***
	(−0.363 – −0.099)	(0.088 – 0.356)	(0.537 – 1.557)	(−0.738 – 0.120)	(−0.228 – 0.401)	(0.406 – 1.281)
Woman	−0.192***	0.186***	0.226	0.265	0.343**	−0.033
	(−0.327 – −0.057)	(0.049 – 0.323)	(−0.293 – 0.746)	(−0.172 – 0.702)	(0.023 – 0.664)	(−0.365 – 0.299)

Continued

Table A.3:
Continued

	Model (1) Health status	Model (2) Sleep disorders	Model (3) Activity limitations	Model (4) Depression	Model (5) Personal relations Satisfaction	Model (6) Job satisfaction
Age 30–34 ∗ Woman	0.166*	−0.094	0.064	0.245	0.000	0.281
	(−0.015 – 0.347)	(−0.278 – 0.091)	(−0.635 – 0.763)	(−0.343 – 0.834)	(−0.431 – 0.432)	(−0.164 – 0.726)
Age 35–39 ∗ Woman	0.225***	−0.078	−0.024	0.617**	−0.065	0.305
	(0.057 – 0.394)	(−0.249 – 0.094)	(−0.673 – 0.626)	(0.070 – 1.164)	(−0.466 – 0.336)	(−0.108 – 0.718)
Age 40–44 ∗ Woman	0.146*	−0.112	−0.040	0.271	−0.058	0.084
	(−0.015 – 0.308)	(−0.276 – 0.052)	(−0.662 – 0.583)	(−0.252 – 0.795)	(−0.442 – 0.326)	(−0.310 – 0.479)
Age 45–49 ∗ Woman	0.102	0.001	0.498	0.297	0.138	0.120
	(−0.062 – 0.267)	(−0.166 – 0.168)	(−0.136 – 1.131)	(−0.237 – 0.830)	(−0.253 – 0.530)	(−0.283 – 0.522)
Age 50–54 ∗ Woman	0.087	0.055	0.157	0.350	0.013	0.161
	(−0.082 – 0.257)	(−0.117 – 0.227)	(−0.495 – 0.809)	(−0.199 – 0.899)	(−0.389 – 0.416)	(−0.252 – 0.575)
Age 55–59 ∗ Woman	0.015	−0.022	0.613*	0.550*	−0.037	0.262
	(−0.158 – 0.188)	(−0.198 – 0.155)	(−0.056 – 1.281)	(−0.013 – 1.112)	(−0.450 – 0.375)	(−0.169 – 0.693)
Age 60–64 ∗ Woman	0.076	0.118	0.542	0.580**	0.011	0.349
	(−0.102 – 0.253)	(−0.062 – 0.298)	(−0.142 – 1.226)	(0.004 – 1.155)	(−0.411 – 0.433)	(−0.130 – 0.828)
Age 65–69 ∗ Woman	0.050	−0.042	0.379	0.564*	−0.271	0.633*
	(−0.133 – 0.233)	(−0.228 – 0.144)	(−0.326 – 1.084)	(−0.030 – 1.157)	(−0.707 – 0.164)	(−0.079 – 1.346)
Constant	4.275***	1.234***	1.135***	1.924***	8.023***	7.750***
	(4.180 – 4.370)	(1.137 – 1.330)	(0.768 – 1.501)	(1.616 – 2.232)	(7.797 – 8.249)	(7.516 – 7.984)
Observations	4889	4889	4889	4889	4889	3863
R-squared	0.022	0.030	0.035	0.030	0.014	0.022

Source: Elaboration of the authors based on SHP 2006.
Note: Confidence intervals in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table A.4:
Health and wellbeing outcomes by age groups and education. France

	Model (1) Health status	Model (2) Sleep disorders	Model (3) Activity limitations	Model (4) Depression	Model (5) Personal relations Satisfaction	Model (6) Job satisfaction
Age groups (25–29 Ref)						
Age 30–34	−0.165	0.065	−0.054	0.005	0.027	0.264
	(−0.455 – 0.124)	(−0.066 – 0.196)	(−0.212 – 0.104)	(−0.141 – 0.151)	(−0.077 – 0.132)	(−0.554 – 1.082)
Age 35–39	−0.221	0.061	0.004	−0.050	−0.004	0.792**
	(−0.490 – 0.048)	(−0.061 – 0.182)	(−0.142 – 0.151)	(−0.185 – 0.086)	(−0.101 – 0.093)	(0.039 – 1.546)
Age 40–44	−0.441***	0.120**	0.068	0.056	−0.025	0.488
	(−0.702 – −0.179)	(0.002 – 0.239)	(−0.074 – 0.211)	(−0.075 – 0.188)	(−0.119 – 0.069)	(−0.246 – 1.222)
Age 45-49	−0.523***	0.105*	0.048	0.012	−0.051	0.626*
	(−0.777 – −0.269)	(−0.010 – 0.220)	(−0.090 – 0.186)	(−0.116 – 0.140)	(−0.142 – 0.040)	(−0.086 – 1.338)
Age 50–54	−0.629***	0.130**	0.121*	0.013	−0.015	0.457
	(−0.875 – −0.382)	(0.019 – 0.242)	(−0.013 – 0.255)	(−0.111 – 0.137)	(−0.104 – 0.073)	(−0.234 – 1.149)
Age 55–59	−0.604***	0.153***	0.101	0.027	−0.044	0.900***
	(−0.848 – −0.361)	(0.043 – 0.263)	(−0.031 – 0.234)	(−0.096 – 0.150)	(−0.131 – 0.044)	(0.217 – 1.583)
Age 60–64	−0.672***	0.168***	0.089	0.001	0.004	1.095***
	(−0.917 – −0.426)	(0.056 – 0.279)	(−0.045 – 0.223)	(−0.123 – 0.125)	(−0.084 – 0.092)	(0.407 – 1.783)
Age 65–69	−0.676***	0.199***	0.100	−0.016	−0.004	1.301***
	(−0.921 – −0.432)	(0.088 – 0.310)	(−0.033 – 0.234)	(−0.140 – 0.107)	(−0.092 – 0.084)	(0.614 – 1.988)
Education (Ref. Primary or Lower secondary)						
Upper Secondary	0.016	0.011	0.020	−0.040	0.076*	0.345
	(−0.237 – 0.269)	(−0.103 – 0.126)	(−0.117 – 0.158)	(−0.167 – 0.087)	(−0.014 – 0.167)	(−0.363 – 1.052)
Tertiary	0.163	−0.025	−0.085	−0.087	0.102**	1.288***
	(−0.087 – 0.413)	(−0.138 – 0.088)	(−0.221 – 0.051)	(−0.213 – 0.039)	(0.012 – 0.192)	(0.587 – 1.988)

Continued

Table A.4:
Continued

	Model (1) Health status	Model (2) Sleep disorders	Model (3) Activity limitations	Model (4) Depression	Model (5) Personal relations Satisfaction	Model (6) Job satisfaction
Age 30–34 ∗ Upper Secondary	0.097	−0.081	0.036	−0.021	−0.047	−0.041
	(−0.218 – 0.413)	(−0.224 – 0.062)	(−0.136 – 0.208)	(−0.180 – 0.138)	(−0.160 – 0.067)	(−0.929 – 0.846)
Age 30–34 ∗ Tertiary	0.043	−0.059	0.078	−0.026	−0.028	−0.694
	(−0.271 – 0.357)	(−0.201 – 0.083)	(−0.093 – 0.249)	(−0.184 – 0.132)	(−0.141 – 0.085)	(−1.577 – 0.188)
Age 35–39 ∗ Upper Secondary	0.116	−0.058	−0.004	0.045	−0.036	−0.429
	(−0.176 – 0.409)	(−0.191 – 0.074)	(−0.164 – 0.156)	(−0.102 – 0.193)	(−0.141 – 0.070)	(−1.248 – 0.391)
Age 35–39 ∗ Tertiary	0.132	−0.068	0.031	0.045	−0.028	−0.779*
	(−0.161 – 0.426)	(−0.201 – 0.065)	(−0.129 – 0.191)	(−0.103 – 0.193)	(−0.133 – 0.078)	(−1.600 – 0.043)
Age 40–44 ∗ Upper Secondary	0.212	−0.102	−0.096	−0.076	−0.022	−0.022
	(−0.073 – 0.497)	(−0.231 – 0.027)	(−0.251 – 0.060)	(−0.220 – 0.067)	(−0.124 – 0.081)	(−0.820 – 0.775)
Age 40–44 ∗ Tertiary	0.239	−0.072	0.008	−0.012	−0.048	−0.899**
	(−0.051 – 0.529)	(−0.204 – 0.059)	(−0.150 – 0.166)	(−0.158 – 0.134)	(−0.153 – 0.056)	(−1.709 – −0.089)
Age 45–49 ∗ Upper Secondary	0.241*	−0.081	−0.012	−0.006	−0.026	−0.241
	(−0.037 – 0.519)	(−0.207 – 0.045)	(−0.164 – 0.139)	(−0.146 – 0.134)	(−0.126 – 0.074)	(−1.019 – 0.536)
Age 45–49 ∗ Tertiary	0.316**	−0.080	0.046	−0.002	−0.002	−0.602
	(0.030 – 0.601)	(−0.209 – 0.050)	(−0.110 – 0.201)	(−0.146 – 0.142)	(−0.104 – 0.101)	(−1.400 – 0.196)
Age 50–54 ∗ Upper Secondary	0.181	−0.026	−0.060	0.015	−0.061	0.020
	(−0.090 – 0.452)	(−0.149 – 0.097)	(−0.207 – 0.088)	(−0.122 – 0.152)	(−0.158 – 0.037)	(−0.739 – 0.779)
Age 50–54 ∗ Tertiary	0.338**	−0.072	−0.005	−0.033	−0.055	−0.413
	(0.059 – 0.617)	(−0.198 – 0.054)	(−0.157 – 0.147)	(−0.174 – 0.107)	(−0.156 – 0.045)	(−1.192 – 0.367)
Age 55–59 ∗ Upper Secondary	0.076	−0.027	−0.040	−0.004	−0.037	−0.172
	(−0.194 – 0.346)	(−0.150 – 0.095)	(−0.187 – 0.107)	(−0.140 – 0.132)	(−0.134 – 0.060)	(−0.927 – 0.583)

Continued

Table A.4: Continued

	Model (1) Health status	Model (2) Sleep disorders	Model (3) Activity limitations	Model (4) Depression	Model (5) Personal relations Satisfaction	Model (6) Job satisfaction
Age 55–59 ∗ Tertiary	0.231	−0.093	−0.002	−0.058	−0.013	−0.853**
	(−0.045 – 0.506)	(−0.218 – 0.032)	(−0.152 – 0.148)	(−0.197 – 0.081)	(−0.112 – 0.086)	(−1.622 – −0.083)
Age 60–64 ∗ Upper Secondary	0.251*	−0.059	−0.095	−0.023	−0.082	−0.135
	(−0.025 – 0.527)	(−0.185 – 0.066)	(−0.245 – 0.055)	(−0.162 – 0.116)	(−0.181 – 0.018)	(−0.906 – 0.637)
Age 60–64 ∗ Tertiary	0.212	−0.017	−0.046	−0.033	−0.078	−0.861**
	(−0.076 – 0.500)	(−0.147 – 0.114)	(−0.203 – 0.111)	(−0.178 – 0.112)	(−0.182 – 0.025)	(−1.665 – −0.057)
Age 65–69 ∗ Upper Secondary	0.051	−0.004	−0.010	0.025	−0.090∗	−0.147
	(−0.231 – 0.333)	(−0.132 – 0.123)	(−0.163 – 0.144)	(−0.117 – 0.168)	(−0.192 – 0.011)	(−0.935 – 0.641)
Age 65–69 ∗ Tertiary	0.126	−0.162**	0.016	0.024	−0.063	−0.894**
	(−0.189 – 0.441)	(−0.304 – −0.019)	(−0.155 – 0.188)	(−0.134 – 0.183)	(−0.176 – 0.050)	(−1.770 – −0.018)
Sex	−0.063***	0.019**	0.093***	0.116***	−0.005	−0.384***
	(−0.095 – −0.030)	(0.004 – 0.034)	(0.076 – 0.111)	(0.100 – 0.132)	(−0.017 – 0.006)	(−0.473 – −0.294)
Constant	4.253***	0.080	0.177***	0.188***	0.891***	6.667***
	(4.019 – 4.487)	(−0.026 – 0.185)	(0.049 – 0.304)	(0.070 – 0.306)	(0.807 – 0.976)	(6.010 – 7.323)
Observations	9,480	9,480	9,480	9,480	9,480	9,369
R-squared	0.091	0.037	0.025	0.029	0.017	0.028

Source: Elaboration of the authors based on SIP 2006.
Note: Confidence intervals in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table A.5:
Health and wellbeing outcomes by age groups and education. Switzerland

	Model (1) Health status	Model (2) Sleep disorders	Model (3) Activity limitations	Model (4) Depression	Model (5) Personal relations Satisfaction	Model (6) Job satisfaction
Age groups (25–29 Ref)						
Age 30–34	−0.555**	0.312	1.252	−0.105	−0.016	−0.594
	(−1.022 – −0.088)	(−0.164 – 0.789)	(−0.550 – 3.053)	(−1.622 – 1.412)	(−1.129 – 1.098)	(−1.934 – 0.746)
Age 35–39	−0.149	0.432**	−0.421	0.026	0.607	0.923
	(−0.558 – 0.260)	(0.014 – 0.850)	(−2.000 – 1.159)	(−1.304 – 1.357)	(−0.370 – 1.583)	(−0.227 – 2.073)
Age 40–44	0.031	0.029	−0.544	−0.597	0.583	0.269
	(−0.350 – 0.412)	(−0.360 – 0.418)	(−2.015 – 0.927)	(−1.836 – 0.641)	(−0.326 – 1.492)	(−0.785 – 1.323)
Age 45–49	−0.165	0.296	0.586	−0.213	0.365	0.364
	(−0.555 – 0.226)	(−0.102 – 0.695)	(−0.921 – 2.093)	(−1.483 – 1.056)	(−0.566 – 1.297)	(−0.716 – 1.445)
Age 50–54	−0.251	0.360*	0.300	−0.267	−0.273	0.658
	(−0.658 – 0.155)	(−0.055 – 0.775)	(−1.269 – 1.870)	(−1.589 – 1.055)	(−1.243 – 0.697)	(−0.467 – 1.783)
Age 55–59	−0.416**	0.276	1.159	0.344	0.601	0.936*
	(−0.797 – −0.035)	(−0.113 – 0.665)	(−0.312 – 2.630)	(−0.895 – 1.583)	(−0.308 – 1.511)	(−0.169 – 2.041)
Age 60–64	−0.239	0.286	0.505	−0.535	0.556	0.484
	(−0.623 – 0.144)	(−0.105 – 0.678)	(−0.974 – 1.985)	(−1.782 – 0.711)	(−0.359 – 1.471)	(−0.666 – 1.634)
Age 65–69	−0.272	0.421**	0.912	0.177	0.048	0.802
	(−0.644 – 0.100)	(0.042 – 0.801)	(−0.523 – 2.347)	(−1.032 – 1.386)	(−0.839 – 0.936)	(−0.783 – 2.388)
Education (Ref. Primary or Lower secondary)						
Upper Secondary	0.109	0.078	−1.135*	−0.898	0.507	0.033
	(−0.233 – 0.451)	(−0.271 – 0.427)	(−2.455 – 0.185)	(−2.010 – 0.214)	(−0.309 – 1.323)	(−0.941 – 1.007)
Tertiary	0.283	0.023	−1.296*	−0.990*	0.548	0.397
	(−0.064 – 0.629)	(−0.330 – 0.377)	(−2.633 – 0.040)	(−2.116 – 0.136)	(−0.279 – 1.374)	(−0.583 – 1.378)

Continued

Table A.5: Continued

	Model (1) Health status	Model (2) Sleep disorders	Model (3) Activity limitations	Model (4) Depression	Model (5) Personal relations Satisfaction	Model (6) Job satisfaction
Age 30–34 * Upper Secondary	0.567**	−0.265	−1.140	−0.180	0.013	0.944
	(0.085 – 1.049)	(−0.758 – 0.227)	(−3.001 – 0.721)	(−1.748 – 1.387)	(−1.138 – 1.163)	(−0.429 – 2.318)
Age 30–34 * Tertiary	0.431*	−0.239	−1.003	0.150	0.051	0.417
	(−0.057 – 0.918)	(−0.737 – 0.258)	(−2.884 – 0.877)	(−1.434 – 1.734)	(−1.112 – 1.214)	(−0.966 – 1.799)
Age 35–39 * Upper Secondary	0.150	−0.478**	0.473	−0.193	−0.651	−0.431
	(−0.275 – 0.574)	(−0.912 – −0.045)	(−1.166 – 2.112)	(−1.573 – 1.188)	(−1.665 – 0.362)	(−1.615 – 0.754)
Age 35–39 * Tertiary	0.008	−0.420*	0.461	−0.265	−0.647	−0.808
	(−0.422 – 0.437)	(−0.858 – 0.018)	(−1.197 – 2.118)	(−1.660 – 1.131)	(−1.672 – 0.377)	(−2.000 – 0.384)
Age 40–44 * Upper Secondary	−0.088	−0.017	0.767	0.566	−0.634	0.128
	(−0.484 – 0.309)	(−0.421 – 0.387)	(−0.762 – 2.295)	(−0.721 – 1.854)	(−1.579 – 0.311)	(−0.959 – 1.216)
Age 40–44 * Tertiary	−0.202	0.085	0.820	0.767	−0.864*	−0.192
	(−0.603 – 0.200)	(−0.324 – 0.495)	(−0.728 – 2.369)	(−0.538 – 2.071)	(−1.821 – 0.094)	(−1.288 – 0.904)
Age 45–49 * Upper Secondary	0.041	−0.248	−0.090	0.174	−0.339	0.114
	(−0.364 – 0.447)	(−0.661 – 0.166)	(−1.654 – 1.473)	(−1.143 – 1.491)	(−1.305 – 0.628)	(−1.000 – 1.228)
Age 45–49 * Tertiary	−0.096	−0.266	−0.310	0.395	−0.882*	−0.239
	(−0.508 – 0.316)	(−0.686 – 0.155)	(−1.899 – 1.279)	(−0.943 – 1.734)	(−1.864 – 0.100)	(−1.364 – 0.886)
Age 50–54 * Upper Secondary	0.075	−0.270	0.429	0.469	0.200	−0.089
	(−0.347 – 0.496)	(−0.700 – 0.160)	(−1.198 – 2.056)	(−0.901 – 1.839)	(−0.806 – 1.206)	(−1.248 – 1.069)
Age 50–54 * Tertiary	0.036	−0.160	0.186	0.346	0.091	−0.408
	(−0.393 – 0.465)	(−0.597 – 0.278)	(−1.468 – 1.840)	(−1.047 – 1.739)	(−0.932 – 1.114)	(−1.578 – 0.763)
Age 55–59 * Upper Secondary	0.244	−0.130	−0.263	−0.274	−0.546	−0.294
	(−0.154 – 0.641)	(−0.536 – 0.276)	(−1.797 – 1.271)	(−1.566 – 1.018)	(−1.495 – 0.402)	(−1.436 – 0.848)

Continued

Table A.5:
Continued

	Model (1) Health status	Model (2) Sleep disorders	Model (3) Activity limitations	Model (4) Depression	Model (5) Personal relations Satisfaction	Model (6) Job satisfaction
Age 55–59 ∗ Tertiary	0.154	−0.125	−0.386	−0.178	−0.976**	−0.696
	(−0.254 – 0.561)	(−0.540 – 0.291)	(−1.957 – 1.186)	(−1.502 – 1.146)	(−1.948 – −0.004)	(−1.854 – 0.462)
Age 60–64 ∗ Upper Secondary	0.037	−0.110	0.251	0.690	−0.638	0.256
	(−0.364 – 0.437)	(−0.519 – 0.299)	(−1.295 – 1.796)	(−0.612 – 1.992)	(−1.594 – 0.318)	(−0.939 – 1.452)
Age 60–64 ∗ Tertiary	0.010	−0.118	0.213	0.850	−0.622	0.070
	(−0.401 – 0.421)	(−0.538 – 0.301)	(−1.373 – 1.799)	(−0.485 – 2.186)	(−1.602 – 0.359)	(−1.140 – 1.280)
Age 65–69 ∗ Upper Secondary	0.140	−0.314	0.238	−0.284	−0.141	0.427
	(−0.251 – 0.531)	(−0.713 – 0.085)	(−1.270 – 1.746)	(−1.554 – 0.986)	(−1.073 – 0.792)	(−1.232 – 2.087)
Age 65–69 ∗ Tertiary	0.036	−0.126	0.274	−0.407	0.079	0.156
	(−0.368 – 0.439)	(−0.538 – 0.286)	(−1.283 – 1.831)	(−1.718 – 0.905)	(−0.884 – 1.042)	(−1.509 – 1.821)
Sex	−0.059***	0.157***	0.370***	0.617***	0.304***	0.172***
	(−0.096 – −0.022)	(0.119 – 0.195)	(0.228 – 0.512)	(0.498 – 0.737)	(0.217 – 0.392)	(0.078 – 0.266)
Constant	4.092***	1.039***	1.848***	2.030***	7.236***	7.288***
	(3.757 – 4.428)	(0.697 – 1.381)	(0.555 – 3.141)	(0.941 – 3.119)	(6.436 – 8.035)	(6.330 – 8.247)
Observations	4,889	4,889	4,889	4,889	4,889	3,863
R-squared	0.034	0.033	0.046	0.039	0.022	0.024

Source: Elaboration of the authors based on SHP 2006.
Note: Confidence intervals in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 255–291

Understanding women's well-being in Turkey

***Dilek Yıldız*[1,*]*, Hilal Arslan*[2] *and Alanur Çavlin*[3]**

Abstract

The results of empirical studies focusing on gender differences in subjective well-being based on either national or comparative international data are inconclusive. In Turkey, where levels of gender inequality are high, women tend to report higher levels of life satisfaction than men. This study investigates the relationship between factors related to women's empowerment and life satisfaction for both ever-married and never-married women using the 2018 Turkey Demographic and Health Survey (TDHS), which collected data on life satisfaction for the first time in a TDHS series. The results show that in addition to their material resources and living environment, factors related to women's agency – i.e., education and participation in decision-making – are associated with women's levels of life satisfaction.

Keywords: women's empowerment; subjective well-being; life satisfaction; Turkey; 2018 TDHS

1 Introduction

Many countries in the world have committed to achieving gender equality and to empowering all women and girls by providing them with equal rights and opportunities, as well as protecting them from violence and discrimination (Sustainable Development Goals (SDGs), United Nations 2019).[4] Although gender inequality

[1]International Institute for Applied Systems Analysis (IIASA), Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

[2]Department of Social Research Methodology, Institute of Population Studies, Hacettepe University, Ankara, Turkey

[3]Department of Demography, Institute of Population Studies, Hacettepe University, Ankara, Turkey

[*]Correspondence to: Dilek Yıldız, yildiz@iiasa.ac.at

[4] Detailed information on SDG Goal 5: Achieve gender equality and empower all women and girls is available at https://www.un.org/sustainabledevelopment/gender-equality/.

DOI: 10.1553/populationyearbook2021.res2.3

has been a major concern for most societies, and especially for developing countries, the relationship between gender and subjective well-being is one of the most puzzling issues in happiness research. While the results of various studies show that women tend to report higher average levels of well-being than men (Blanchflower and Oswald 2004; Dolan et al. 2008; Haller and Hadler 2006; Lalive and Stutzer 2004); others indicate that the average level of life satisfaction is higher for men than for women (Tesch-Römer et al. 2008; Stevenson and Wolfers 2009).

While Turkey has been investing in gender equality for nearly a hundred years, reports from international gender equality indexes show that Turkey is still far from achieving this goal. Turkey ranks 68 out of 189 countries on gender equality, with a score of 0.306 in the 2019 United Nations Development Program (UNDP) Gender Inequality Index. In the World Economic Forum (WEF) Global Gender Gap Index, Turkey was in 105th place among 115 countries in 2006, but was in 130th place among 150 countries in 2020. Despite the Turkish government's ratification of the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) in 1985 and of the Council of Europe Convention on Violence against Women and Domestic Violence (Istanbul Convention) in 2012,[5] the implementation of these conventions has been flawed, as reports of gender-based discrimination and violence against women, including intimate partner femicide, have continued to rise. After 2010,[6] there was a major backlash against efforts to advance women's human rights and gender equality in Turkey. First, gender equality as a norm has since been abandoned in the state's policies and legal framework, and has been replaced with "gender equity". Furthermore, as the legal framework has been altered by the abandonment of gender equality as a norm for state policies, a significant deterioration of women's rights and gender equality has taken place in Turkey (Güneş-Ayata and Doğangün 2017; Hülagü 2021). For example, changes in the regulations have excluded child marriage from being classified as a criminal act; religious authorities have been given a legal mandate to ratify marriages; and restrictions on alimony rights have been proposed.

This radical gender backlash has also led to deepening inequalities in education and employment, which are the major sources of women empowerment. According to data provided by the Turkish Statistical Institute (TurkStat), Turkey is still falling

[5] Council of Europe Convention on preventing and combating violence against women and domestic violence, available at https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/210/signatures. Recently, Turkey announced the withdrawal from the Istanbul Convention with decree No. 3718 in the Official Gazette on March 20, 2021.

[6] There have been major interventions aimed at improving gender equality since the 1990s. The first set of such interventions were initiatives by the government to improve the administrative structures and the legal framework: e.g., the reform process to improve legislation, including the revision of Article 10 of the Constitution to emphasise that ensuring that gender equality is achieved is the obligation of the states; the adoption of the Family Protection Law, which requires that women be protected against violence inflicted by men in the domestic realm; and the Istanbul Convention on combatting violence against women, which went into force in 2014.

short of the target of 100% female literacy set for 2000, which was one of the commitments the country made under the Beijing Declaration.[7] In Turkey, the literacy rate was 95.1% for males and 80.6% for females in 2008; and was 98.7% for males and 92.4% for females in 2018 (TurkStat 2019a). While the literacy data for 2018 indicate that the gender gap had been closed in primary education (92.1% of girls and 91.8% of boys were enrolled) and in secondary education (93.6% of girls and 92.9% of boys were enrolled), it should be noted that the educational enrolment figures did not cover the full population across age groups, and that there were significant disparities among provinces (TurkStat 2019a). Moreover, in 2018, the percentage of girls over age 14 who were enrolled in distant education had increased (62%); and 56% of the female students were enrolled in religious vocational secondary schools.[8] Furthermore, at age 25, 24.6% of men, but only 16.5% of women, were high school graduates or the equivalent (TurkStat 2019a).

In addition to these gender disparities in education, women in Turkey have much lower rates of labour force participation and employment than their male counterparts. According to Gender Statistics 2019, 72% of men, but just 34% of women, are employed (TurkStat 2019a). In addition, the largest share of women who work are employed in the service sector, with the majority working in irregular, underpaid jobs without social security. Based on the International Labour Organization's (ILO) definition of "vulnerable employment", 32% of women in Turkey are either "unpaid family workers" or are "own-account workers" (Toksöz and Memiş 2018). Moreover, the results of the Structure of Earnings Survey show that there are persisting inequalities in the earnings of men and women who do the same job, and that the gender pay gap is 20% among higher education graduates (TurkStat 2018). In addition, compared to men, women in Turkey have much higher rates of unemployment, particularly in the non-agricultural sectors. Out of 20 million women living in Turkey in 2019, 57%, or 11,359,000, were out of the labour force because they were occupied with housework. Moreover, according to the results of the TurkStat Time Use Survey (2014–2015), household labour has been disproportionately performed by women (TurkStat 2016). The survey of the daily activities of individuals aged 10 and older showed that the average time spent on household chores and care responsibilities was four hours and 35 minutes for women, compared to only 53 minutes for men (TurkStat 2016).

Despite the persistence of gender-based discrimination against women in Turkey, the results of the TurkStat Life Satisfaction Survey showed that the percentage of women who reported being "happy" was higher than that of men, and that

---

[7] The Fourth World Conference on Women, Beijing Declaration and Platform for Action, available at https://www.un.org/en/events/pastevents/pdfs/Beijing_Declaration_and_Platform_for_Action.pdf.

[8] In 2012, the system known as 4 + 4 + 4 was introduced. Under this system, secondary education was made compulsory, but distant education was permitted after the first four years. Until 2012, attending eight years of in-class education was compulsory. After 2012, female students started to enrol in distant education schools, prompting debates about women's exclusion from educational opportunities together with men and chances to attend school with their peers.

the happiness gap between men and women in Turkey increased from 3% to 9% between 2003 and 2019 (TurkStat 2020). When we examine the happiness trends over the last 25 years in Turkey using World Values Survey (WVS) data, we can see that the mean level of life satisfaction has returned to its initial level of 6.5, after increasing to 7.4 in 2011 (Inglehart et al. 2014). Over this period, there was a sharp decrease in average levels of happiness due to the economic crisis of 2001, when the lowest scores for both men (5.8) and women (5.4) were recorded. However, in the wake of the economic recovery and the substantial improvements in material living standards over the following decade, the highest levels of self-reported well-being for women (7.3) and men (7.2) were reached in 2011. Although women living in Turkey have tended to report slightly higher levels of life satisfaction than their male counterparts, the size of the gender gap in happiness was negligible between 1996 and 2018. Just as the findings regarding the relationship between gender and happiness have been contradictory, some empirical studies for Turkey have shown that women report higher levels of life satisfaction than men (Caner 2014, 2016; Eren and Aşıcı 2017; Ekici and Koydemir 2014), whereas other studies found that the opposite is the case (Akın and Şentürk 2012; Dumludag 2013). Nevertheless, only a few demographic and happiness studies have focused on the factors influencing the subjective well-being (SWB) of women living in Turkey (Akay and Timur 2017; Çakıroğlu-Çevik 2016; Ermiş-Mert 2020, Şengül and Lopcu 2020). Moreover, existing research using national representative data has tended to focus only on women's labour market status and living standards, rather than on gender inequalities in all life domains. Thus, little is known about the demographic, economic and social characteristics that influence women's well-being in Turkey in relation to women's empowerment.

Against this background, this paper investigates the factors that influence women's satisfaction with life in Turkey by using the women's empowerment and liveability approach. We focus on women's socio-demographic characteristics, external living environment, material and social resources, and agency in relation to decision-making power; as well as the power dynamics in couples. Our aim in this study is to contribute to a growing area of research on the relationship between gender inequality and women's life satisfaction by concentrating on a country with a socio-political and cultural context that differs from those of western countries.

Using data from the 2018 Turkey Demographic and Health Survey (TDHS) (HUIPS 2019), we employ quantitative methods of enquiry to investigate the relationship between women's empowerment and subjective well-being among women living in Turkey. In this study, the life satisfaction of women is measured using a tried and tested 10-point scale in which a value of one means completely dissatisfied, and a value of 10 means completely satisfied. First, using descriptive statistics and Chi-squared tests for independence, we analyse the association between women's socio-demographic characteristics, external living environment and material and social resources; as well as women's agency in relation to decision-making power, the power dynamics in couples, and women's life satisfaction. Then, we apply generalised ordered logit models (gologit) to investigate the effects of these factors on women's life satisfaction.

The article is structured as follows. The background section critically reviews the literature on gender and subjective well-being by focusing on the conceptualisation of the relationship between women's empowerment and life satisfaction in Turkey, while the second section introduces the data, methods and analytical strategy for the data analysis. The third section presents the results of both descriptive and multivariate statistical analyses on the factors affecting women's life satisfaction. The last section concludes with a discussion of the empirical findings on improvements in the subjective well-being of women living in Turkey.

2 Literature on subjective well-being and women's empowerment

Gender inequalities are evident throughout the life courses of women; and because these inequalities affect factors related to women's empowerment, they in turn affect women's well-being. Accordingly, in this section, we first present the relevant literature on gender and subjective well-being. This is followed by an introduction of the conceptual model of the study on women's empowerment and life satisfaction, and then by a discussion of the results of the existing empirical studies on this topic.

2.1 Gender and subjective well-being

Subjective well-being is defined as an individual's positive judgment of his or her overall quality of life (Veenhoven 2000). The fundamental components of subjective well-being are as follows: an individual's personal assessment of his or her life at the cognitive level, i.e., satisfaction with life; and an individual's emotional reaction to life events at the affective or hedonic level, i.e., happiness (Andrews and Withey 2005; Campbell et al. 1976; Diener 1994; Michalos 1980). In order to overcome the difficulties associated with drawing clear-cut boundaries between subjective well-being, life satisfaction and happiness, we adopt a hybrid view and use the happiness and SWB terms interchangeably, while measuring them under the umbrella term "life satisfaction". In line with the existing knowledge on this issue, we expect to find that rather than representing separate constructs, there is a considerable degree of correlation between life satisfaction and happiness, since both involve affective and cognitive evaluations of life events and conditions (Diener 1994; Graham 2005; Schyns 1998).

In subjective well-being research, gender has generally been treated not as an explanatory factor in the universal happiness formula, like income, age, education and having partner; but as a "control" variable used to limit measurement error. While previous research on this topic has mainly concentrated on gender differences in subjective well-being, the results of empirical studies based on either national or comparative cross-national data have been inconclusive. If gender is a direct reflection of people's assessments of their lives, we would expect to find that women report lower levels of well-being than men, particularly in societies with strong patriarchal regimes that foster gender-based inequalities in all domains of men's and women's lives. Accordingly, the findings of some studies have shown that men have much higher life satisfaction levels than women (e.g., Boncompte and Paredes 2020; Haring et al. 1984; Meisenberg and Woodley 2015; Stevenson and Wolfers 2009; Wood et al. 1989; Zuckerman et al. 2017). Nevertheless, a number of other studies have reported statistically significant results indicating that women have higher levels of subjective well-being than men (e.g., Arrondo et al. 2020; Arrosa and Gandelman 2016; Blanchflower and Oswald 2004; Dolan et al. 2008; Fujita et al. 1991; Graham and Chattopadhyay 2013; Haller and Hadler 2006; Van Praag and Ferrer-i-Carbonell 2008; Lalive and Stutzer 2004; Zweig 2015). Furthermore, a few studies have found no differences in self-reports of individual well-being between women and men, particularly after controlling for the relationship between well-being and other socio-economic and demographic characteristics (e.g., Clemente and Sauer 1976; Inglehart 1990; Mayungbo 2016; Okun and George 1984; Roothman et al. 2003; Shmotkin 1990; Tiefenbach and Kohlbacher 2013).

In addition, at a macro level, it appears that gender differences in subjective well-being may be paradoxically associated with cultural, economic and political conditions that produce gender (in)equalities in various societal contexts. There is, for example, empirical evidence that people's assessments of their well-being are irrespective of their gender in societies that foster gender equality in different domains of life, such as work-life balance, women's participation in decision-making, and equal access to high-quality health and educational services (Audette et al. 2019; Bjørnskov et al. 2007; Inglehart and Welzel 2005; Inglehart et al. 2008; Jorm and Ryan 2014; Ruth and Napier 2014; Tesch-Römer et al. 2008). On the other hand, there is also evidence that women tend to report lower levels of subjective well-being in certain countries with political structures, economic sources and opportunities, and cultural norms that are supportive of gender equality than women in some African, Muslim and East Asian societies with strong patriarchal values (Stevenson and Wolfers 2009; Tiefenbach and Kohlbacher 2013; Vieira Lima 2011).

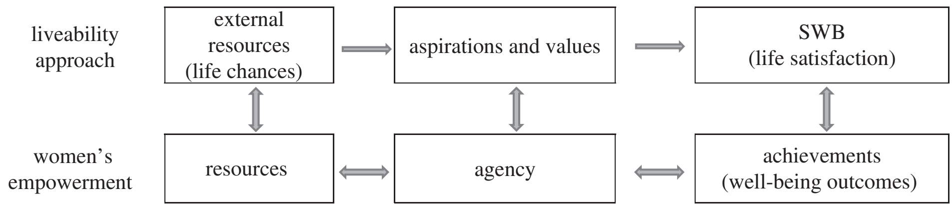
In order to move beyond analyses of the nature of the relationship between gender and life satisfaction based on disaggregating happiness by sex, our study contextualises the gender-happiness puzzle under conditions of gender (in)equality by focusing on subjective well-being and women's empowerment.

2.2 Subjective well-being and women's empowerment

In this study, we conceptualise subjective well-being using the "liveability approach", which focuses on the argument that individuals' living circumstances play a significant role in their appraisals of their quality of life contrary to hedonic treadmill or set-point theory[9] (Veenhoven 1996). This approach explains life satisfaction in terms of the relationship between people's *external resources,* namely, their objective living conditions and their evaluation of them; and people's *aspirations and values,* which play a mediating role in their perceptions of their life chances. Furthermore, this approach explains the link between the distribution of life chances in the social environment and how people feel about their lives. Veenhoven (2006) extended his liveability theory so that it referred not only to the characteristics of an individual's social surroundings or the quality of the society, but also to an individual's position in this social setting, which is closely associated with the inherently disadvantaged status of women in unequal gender regimes. Second, from a broader perspective, women's empowerment can be defined as enabling women to act socially, economically, politically and legally by enabling them to access to rights, services, facilities and resources that were previously denied or restricted. According to Rowlands (1995), empowerment denotes women's maximisation of opportunities independent of external restrictions and interventions. Another perspective on empowerment is set within a feminist framework, in which embedded power relations are reversed by giving the control over resources to women. This can strengthen a woman's ability to choose; i.e., it can enable a woman to close the gap between what she desires and what she actually has (Batliwala 1994). In our study, we adopted Kabeer's (1999) conceptualisation of women's empowerment in relation to gender and subjective well-being. In other words, the concept of empowerment that is introduced into our study is based on a feminist approach. Indeed, women's empowerment is a concept that is mostly used in the development literature, with its liberal meanings and connotations. In contrast to this mainstream use of the concept, we prefer to adopt Kabeer's conceptualisation in which she underlines the significance of power, and of women's liberation from the patriarchal societal order. In Kabeer's modelling of empowerment, she emphasises the role of the resources, agency and achievements that enable women to pursue their lives based on their own life choices. Here, "*resources*" refer to the pre-conditions, which are composed of material (e.g., income) and non-material (e.g., human and social) means and forms of capital. As another component of women's

9 Nevertheless, whether socio-demographic factors directly influence how people interact with their social environment, and the significance of these factors in estimating individual well-being, are open to debate. Some psychological theories of well-being have argued that people's socio-economic environment plays a negligible role in their subjective well-being. They claim that people's assessments of their quality of life depend solely on personal and relational factors, such as on their personality traits and social relations (Kahneman and Krueger 2006). Nevertheless, liveability theory has defended the salience of individuals' living circumstances in shaping their perceived quality of life. In parallel with this approach, there is a substantial amount of empirical evidence that living conditions have decisive effects on individual well-being. It has, for example, been reported that nearly one-third of the variation in life satisfaction scores is due to properties of people's external circumstances (Noll 2002; Saris 2001; Seghieri et al. 2006; Yetim 1993).

Figure 1:
The conceptual model of the relationship between gender and happiness



empowerment, *agency* is defined as a woman's potential to both act in alignment with her goals, and to plan and attribute meaning to her activities. In other words, a woman's agency comprises her capabilities of decision-making and of handling her relationships in everyday life. Finally, a woman's *achievements* are the well-being outcomes during her life course that are related to both resources and agency in Kabeer's conceptual model of women's empowerment. In order to overcome the inconclusive nature of the relationship between gender and happiness, and to discuss women's well-being in relation to gender inequality, we propose a common conceptual model that combines both the liveability and the women's empowerment approach (See Figure 1).

According to this combined model, a woman's subjective well-being is related not only to securing her basic standard of living by accessing material resources – i.e., income, wealth and paid employment – but also to having a say over how the household income is spent, and the self-evaluation of those resources; and to having the power and the capability to make life choices and decisions that result in a decent quality of life. Few of the existing empirical studies have discussed the role of women's resources and agency in explaining both women's empowerment and subjective well-being. For instance, as a material resource, having adequate income is important for a woman's autonomy and freedom to make choices, and can thus affect her sense of well-being. Some studies have provided evidence that having an income can empower a married woman to call for a fairer division of household labour with her husband (Ball and Chernova 2008; Treas et al. 2011). Nevertheless, considering the relationship between resources and agency, if a married woman adopts traditional gender roles and sees her husband as a breadwinner, her own income would not affect her personal assessment of her quality of life (Clark 1997). Similar to income, the relationship between employment status and happiness among women is a controversial issue, because even though having paid work can contribute to a woman's personal autonomy and self-fulfilment, it does not always produce concrete well-being outcomes. While some studies have shown that being employed has a positive impact on women's subjective well-being (e.g., Tay et al. 2014), others have found that being in paid work can have a negative impact on women's life satisfaction, because women may experience work-family conflicts, as well as pressure to take on nonconforming gender roles, especially for motherhood (e.g., Booth and Van Ours 2009; Böhnke 2005; Mencarini and Sironi 2012; Mitsuyama and Shimizutani 2019; Stutzer and Frey 2006; Torosyan and Pignatti 2020). Thus, there is evidence that, on average, housewives report having higher levels of life satisfaction than women with paid work and a regular employment status (Haller and Hadler 2006; Treas et al. 2011). Furthermore, education has been shown to have a positive influence on the subjective well-being of women, as it increases their decision-making power within the household and their capability to control their own life (Mitsuyama and Shimizutani 2019). Regarding agency, Ali and ul Haq (2006) have pointed out the contribution of women's autonomy to levels of self-reported life satisfaction among Pakistani women.

There are only a few existing nationally representative studies that have examined how each component of women's empowerment is reflected in women's individual assessments of their overall quality of life in Turkey. In her empirical analysis of data from the European Quality of Life Survey (2007), Çakıroğlu-Çevik (2016) found that having a higher level of education and being married, rather than material living conditions, were positively associated with women's quality of life and happiness. In addition, her findings underlined the positive effects on women's subjective well-being of social ties and family relationships (family, parents, siblings, relatives and neighbours) within the private domain. Using more recent data from the TurkStat Life Satisfaction Survey (LFS), Ermiş-Mert (2020) examined the relationship between gender, income and job satisfaction, and their influence on the self-reported happiness levels of women. The results of the study showed that, on the one hand, being an unpaid family worker was positively associated with women's life satisfaction; and, on the other hand, the job satisfaction of working women contributed to their global life satisfaction. In a similar vein, in an analysis of the distribution of happiness across different employment status groups by using World Values Survey Turkey data (1996–2011), Arslan (2020) discussed the paradox in Turkey of happy housewives who reported higher levels of life satisfaction than employed and retired women. Furthermore, in their analysis of the Income and Living Conditions Survey of Turkey conducted by TurkStat, Şengül and Lopcu (2020) found that in Turkey, widowed/divorced female household heads had lower levels of subjective well-being than married women. In addition, their results indicated that, on average, the higher a woman's level of education, the higher her level of life satisfaction.

Despite the findings of these empirical studies, there is still considerable uncertainty about the relationship between gender and subjective well-being, and little is known about the factors that explain women's life satisfaction in relation to women's empowerment, particularly in Turkey. Therefore, against this background, we examine to what extent women's empowerment affects the life satisfaction of women living in Turkey by analysing the effects that women's material and social resources, as well as their agency, have on their subjective well-being outcomes.

3 Data, variables and methods

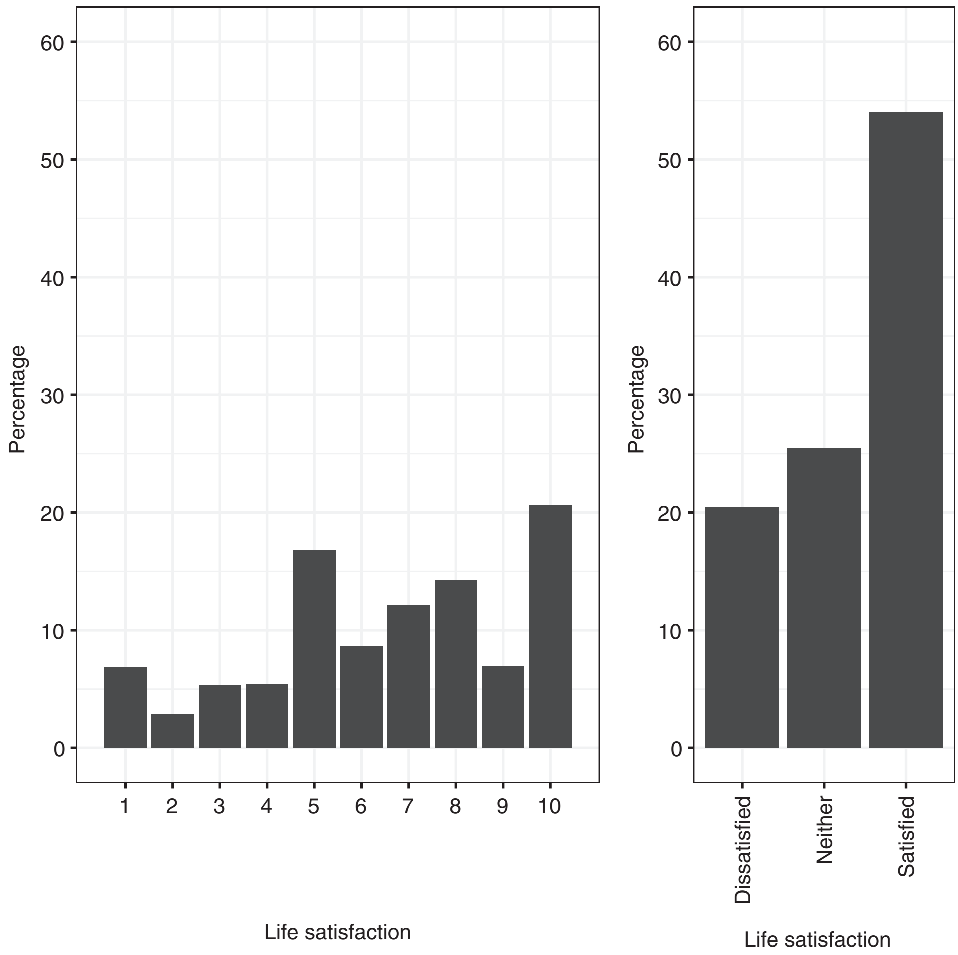
This research uses individual-level data collected by the 2018 Turkey Demographic and Health Survey (TDHS). Similar to other DHS surveys, the TDHS collects very rich information on women aged 15 to 49, including on their sexual behaviour and reproductive health, migration, work and marriage histories, among many other characteristics. The 2018 TDHS collected data on 7,346 women. Of these respondents, slightly more than 25% were never-married, while 75% had been married at least once. A majority of the women surveyed (78%) were living in an urban area.

The 2018 TDHS collected information on life satisfaction for the first time in a DHS survey conducted in Turkey. This newly collected indicator opened a new door for a more detailed investigation of women's subjective well-being, its association with different characteristics, and the inequalities in well-being among different groups. The information for the dependent variable of life satisfaction was collected with the question: "All things considered, how satisfied are you with your life as a whole?" In the current analysis, we investigate independent variables in five groups: socio-demographic characteristics, material resources, social resources, decision-making and couple power dynamics. The variables in the last two groups are only relevant for ever-married women.

In order to investigate the determinants of life satisfaction, we assign the life satisfaction variable to three groups: dissatisfied (1–4), neither satisfied nor dissatisfied (neither) (5 and 6) and satisfied (7–10). The percentage distributions of the 10-point scale and the grouped variables are shown in Figure 2 in Panel A and Panel B, respectively. In both groups, about 54% of women reported that they were satisfied with their life. However, slightly more ever-married women (22%) than never-married women (18%) reported that they were dissatisfied with their life. Table A.1 provides the percentages and the frequencies of life satisfaction of never-married and ever-married women in the sample.

The analysis presented in this paper is performed separately for ever-married and never-married women. This is because the preliminary analyses showed that never-married women and ever-married women differed in terms of their socio-demographic characteristics, resources and agency. The greatest difference between these two groups of women was in their age distributions. In other words, these two groups consisted of women from different birth cohorts. While more than 80% of the ever-married women were aged 25 or older at the time of the survey, 78% of the never-married women were younger than 25 years old. This difference in the age distributions of these two groups was also reflected in the distributions of other variables, such as educational attainment, external living environment and internet use. Furthermore, the TDHS collected detailed information about ever-married women, and about their current and previous partners. Therefore, some of the variables are only available for the ever-married women sample. The rest of the characteristics of the surveyed women are provided in detail in Table A.2 in the appendix.

Figure 2:
Life satisfaction, (A) 10-point Likert scale, (B) 3-point scale



First, based on the conceptual model of the relationship between women's empowerment and subjective well-being, the explanatory variables for the *resource* dimension are categorised into three groups: (1) socio-demographic characteristics and external living environment, (2) material resources, and (3) social resources. A respondent's *socio-demographic characteristics* consist of her age, educational attainment, mother tongue (a proxy for ethnic identity) and number of children (only for ever-married women). A respondent's external living environment includes her region and place of residence. A respondent's *material resources* include her employment, the money she has to spend by herself, and the assets (house, land and car) she owns. A respondent's *social resources* are measured by the extent to which she goes out for meals, organises home meetings, uses the internet and goes to the cinema/theatre. Second, for the *agency* dimension of the model, the respondent's decision-making power and couple power dynamics are considered. These characteristics are also indicators for Goal 5 (Gender Equality) of the Sustainable Development Goals (UN 2015). The variables used to measure *decision-making power* are the respondent's decisions regarding contraception use and health care. Finally, the *couple power dynamics* variables take into account the characteristics of the respondent's last partner; i.e., the partner's educational attainment, social security status and mother tongue.

For the data analysis, we start with descriptive analysis and Chi-squared tests for the independence of rows and columns in which the rows are the selected variables and the columns are the life satisfaction variables. Second, for the multivariate analysis, we apply gologit models for never-married and ever-married women separately.

The natural starting point for investigating the association between women's life satisfaction and their characteristics is ordinal logistic regression. However, for this particular data set, the proportional odds assumption that is required for ordinal logistic regression – which is also referred to as parallel lines – is not held. Therefore, in this paper we employ several different gologit models. Gologit models are often used as alternatives to ordinal logistic regression since the assumptions of the ordinal logistic regression model can be relaxed for selected variables (Williams 2016). We present five additive models in which we start with the sociodemographic characteristics (Model 1) and add one-by-one the remaining four groups of variables categorised as material resources, social resources, decision-making power, and couple power dynamics. Finally, we present a parsimonious model that only includes the variables that are significant in at least one of the previous five models.

4 Data analysis

4.1 Descriptive statistics

Figure 3 presents the distribution of life satisfaction by the characteristics of ever-married women along with the Chi-squared test for the independence of the rows and columns. The same figure for never-married women is provided in the appendix. Figure 3 shows that younger women, women with higher education and women working with social security had higher levels of life satisfaction than the other respective groups. The Chi-squared test does not show any significant difference between the place of residence and life satisfaction. However, all of the other variables related to women's basic characteristics and their economic status are found to be significant.

Figure 3:
Distribution of life satisfaction by ever-married women's characteristics

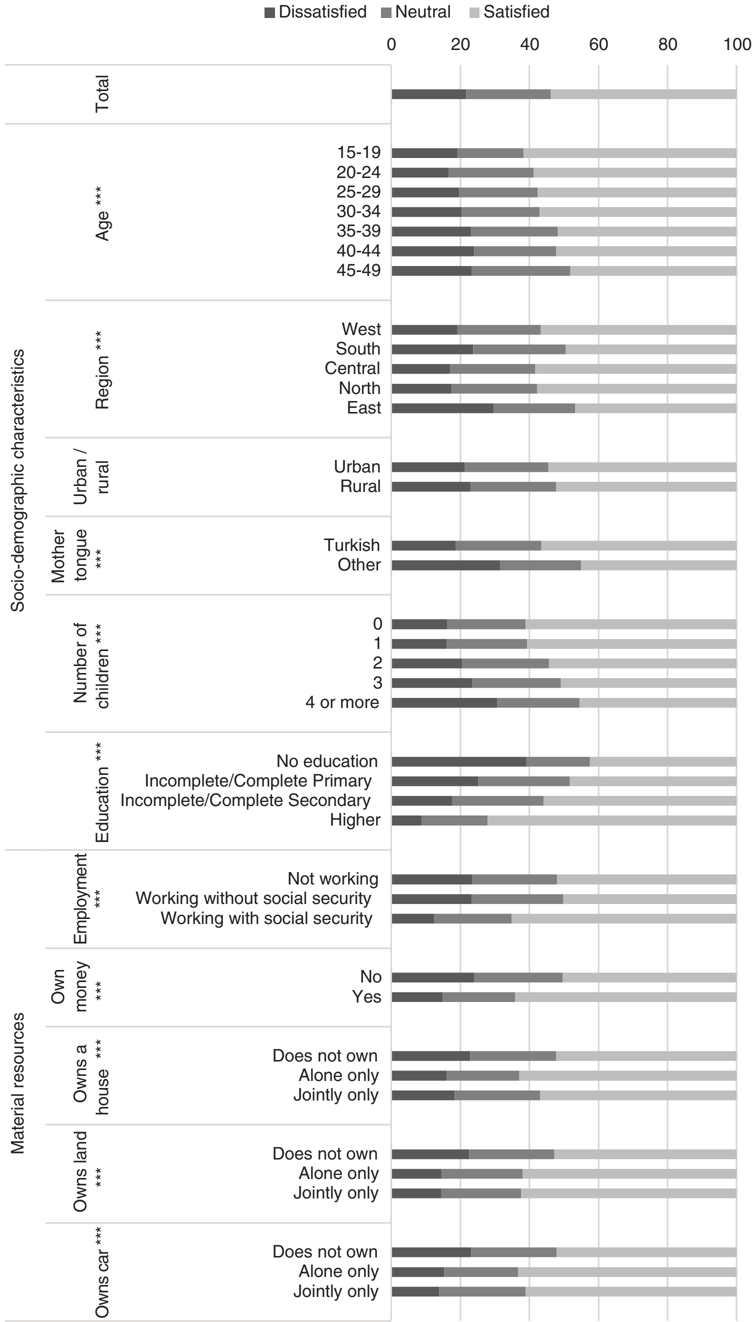
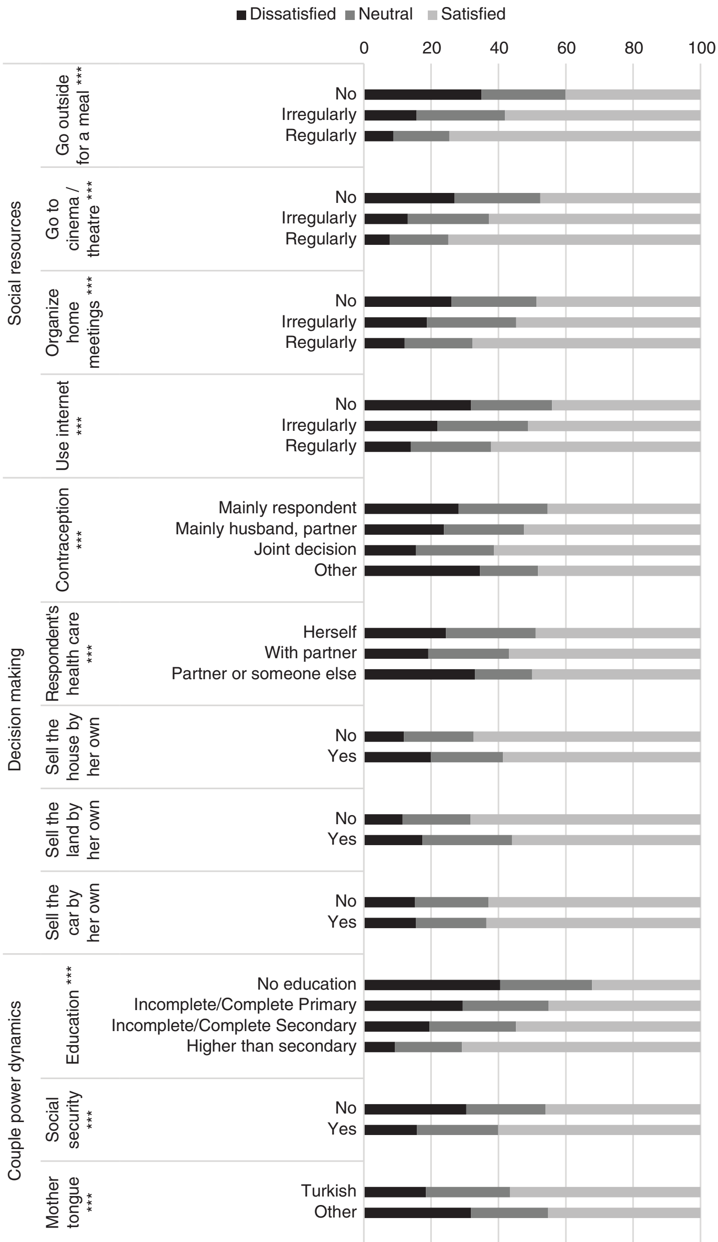


**Figure 3:
Continued**



Note: Chi-squared test for independence: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

The second half of Figure 3 shows the distribution of life satisfaction and selected variables related to the respondents' socio-demographic characteristics, material and social resources, decision-making power and agency. As expected, the results indicate that women who socialised regularly had higher levels of life satisfaction than the other groups. Similarly, women who had a partner with higher than secondary education had higher levels of life satisfaction than other groups of women. The observed differences between all variables except being the sole decision-maker for the sale of assets and life satisfaction are shown to be significant. It is worth noting that the majority of women did not hold their own assets. Less than a quarter of ever-married women (23.4%) owned a house alone or jointly, and even smaller shares of women owned land (10.9%) or a car (17.9%).

The distribution of life satisfaction by the characteristics of never-married women is provided in Appendix Figure A.1. The age, region, education, mother tongue, employment, having money to spend and living standards variables are found to be significant. Surprisingly, the age group in which the largest proportion of women had a high level of life satisfaction (65.8%) is shown to be ages 30–34.

4.2 Results of the multivariate analysis

As was mentioned above, we employ gologit models to investigate the association between women's life satisfaction and their characteristics. Gologit models collapse categories of the dependent variable and run a series of binary logistic regressions. In our case, in which the dependent variable has three categories, the model first combines the neither satisfied nor dissatisfied category (2) with the satisfied category (3), and then calculates the odds ratios for this new category relative to those for the dissatisfied category (1). Tables 1 and 2 present a comparison of the odds ratios for category 1 with those for categories 2 and 3. The model also collapses the categories of dissatisfied (1) and neither satisfied nor dissatisfied (2), and compares this new category with the satisfied category (3). In other words, categories 1 and 2 are compared with category 3. We present the odds ratios of these comparisons for ever-married and never-married women in the appendix. In order to avoid the unnecessary complexity of the original gologit model, we use the gologit2 routine in Stata, which relaxes the parallel line assumption only when it is violated with the autofit option (Williams 2005). In the next two subsections, we present the results of the gologit models separately for ever-married and never-married women. As was mentioned above, for both groups, the presented referent group consists of dissatisfied women. Tables 1 and 2 show the odds ratios and their respective significance levels. Three different regression models are employed for both ever-married and never-married women: (1) the respondent's socio-demographic characteristics, (2) the respondent's material resources and (3) the respondent's social resources. For

Table 1:
Gologit model results (Odds ratio) for ever-married women

Dissatisfied	Model 1	Model 2	Model 3	Model 4	Model 5	Parsimonious
Socio-demographic characteristics						
Respondent's current age	0.991	0.986**	0.983***	0.981*	0.984	0.981*
Region (East)						
West	1.174	1.154	1.010	1.164	1.102	1.100
South	0.975	0.954	0.918	1.177	1.18	1.112
Central	1.299**	1.278*	1.187	1.395	1.399	1.333
North	1.190	1.163	1.016	1.280	1.123	1.219
Type of place of residence (Urban)						
Rural	1.199**	1.216**	1.435***	1.329*	1.329	1.367*
Number of children (4 or more)						
0	1.101	1.083	0.905	1.597	1.794	
1	1.106	1.089	0.911	0.753	0.808	
2	1.017	1.001	0.838	0.699	0.716	
3	1.048	1.050	0.949	0.872	0.951	
Mother tongue (Turkish)						
Other	0.740***	0.741**	0.783**	0.875	0.862	0.909
Educational attainment (No education)						
Incomplete/Complete primary	1.516***	1.471**	1.276	1.543	1.568	1.477
Incomplete/Complete secondary	2.313***	2.136***	1.466*	1.441	1.405	1.312
Higher than secondary	4.647***	3.590***	2.074***	1.804	1.385	1.862*
Material resources						
Employment (Working with social security)						
Not working		0.961	0.985	0.811	0.862	
Working without social security		0.829	0.866	0.830	0.857	
Has own money to spend (No)						
Yes		1.330***	1.142	1.310	1.343	1.368*
Owns a house alone or jointly (Does not own)						
Alone only		1.226	1.170	1.135	1.106	1.134
Jointly only		1.145	1.088	1.099	1.026	1.092
Owns land alone or jointly (Does not own)						
Alone only		1.139	1.133	1.059	0.942	1.04
Jointly only		1.384**	1.364*	1.099	1.075	1.074
Owns a car alone or jointly (Does not own)						
Alone only		1.067	0.924	0.876	0.876	0.925
Jointly only		1.465**	1.387*	0.858	0.803	0.875
Social resources						
Goes out for meals (No)						
Irregularly			2.197***	2.041***	1.882***	2.018***
Regularly			2.584***	3.136***	2.809***	3.041***

Continued

Table 1:
Continued

Dissatisfied	Model 1	Model 2	Model 3	Model 4	Model 5	Parsimonious
Organises home meetings (No)						
Irregularly			1.217*	1.104	1.071	1.082
Regularly			1.674***	1.631**	1.607**	1.545**
Uses the internet (No)						
Irregularly			1.103	0.993	0.916	
Regularly			1.065	1.017	0.952	
Go to the cinema/theatre (No)						
Irregularly			1.238*	2.054***	1.609**	2.074***
Regularly			1.361	3.361*	1.452	3.578**
Decision-making						
Decision-maker for contraception use (Mainly respondent)						
Mainly husband, partner				1.649	1.658	1.662
Joint decision				1.780***	1.781***	1.773***
Other				1.001	1.058	1.014
Person who usually decides on the respondent's health care (Herself)						
With partner				1.516***	1.390**	1.512***
Partner or someone else				0.820	0.733	0.837
Couple power dynamics						
Partner's educational attainment (No education)						
Incomplete/Complete primary					1.501	
Incomplete/Complete secondary					1.662	
Higher than secondary					2.214	
Partner has social security (No)						
Yes					1.373	
Partner's mother tongue (Turkish)						
Other					1.033	
Observations	5484	5480	5480	1912	1755	1912
Pseudo R-squared	0.029	0.0348	0.0594	0.0921	0.0860	0.0889

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

ever-married women, two additional models are considered: (4) decision-making within the relationship and (5) couple power dynamics.

It is worth noting that the models are additive in nature. In other words, we start with the socio-demographic characteristics of women in Model 1 and add variables at each step. Hence, Model 5 includes all variables from the five categories. Considering more variables in each consecutive model results in minor changes in both the value and the significance of odds ratios. These changes happen when the new variable is correlated with either another independent variable that is already in the model or the dependent variable. For example, such a change is expected and

Table 2:
Gologit model results (odds ratio) for never-married women

Dissatisfied	**Model 1**	**Model 2**	**Model 3**	**Parsimonious**
Socio-demographic characteristics				
Respondent's current age	0.991	0.985	0.985	
Region (East)				
West	1.144	1.099	1.019	1.059
South	0.812	0.818	0.773	0.802
Central	1.410*	1.405*	1.342	1.426*
North	1.431*	1.398	1.336	1.430*
Type of place of residence (Urban)				
Rural	0.998	1.088	1.126	
Mother tongue (Turkish)				
Other	0.761	0.804	0.843	
Educational attainment (No education)				
Incomplete/Complete primary	0.792	0.824	0.813	0.851
Incomplete/Complete secondary	1.579	1.507	1.299	1.741
Higher than secondary	2.470*	2.15	1.691	2.222
Material resources				
Employment (Working with social security)				
Not working		0.984	1.011	
Working without social security		0.68	0.727	
Has own money to spend (No)				
Yes		1.414**	1.337*	1.305*
Owns a house alone or jointly (Does not own)				
Alone only		1.869	1.972	
Jointly only		1.361	1.39	
Owns land alone or jointly (Does not own)				
Alone only		0.434	0.435	
Jointly only		0.842	0.811	
Owns a car alone or jointly (Does not own)				
Alone only		1.337	1.229	
Jointly only		1.116	1.04	
Social resources				
Goes out for meals (No)				
Irregularly			1.436**	1.784***
Regularly			1.569	1.573*
Organises home meetings (No)				
Irregularly			1.117	1.108
Regularly			1.355	1.349
Uses the internet (No)				
Irregularly			0.912	
Regularly			1.254	
Goes to the cinema/theatre (No)				
Irregularly			0.944	
Regularly			0.966	
Observations	1862	1862	1862	1862
Pseudo R-squared	0.0199	0.0259	0.0328	0.029

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

observed in the odds ratios of educational attainment categories since it is correlated with many other characteristics that are considered in the model.

4.2.1 Ever-married women

In this section, we discuss the determinants of being satisfied or being neutral (hereafter being satisfied/neutral) rather than being dissatisfied among ever-married women. Table 1 shows the results for the satisfied/neutral category compared to the dissatisfied category. The results for the dissatisfied category and for the neither satisfied nor dissatisfied (neutral/dissatisfied) category relative to the satisfied category can be found in Table A.3 in the appendix. We find that most variables, except for education and going out for meals, did not violate the parallel lines assumption; therefore, they had the same odds ratio in both tables, irrespective of which two categories of life satisfaction were combined.

The first model investigates the socio-demographic characteristics of the respondent. According to Model 1, living in the central region (compared to the eastern region) and in a rural area (compared to an urban area) increased the odds of being satisfied/neutral with life (compared to being dissatisfied) for ever-married women. Similarly, having at least some primary education increased the odds of being satisfied/neutral when other variables in the model were held constant. More specifically, compared to women with no education, women with incomplete or complete secondary education were more than twice as likely to be satisfied/neutral rather than dissatisfied compared to women with no education, while women with higher education were 4.6 times more likely to be satisfied/neutral than women with no education. Speaking a mother tongue other than Turkish decreased the odds of being satisfied/neutral with life. The other three variables – i.e., the respondent's age and number of children – were found to be insignificant.

Model 2 explores additional variables related to women's material resources. With the inclusion of new variables, the respondent's age becomes a significant determinant. This could be the result of people valuing material resources differently at different ages. According to this model, women who had their own money to spend had 1.330 higher odds of being satisfied/neutral relative to being dissatisfied than women who did not have their own money to spend. Furthermore, the model shows that women who had their own or jointly owned land or a car are more likely to be satisfied/neutral with their lives than women who did not own such assets. As expected, the inclusion of material resources in Model 2 slightly decreased the importance of educational attainment.

Model 3 includes variables related to social resources, in addition to variables related to socio-demographic characteristics and material resources. The results show that whether women had their own money to spend and whether they were living in the central region became insignificant when the new variables were included, and the odds related to educational attainment were further decreased. The odds of being satisfied/neutral relative to being dissatisfied increased for women who reported going out for meals, going to the cinema/theatre or organising home meetings.

Model 4 investigates how women's empowerment through decision-making affected their life satisfaction levels. Surprisingly, according to this model, the odds of being satisfied/neutral relative to being dissatisfied were higher for women who made decisions on contraceptive use and their health jointly with their partner than they were for women who made these decisions themselves.

In Model 5, none of the characteristics of the women's last partner were found to be significant. This was because the partner's education status (0.86) and mother tongue (0.58) were highly correlated with the women's respective characteristics.

Finally, we run a parsimonious model with the variables that were shown to be significant in the previous models. The results of the parsimonious model indicate that among ever-married women, living in a rural area, having higher than secondary education, having their own money to spend, socialising and having decision-making power increased the odds of being satisfied/neutral; while being older decreased the odds of being satisfied/neutral.

4.2.2 Never-married women

In this section, we discuss the determinants of being satisfied/neutral compared to being dissatisfied for never-married women. Slightly more than 25% of the sample consisted of never-married women. It is important to note that 90% of these never-married women were younger than 30 years old. Three different models are used to investigate the association between life satisfaction and other characteristics among never-married women in Turkey. Table 2 shows the results for the combined satisfied/neutral category relative to the dissatisfied category. The results for the dissatisfied category and for the neither satisfied nor dissatisfied (neutral/dissatisfied) category relative to the satisfied category can be found in Table A.4 in the appendix.

In Model 1, which includes the socio-demographic characteristics of never-married women, living in the central or the northern region and having higher than secondary education were found to be significant. Living in the central or northern region rather than in the eastern region increased the odds of being satisfied/neutral by a factor of 1.4, while having secondary or higher education compared to having no education increased the odds of being satisfied/neutral by 2.47.

In Model 2, the variables related to material resources are included in addition to the socio-demographic characteristics. In line with the results for ever-married women, the model shows that having their own money to spend increased the odds of being satisfied/neutral among never-married women (by a factor of 1.41).

In Model 3, the variables associated with social resources are added to the model. Women who go out for meals were more likely to be satisfied or neutral than dissatisfied. The internet use variable was not significant in this model. This could be because it was not a distinctive variable, since only 14% of the never-married women in the sample did not use internet.

Finally, the results of the parsimonious model show that living in the central region, having their own money to spend and socialising increased the odds of being satisfied/neutral among never-married women in Turkey.

In this study, we present as a measure of fit the Pseudo R-squared statistics (McFadden's R^2 at the end of Tables 1 and 2). Pseudo R-squared measures are not same as the R-squared measure, and they should be interpreted with caution. However, in both tables, they show that adding additional variables improved the model fit. A considerable number of empirical studies have observed that people's socio-demographic factors and living conditions explain only between 5% and 10% of the total variation in individual happiness scores, with some studies reporting even lower figures (Argyle 1999; Diener 1994; Graham 2004).

5 Discussion and Conclusion

The aim of this article was to shed light on the relationship between women's life satisfaction and women's empowerment in Turkey. To investigate this relationship, we used data from the most recent DHS survey conducted in Turkey, which collected information on women's life satisfaction for the first time. By combining the liveability approach – which emphasises the role of external resources, such as objective living conditions, and people's own evaluations of their quality of life – and the women's empowerment model, we investigated the factors influencing the life satisfaction levels of women living in Turkey.

This is the first study that has explored the relationship between life satisfaction and the detailed demographic and socio-economic information on women in Turkey. As we mentioned above, we conducted our analysis separately for the ever-married and the never-married women since they tended to be from different birth cohorts and to have different socio-economic and demographic profiles. On average, the never-married women were younger and better educated than the ever-married women. Surprisingly, however, the factors associated with life satisfaction were found the similar for the two groups.

Our investigation showed that for ever-married women who were living in a rural area, having education, material and social resources, and decision-making power were associated with life satisfaction. Our results also indicated that for never-married women who were living in the central or the northern region, having material resources (such as their own money to spend) and social resources increased their levels of life satisfaction. These findings suggest that factors associated with empowerment increased the life satisfaction levels of women in both groups. Furthermore, this study produced results that corroborate the findings of a large number of previous studies in this field (e.g., Ali and ul Haq 2006; Booth and Van Ours 2009; Böhnke 2005; Çakıroğlu-Çevik 2016; Mencarini and Sironi 2012; Mitsuyama and Shimizutani 2019).

The central region, which includes Ankara, the capital of Turkey, had the second-highest proportion of highly educated women (18%), while the western region had the highest proportion (20%). The central region also had the second-highest proportion of women in the richest wealth quintile (28%). However, women living in the central region reported the highest levels of life satisfaction. This finding is contrary to those of previous surveys on happiness (TurkStat 2019b), which showed that women in Ankara tended to report lower happiness levels than the average levels for Turkey.

It is worth noting that the educational system in Turkey has been altered several times in recent decades. Thus, women in different cohorts were exposed to different number of mandatory years of compulsory schooling. Regardless of these changes, both our descriptive and multivariate analyses showed that education was positively associated with women's life satisfaction. A potential explanation for this observed relationship is that in addition to being correlated with economic well-being and mental and physical health, education increases women's agency and capability to control their own lives. These findings are in line with those of Mitsuyama and Shimizutani (2019), who found that higher levels of education were positively associated with self-reports of life satisfaction among women living in Japan. Therefore, education can be considered a key resource for women's empowerment.

In addition, we found that two factors were especially important for the economic well-being and life satisfaction of the surveyed women: 1. having their own money to spend; and 2. having assets. Without having control over money and assets, a woman can easily become a victim of economic violence or dependent on her partner, her family or others. In such cases, women will also be less likely to have the means to get out of unsatisfactory relationships, which will ultimately decrease their levels of life satisfaction. Unfortunately, very few women in Turkey have their own money to spend or their own assets. Indeed, the majority of married women living in Turkey share their finances with other family members. Additionally, in most Turkish households, the husband controls the family's money, and decides on how it is spent (Çakıroğlu-Çevik 2016; TAYA 2011). We also found that the never-married women in the survey, who tended to be younger than the married women, were even less likely to have their own assets.

When we looked at the impact of women's employment status on life satisfaction, we found that this factor was not significant. This may be because among the surveyed women, accepting a more traditional gender role and household wealth status was a more important factor in their life satisfaction than their own economic status (Clark 1997; Mitsuyama and Shimizutani 2019; Torosyan and Pignatti 2020).

Among the other factors that were shown to be associated with life satisfaction for both ever-married and never-married women were factors related to women's social resources, including going out for meals, organising home meetings and going to the cinema/theatre. Going out for meals and going to the cinema/theatre may be indirect indicators of household economic status, whereas organising home meetings is more likely to be related to socialising with other women. This is because in Turkey, women tend to socialise in their homes. These home meetings –

which are also known as reception days, *gün* meetings, gold days or money days (Tuncer 2018) – can also provide an income for the women who are organising them. At these reception days, which are held at regular intervals, the women who are guests generally give a certain amount of money or a piece of gold to the host. Thus, many Turkish women strengthen their social capital in the domestic sphere through these *gün* meetings. Moreover, as the findings of Çakıroğlu-Çevik (2016) indicate, Turkish women generally enjoy socialising with their neighbours, friends and relatives, which, in turn, contributes to their tendency to have more agency and life satisfaction than men.

Furthermore, our regression analyses showed that decision-making had a positive effect on women's life satisfaction. This result is also in line with the findings of Ali and ul Haq (2006), who found that having autonomy increased women's life satisfaction.

The demographic, social and economic characteristics of the different ethnic groups in Turkey vary considerably. When asked to identify their mother tongue, 76.4% of the women in our sample said Turkish, 18.3% cited Kurdish, 2.7% said Arabic and 2.6% named another language. The respondents whose mother tongue was not Turkish reported lower levels of life satisfaction. A possible explanation for this finding is that factors related to women's empowerment are highly associated with ethnicity. For example, 29.4% of women who cited Turkish as their mother tongue had their own money to spend, compared to 11.1% and 18.1% of women who said their mother tongue was Kurdish or Arabic, respectively. In terms of socialising, women whose mother tongue was not Turkish were also less likely to say that they go out for meals, organise home meetings or go to the cinema/theatre.

A major limitation of this study was our use of cross-sectional data, which did not allow us to test reverse causality between life satisfaction and the demographic and socio-economic characteristics of the women in the sample. Unfortunately, since longitudinal life satisfaction information was not provided in the TDHS, we were unable to set up an ideal causal model to control for this type of endogeneity.

Acknowledgements

The research is funded by the TÜBİTAK-1007 Support Program for Research and Development Projects of Public Institutions Project Number:118G001. The research project is funded by the European Research Council (EmpoweredLifeYears, H2020 741105).

References

Akay, E. Ç. and B. Timur 2017. Analysis of the factors affecting the happiness of women and men using the generalized ordered logit model. *Sosyal Bilimler Araştırma Dergisi* 6(3): 88–105. https://doi.org/10.17130/ijmeb.803957

Akın, H. B. and E. Şentürk 2012. Analysing levels of happiness of individuals with ordinal logistic analysis. *Öneri* 37(10): 183–193.

Ali, S. M. and R. ul Haq 2006. Women's autonomy and happiness: The case of Pakistan. *The Pakistan Development Review* 45(1): 121–136.

Andrews, F. M. and S. B. Withey 2005. Developing measures of perceived life quality: Results from several national surveys. In *Citation classics from social indicators research*, ed A. Michalos, 75–100. Netherlands: Springer.

Argyle, M. 1999. Causes and correlates of happiness. In *Well-being: The foundations of hedonic psychology*, eds D. Kahneman, E. Diener and N. Schwarz, 353–373. New York: Russel Sage Foundation.

Arrosa, M. L. and N. Gandelman 2016. Happiness decomposition: Female optimism. *Journal of Happiness Studies* 17(2): 731–756. https://doi.org/10.1007/s10902-015-9618-8

Arslan, H. 2020. Mutluluk ve mutluluk eşitsizliği. içinde Türkiye'de yoksulluk ve eşitsizlik nedenler, süreçler, çözümler, ed S. Ayata, 115–125. Ankara: Sosyal Demokrasi Derneği Yayınları.

Audette, A. P., S. Lam, H. O'Connor and B. Radcliff 2019. (E)Quality of life: A cross-national analysis of the effect of gender equality on life satisfaction. *Journal of Happiness Studies* 20(7): 2173–2188. https://doi.org/10.1007/s10902-018-0042-8

Arrondo, R., A. Cárcaba and E. González 2020. Drivers of subjective well-being in Spain: Are there gender differences? *Applied Research in Quality of Life*. https://doi.org/10.1007/s11482-020-09862-x

Ball, R. and K. Chernova 2008. Absolute income, relative income, and happiness. *Social Indicators Research* 88(3): 497–529. https://doi.org/10.1007/s11205-007-9217-0

Batliwala, S. 1994. The meaning of women's empowerment: New concepts from action. In *Population policies reconsidered: Health, empowerment and rights*, eds G. Sen, A. Germain and L. C. Chen, 127–138. Boston, MA: Harvard Center for Population and Development Studies.

Bjørnskov, C., A. Dreher and J. A. Fischer 2007. On gender inequality and life satisfaction: Does discrimination matter? SSE/EFI Working Paper Series in Economics and Finance, no. 657. Stockholm: The Economic Research Institute (EFI), Stockholm School of Economics. http://hdl.handle.net/10419/56302

Blanchflower, D. G. and A. J. Oswald 2004. Well-being over time in Britain and the USA. *Journal of Public Economics* 88(7): 1359–1386. https://doi.org/10.1016/S0047-2727(02)00168-8

Boncompte, J. G. and R. D. Paredes 2020. Human capital endowments and gender differences in subjective well-being in Chile. *Journal of Happiness Studies* 21(1): 241–269. https://doi.org/10.1007/s10902-019-00085-y

Booth, A. L. and J. C. Van Ours 2009. Hours of work and gender identity: Does part-time work make the family happier? *Economica* 76(301): 176–196. https://doi.org/10.1111/j.1468-0335.2007.00670.x.

Böhnke, P. 2005. *First European quality of life survey: Life satisfaction, happiness and sense of belonging*. Luxembourg: Office for Official Publications in the European Communities.

Campbell, A., P. E. Converse and W. L. Rodgers 1976. *The quality of American life: Perceptions, evaluations, and satisfactions*. New York: Russell Sage Foundation.

Caner, A. 2014. Happiness, comparison effects, and expectations in Turkey. *Journal of Happiness Studies* 16(5): 1323–1345. https://doi.org/10.1007/s10902-014-9562-z

Caner, A. 2016. Happiness and life satisfaction in Turkey in recent years. *Social Indicators Research* 127(1): 361–399. https://doi.org/10.1007/s11205-015-0948-z

Clark, A. E. 1997. Job satisfaction and gender: Why are women so happy at work? *Labour Economics* 4(4): 341–372. https://doi.org/10.1016/S0927-5371(97)00010-9

Clemente, F. and W. Sauer 1976. Life satisfaction in the United States. *Social Forces* 54(3): 621–631. https://doi.org/10.1093/sf/54.3.621

Çakıroğlu-Çevik, A. 2016. Yaşam kalitesi ve mutluluk ilişkisi çerçevesinde Türkiye'de kadının durumu. *Kadin/Woman 2000* 17(2): 133–160.

Diener, E. 1994. Assessing subjective well-being: Progress and opportunities. *Social Indicators Research* 31(2): 103–157. https://doi.org/10.1007/bf01207052

Dolan, P., T. Peasgood and M. White 2008. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology* 29(1): 94–122. https://doi.org/10.1016/j.joep.2007.09.001

Dumludag, D. 2013. Life satisfaction and income comparison effects in Turkey. *Social Indicators Research* 114(3): 1199–1210. https://doi.org/10.1007/s11205-012-0197-3

Ekici, and S. Koydemir 2014. Social capital, government and democracy satisfaction, and happiness in Turkey: A comparison of surveys in 1999 and 2008. *Social Indicators Research* 118(3): 1031–1053. https://doi.org/10.1007/s11205-013-0464-y

Eren, K. A. and A. A. Aşıcı 2017. The determinants of happiness in Turkey: Evidence from city-level data. *Journal of Happiness Studies* 18(3): 647–669. https://doi.org/10.1007/s10902-016-9746-9

Ermiş-Mert, A. 2020. A multivariate investigation of overall happiness, job satisfaction and income satisfaction of women and men in Turkey. *Sosyoekonomi* 28(44): 11–26. https://doi.org/10.17233/sosyoekonomi.2020.02.01

Fujita, F., E. Diener and E. Sandvik 1991. Gender differences in negative affect and well-being: the case for emotional intensity. *Journal of Personality and Social Psychology* 61(3): 427–434. https://doi.org/10.1037//0022-3514.61.3.427

Graham, C. 2004. Globalization, poverty, inequality, and insecurity: Some insights from the economics of happiness. Paper presented at the UNU/WIDER Conference on The Impact of Globalization on the World's Poor, Helsinki Finland, October 2004. https://www.wider.unu.edu/sites/default/files/rp2005-33.pdf, accessed 16 July 2021.

Graham, C. 2005. The economics of happiness: Insights on globalization from a novel approach. *World Economics 6*(3): 41–55.

Graham, C. and S. Chattopadhyay 2013. Gender and well-being around the world. *International Journal of Happiness and Development* 1(2): 212–232. https://doi.org/10.1504/IJHD.2013.055648

Güneş-Ayata, A. and G. Doğangün 2017. Gender politics of the AKP: Restoration of a religio-conservative gender climate. *Journal of Balkan and Near Eastern Studies* 19(6): 610–627. https://doi.org/10.1080/19448953.2017.1328887

Haller, M. and M. Hadler 2006. How social relations and structures can produce happiness and unhappiness: An international comparative analysis. *Social Indicators Research* 75(2):

169–216. https://doi.org/10.1007/s11205-004-6297-y

Haring, M. J., W. A. Stock and M. A. Okun 1984. A research synthesis of gender and social class as correlates of subjective well-being. *Human Relations* 37(8): 645–657. https://doi.org/10.1177/001872678403700805

Hacettepe University Institute of Population Studies (HUIPS) 2019. *2018 Turkey demographic and health survey*. Ankara: Elma Publishing.

Hülagü, F. 2021. *Police reform in Turkey: Human security, gender and state violence under Erdogan*. Bloomsbury Publishing.

Inglehart, R. 1990. *Culture shift in advanced industrial society*. Princeton University Press.

Inglehart, R. and C. Welzel 2005. *Modernization, cultural change, and democracy: The human development sequence*. Cambridge University Press.

Inglehart, R., R. Foa, C. Peterson and C. Welzel 2008. Development, freedom, and rising happiness: A global perspective (1981–2007). *Perspectives on Psychological Science 3*(4): 264–285. https://doi.org/10.1111/j.1745-6924.2008.00078.x

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin and B. Puranen et al. (eds) 2014. *World values survey: All rounds - country-pooled datafile version*. Madrid: JD Systems Institute. https://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp

Jorm, A. F. and S. M. Ryan 2014. Cross-national and historical differences in subjective well-being. *International Journal of Epidemiology* 43(2): 330–340. https://doi.org/10.1093/ije/dyt188

Kabeer, N. 1999. Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change* 30(3): 435–464. https://doi.org/10.1111/1467-7660.00125

Kahneman, D. and A. B. Krueger 2006. Developments in the measurement of subjective well-being. *The Journal of Economic Perspectives* 20(1): 3–24. https://doi.org/10.1257/089533006776526030

Lalive, R. and A. Stutzer 2004. Approval of equal rights and gender differences in well-being. IZA Discussion Paper 1202. Bonn: Forschungsinstitut zur Zukunft der Arbeit.

Mayungbo, O. A. 2016. Gender differences, self-esteem and life satisfaction. *African Journal for The Psychological Studies of Social Issues* 19(2): 1–11.

Meisenberg, G. and M. A. Woodley 2015. Gender differences in subjective well-being and their relationships with gender equality. *Journal of Happiness Studies* 16(6): 1539–1555. https://doi.org/10.1007/s10902-014-9577-5

Mencarini, L. and M. Sironi 2012. Happiness, housework and gender inequality in Europe. *European Sociological Review* 28(2): 203–219. https://doi.org/10.1093/esr/jcq059

Michalos, A. C. 1980. Satisfaction and happiness. *Social Indicators Research* 8(4): 385–422. https://doi.org/10.1007/bf00461152

Mitsuyama, N. and S. Shimizutani 2019. Male and female happiness in Japan during the 2000s: Trends during era of promotion of active participation by women in society. *The Japanese Economic Review* 70(2): 189–209. https://doi.org/10.1111/jere.12183

Noll, H. H. 2002. Towards a European system of social indicators: Theoretical framework and system architecture. *Social Indicators Research* 58(1–3): 47–87. https://doi.org/10.1023/A:1015775631413

Okun, M. A. and L. K. George 1984. Physician-and self-ratings of health, neuroticism and subjective well-being among men and women. *Personality and Individual Differences* 5(5): 533–539. https://doi.org/10.1016/0191-8869(84)90027-8

Roothman, B., D. K. Kirsten and M. P. Wissing 2003. Gender differences in aspects of psychological well-being. *South African Journal of Psychology* 33(4): 212–218. https://doi.org/10.1177/008124630303300403

Rowlands, J. 1995. Empowerment examined. *Development in Practice* 5(2): 101–107. https://doi.org/10.1080/0961452951000157074

Ruth, L. and J. Napier 2014. Gender inequality and subjective well-being. In *Encyclopedia of quality of life and well-being research*, ed A. C. Michalos, 2446–2451. Netherlands: Springer.

Saris, W. E. 2001. The relationship between income and satisfaction: The effect of measurement error and suppressor variables. *Social Indicators Research* 53(2): 117–136. https://doi.org/10.1023/a:1026785328244

Schyns, P. 1998. Cross national differences in happiness: Economic and cultural factors explored. *Social Indicators Research* 43(1): 3–26. https://doi.org/10.1023/a:1006814424293

Seghieri, C., G. Desantis and M. Tanturri 2006. The richer, the happier? An empirical investigation in selected European countries. *Social Indicators Research* 79(3): 455–476. https://doi.org/10.1007/s11205-005-5394-x

Şengül, S. and K. Lopcu 2020. Does gender affect subjective well-being? The case of Turkey. Topics in middle Eastern and African economies. *Proceedings of Middle East Economic Association* 22(1): 179–191.

Shmotkin, D. 1990. Subjective well-being as a function of age and gender: A multivariate look for differentiated trends. *Social Indicators Research* 23(3): 201–230. https://doi.org/10.1007/BF00293643

Stevenson, B. and J. Wolfers 2009. The paradox of declining female happiness. *American Economic Journal: Economic Policy* 1(2): 190–225. https://doi.org/10.1257/pol.1.2.190

Stutzer, A. and B. S. Frey 2006. What happiness research can tell us about self-control problems and utility misprediction. IZA Discussion Paper 1952. Bonn: Forschungsinstitut zur Zukunft der Arbeit.

Tay, L., V. Ng, L. Kuykendall and E. Diener 2014. Demographic factors and worker well-being: An empirical review using representative data from the United States and across the world. In *The role of demographics in occupational stress and wellbeing,* eds P. L. Perrewé, C. C. Rosen and J. R. B. Halbesleben, 235–283. Emerald Group Publishing Limited.

TAYA 2011. *Türkiye Aile Yapısı Araştırması*. T.C. Aile ve Sosyal Politikalar Bakanlığı Aile ve Toplum Hizmetleri Genel Müdürlüğü. Ankara, Turkey.

Tesch-Römer, C., A. Motel-Klingebiel and M. J. Tomasik 2008. Gender differences in subjective well-being: Comparing societies with respect to gender equality. *Social Indicators Research* 85(2): 329–349. https://doi.org/10.1007/s11205-007-9133-3.

Tiefenbach, T. and F. Kohlbacher 2013. Happiness and life satisfaction in Japan by gender and age. Working Paper 13/2. Tokyo: German Institute for Japanese Studies.

Toksöz, G. and E. Memiş 2018. *İstihdamda toplumsal cinsiyet eşitliği haritalama ve izleme çalışması.* CEİD Publications: 5. https://dspace.ceid.org.tr/xmlui/handle/1/171

Torosyan, K. and N. Pignatti 2020. Employment vs. homestay and the happiness of women in the South Caucasus, IZA Discussion Paper 12888. Bonn: Forschungsinstitut zur Zukunft der Arbeit.

Treas, J., T. van der Lippe and T. O. C. Tai 2011. The happy homemaker? Married women's well-being in cross-national perspective. *Social Forces* 90(1):111–132. https://doi.org/10.1093/sf/90.1.111

Tuncer, S. 2018. *Women and public space in Turkey: Gender, modernity and the urban experience*. I.B. Tauris.

TurkStat 2016. *Time Use Survey of Turkey for 2014–2015*. Turkish Statistical Institute. Available at: https://www.tuik.gov.tr/Kurumsal/Mikro_Veri

TurkStat 2018. *Structure of Earnings Survey.* Turkish Statistical Institute. Available at https://www.tuik.gov.tr/Kurumsal/Mikro_Veri

TurkStat 2019a. *Gender Statistics.* Available at https://data.tuik.gov.tr/Kategori/GetKategori?p=nufus-ve-demografi-109&dil=1, accessed on 10 July 2020.

TurkStat 2019b. *Merkezi Dagitim Sistemi*, *General level of happiness*. Available at https://biruni.tuik.gov.tr/medas/?kn=100&locale=en, accessed on 10 August 2021.

TurkStat 2020. *Life Satisfaction Survey*. Turkish Statistical Institute. Available at https://www.tuik.gov.tr/Kurumsal/Mikro_Veri

UN 2019. *Sustainable Development Goals: Goal 5 Achieve gender equality and empower all women and girls*. Available at https://www.un.org/sustainabledevelopment/gender-equality/, accessed on 25 November 2020.

UN 2015. *Transforming our world: The 2030 agenda for sustainable development* (A/RES/70/1). New York, USA: United Nations.

UNDP 2020. *Gender Inequality Index 2018*. Available at http://hdr.undp.org/, accessed on 25 November 2020.

Van Praag, B. M. and A. Ferrer-i-Carbonell 2008. A multidimensional approach to subjective poverty. In *Quantitative approaches to multidimensional poverty measurement*, eds N. Kakwani and J. Silber, 135–154. London: Palgrave Macmillan.

Veenhoven, R. 1996. The study of life satisfaction. In *A comparative study of satisfaction with life in Europe*, eds W. E. Saris, R. Veenhoven, A. C. Scherpenzeel and B. Bunting, 11–48. Eötvös University Press.

Veenhoven, R. 2000. The four qualities of life. *Journal of Happiness Studies* 1(1): 1–39. https://doi.org/10.1023/a:1010072010360

Veenhoven, R. 2006. What we know about happiness? *Positiv Psykologi* 16(60): 14–50.

Vieira Lima, S. 2011. A cross-country investigation of the determinants of the happiness gender gap. Available at https://www.scribd.com/doc/294258071/A-Cross-Country-Investigation-of-the-Determinants-of-the-Happiness-Gender-Gap, accessed on 16 July 2021.

Williams, R. 2016. Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology* 40(1): 7–20. https://doi.org/10.1080/0022250X.2015.1112384

Williams, R. 2005. *gologit2: Generalized logistic regression models for ordinal dependent variables*. Proceedings of 2005 North American Stata Users Group meeting. 11–12 July 2005. Massachusetts, USA. Available at https://www.stata.com/meeting/4nasug/Williams_NASUG.pdf, accessed on 8 February 2021.

Wood, W., N. Rhodes and M. Whelan 1989. Sex differences in positive well-being: A consideration of emotional style and marital status. *Psychological Bulletin* 106(2): 249.

World Economic Forum 2020. *Global Gender Gap Index Report 2020*. Available at: http://www3.weforum.org/docs/WEF_GGGR_2020.pdf, accessed 25 November 2020.

Yetim, Ü. 1993. Life satisfaction: A study based on the organization of personal projects. *Social Indicators Research* 29(3): 277–289. https://doi.org/10.1007/bf01079516

Zuckerman, M., C. Li and E. F. Diener 2017. Societal conditions and the gender difference in well-being: Testing a three-stage model. *Personality and Social Psychology Bulletin* 43(3): 329–336. https://doi.org/10.1177/0146167216684133

Zweig, J. S. 2015. Are women happier than men? Evidence from the Gallup World Poll. *Journal of Happiness Studies* 16(2): 515–541. https://doi.org/10.1007/s10902-014-9521-8

Appendix

Table A.1:
Life satisfaction of never-married and ever-married women

Life satisfaction	Never married	Ever married	Total
Dissatisfied			
Frequency	340	1188	1528
Column percentage	18.26	21.66	20.8
Neither satisfied nor dissatisfied			
Frequency	505	1344	1849
Column percentage	27.12	24.51	25.17
Satisfied			
Frequency	1017	2952	3969
Column percentage	54.62	53.83	54.03
Total			
Frequency	1862	5484	7346
Column percentage	100	100	100

Table A.2:
Characteristics of never-married and ever-married women

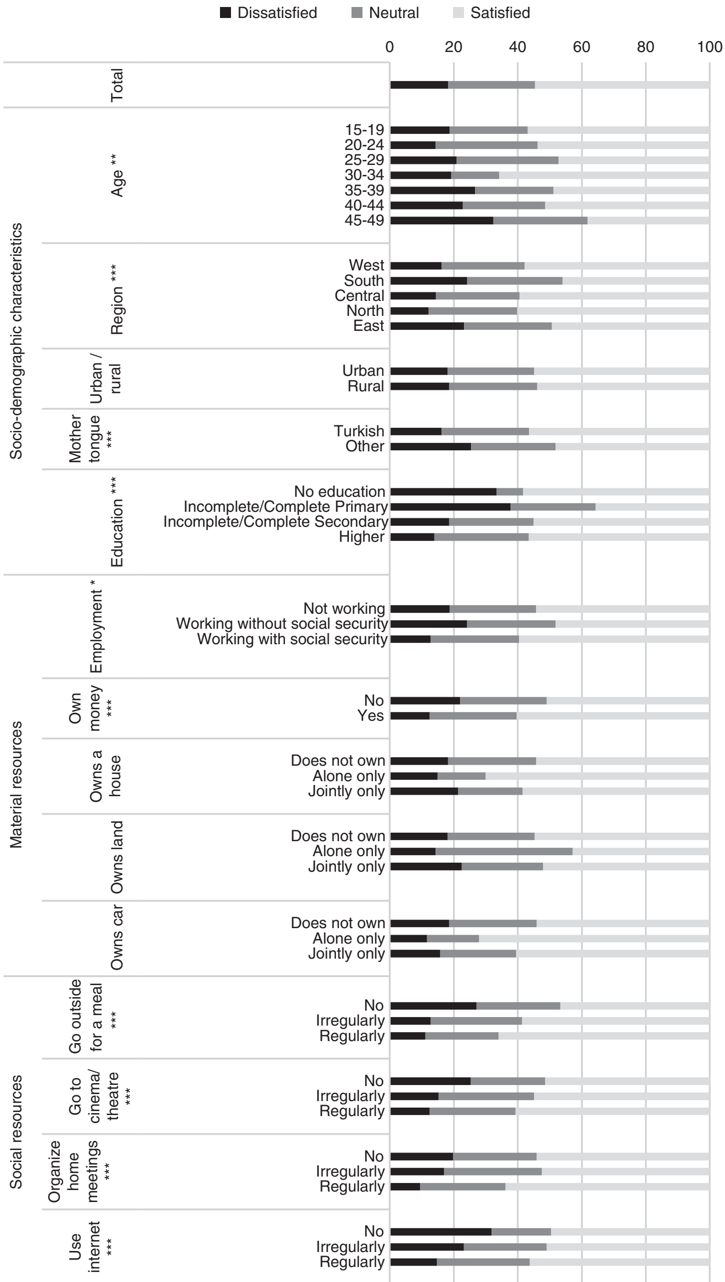
Characteristics	Never married	Ever married	Total
Socio-demographic characteristics			
Age in five-year groups			
15–19	50.0	1.2	15.8
20–24	27.7	8.2	14.1
25–29	12.2	14.9	14.1
30–34	4.1	19.0	14.5
35–39	2.4	20.5	15.0
40–44	1.7	19.2	14.0
45–49	1.9	17.0	12.5
Region			
West	42.0	44.3	43.6
South	12.0	12.6	12.4
Central	20.0	21.1	20.7
North	5.8	5.3	5.5
East	20.2	16.7	17.8
Type of place of residence			
Urban	78.2	78.2	78.2
Rural	21.9	21.8	21.8
Educational attainment			
No education	1.7	9.2	6.9
Incomplete/Complete primary	4.4	39.2	28.7
Incomplete/Complete secondary	56.9	34.8	41.4
Higher	37.1	16.9	22.9
Mother tongue			
Turkish	81.6	78.5	79.5
Other	18.4	21.5	20.6
Material resources			
Employment			
Not working	75.2	70.1	71.7
Working without social security	7.5	12.3	10.9
Working with social security	17.3	17.5	17.5
Has own money to spend			
No	56.3	72.5	67.6
Yes	43.7	27.5	32.4
Owns a house alone or jointly			
Does not own	94.1	76.5	81.8
Alone only	1.3	7.7	5.8
Jointly only	4.7	15.7	12.4
Owns land alone or jointly			
Does not own	96.3	89.1	91.2
Alone only	0.3	4.1	3.0
Jointly only	3.4	6.8	5.8

Continued

Table A.2:
Continued

Characteristics	Never married	Ever married	Total
Owns a car alone or jointly			
Does not own	95.8	82.0	86.2
Alone only	2.1	6.5	5.2
Jointly only	2.1	11.4	8.7
Social resources			
Goes out for meals			
No	35.6	33.0	33.8
Irregularly	54.7	51.7	52.6
Regularly	9.7	15.3	13.6
Organises home meetings			
No	70.5	58.6	62.1
Irregularly	20.2	22.3	21.7
Regularly	9.4	19.1	16.2
Uses the internet			
No	12.2	30.9	25.2
Irregularly	12.0	18.1	16.3
Regularly	75.9	51.0	58.5
Goes to the cinema/theatre			
No	28.8	60.0	50.7
Irregularly	52.2	31.4	37.7
Regularly	19.0	8.5	11.7
Decision-making			
Decision-maker for contraception use			
Mainly respondent	11.3	22.4	22.2
Mainly husband, partner	NA	2.1	2.0
Joint decision	88.7	74.0	74.2
Other	0	1.6	1.6
Person who usually decides on the respondent's health care			
Herself	NA	37.8	37.8
With partner	NA	57.6	57.6
Partner or someone else	NA	4.5	4.5
Couple power dynamics			
Partner's educational attainment			
No education	NA	1.9	1.9
Incomplete/Complete primary	NA	33.7	33.7
Incomplete/Complete secondary	NA	44.4	44.4
Higher than secondary	NA	20	20
Partner has social security			
No	NA	20.7	20.7
Yes	NA	79.3	79.3
Partner's mother tongue			
Turkish	NA	78.2	78.2
Other	NA	21.8	21.8
Total	100.0	100.0	100.0

Figure A.1:
Correlation between life satisfaction and never-married women's characteristics



Note: Chi-squared test for independence: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A.3:
Gologit model results (odds ratio) for ever-married women

Neither satisfied nor dissatisfied	Model 1	Model 2	Model 3	Model 4	Model 5	Parsimonious
Socio-demographic characteristics						
Respondent's current age	0.991	0.986**	0.983***	0.981*	0.984	0.981*
Region (East)						
West	1.174	1.154	1.010	1.164	1.102	1.10
South	0.975	0.954	0.918	1.177	1.18	1.112
Central	1.299**	1.278*	1.187	1.395	1.399	1.333
North	1.190	1.163	1.016	1.28	1.123	1.219
Type of place of residence (Urban)						
Rural	1.199**	1.216**	1.435***	1.329*	1.329	1.367*
Number of children (4 or more)						
0	1.101	1.083	0.905	1.597	1.794	
1	1.106	1.089	0.911	0.753	0.808	
2	1.017	1.001	0.838	0.699	0.716	
3	1.048	1.050	0.949	0.872	0.951	
Mother tongue (Turkish)						
Other	0.740***	0.741**	0.783**	0.875	0.862	0.909
Educational attainment (No education)						
Incomplete/Complete primary	0.985	0.972	0.861	1.543	1.568	1.477
Incomplete/Complete secondary	1.268	1.174	0.850	1.441	1.405	1.312
Higher than secondary	2.323***	1.879***	1.170	1.804	1.385	1.862*
Material resources						
Employment (Working with social security)						
Not working		0.961	0.985	0.811	0.862	
Working without social security		0.829	0.866	0.830	0.857	
Has own money to spend (No)						
Yes		1.330***	1.142	1.310	1.343	1.368*
Owns a house alone or jointly (Does not own)						
Alone only		1.226	1.17	1.972**	1.881*	1.951**
Jointly only		1.145	1.088	1.099	1.026	1.092
Owns land alone or jointly (Does not own)						
Alone only		1.139	1.133	1.059	0.942	1.04
Jointly only		1.384**	1.364*	1.099	1.075	1.074
Owns a car alone or jointly (Does not own)						
Alone only		1.067	0.924	0.876	0.876	0.925
Jointly only		1.102	1.028	0.858	0.803	0.875

Continued

Table A.3: Continued

Neither satisfied nor dissatisfied	Model 1	Model 2	Model 3	Model 4	Model 5	Parsimonious
Social resources						
Goes out for meals (No)						
Irregularly			1.706***	2.041***	1.882***	2.018***
Regularly			2.584***	3.136***	2.809***	3.041***
Organises home meetings (No)						
Irregularly			1.217*	1.104	1.071	1.082
Regularly			1.674***	1.631**	1.607**	1.545**
Use internet (No)						
Irregularly			1.103	0.993	0.916	
Regularly			1.065	1.017	0.952	
Goes to the cinema/theatre (No)						
Irregularly			1.238*	1.464*	1.609**	1.486**
Regularly			1.361	1.264	1.452	1.381
Decision-making						
Decision-maker for contraception use (Mainly respondent)						
Mainly husband, partner				1.649	1.658	1.662
Joint decision				1.780***	1.781***	1.773***
Other				1.001	1.058	1.014
Person who usually decides on the respondent's health care (Herself)						
With partner				1.516***	1.390**	1.512***
Partner or someone else				0.820	0.733	0.837
Couple power dynamics						
Partner's educational attainment (No education)						
Incomplete/Complete primary					1.501	
Incomplete/Complete secondary					1.662	
Higher than secondary					2.214	
Partner has social security (No)						
Yes					0.811	
Partner's mother tongue (Turkish)						
Other					1.033	
Observations	5484	5480	5480	1912	1755	1912
Pseudo R-squared	0.029	0.0348	0.0594	0.0921	0.0860	0.0889

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A.4:
Gologit model results (odds ratio) for never-married women

Neither satisfied nor dissatisfied	Model 1	Model 2	Model 3	Model 4	Model 5	Parsimonious
Socio-demographic characteristics						
Respondent's current age	0.991	0.986**	0.983***	0.981*	0.984	0.981*
Region (East)						
West	1.174	1.154	1.010	1.164	1.102	1.10
South	0.975	0.954	0.918	1.177	1.18	1.112
Central	1.299**	1.278*	1.187	1.395	1.399	1.333
North	1.190	1.163	1.016	1.28	1.123	1.219
Type of place of residence (Urban)						
Rural	1.199**	1.216**	1.435***	1.329*	1.329	1.367*
Number of children (4 or more)						
0	1.101	1.083	0.905	1.597	1.794	
1	1.106	1.089	0.911	0.753	0.808	
2	1.017	1.001	0.838	0.699	0.716	
3	1.048	1.050	0.949	0.872	0.951	
Mother tongue (Turkish)						
Other	0.740***	0.741**	0.783**	0.875	0.862	0.909
Educational attainment (No education)						
Incomplete/Complete primary	0.985	0.972	0.861	1.543	1.568	1.477
Incomplete/Complete secondary	1.268	1.174	0.850	1.441	1.405	1.312
Higher than secondary	2.323***	1.879***	1.170	1.804	1.385	1.862*
Material resources						
Employment (Working with social security)						
Not working		0.961	0.985	0.811	0.862	
Working without social security		0.829	0.866	0.830	0.857	
Has own money to spend (No)						
Yes		1.330***	1.142	1.310	1.343	1.368*
Owns a house alone or jointly (Does not own)						
Alone only		1.226	1.17	1.972**	1.881*	1.951**
Jointly only		1.145	1.088	1.099	1.026	1.092
Owns land alone or jointly (Does not own)						
Alone only		1.139	1.133	1.059	0.942	1.04
Jointly only		1.384**	1.364*	1.099	1.075	1.074
Owns a car alone or jointly (Does not own)						
Alone only		1.067	0.924	0.876	0.876	0.925
Jointly only		1.102	1.028	0.858	0.803	0.875

Continued

Table A.4: Continued

Neither satisfied nor dissatisfied	Model 1	Model 2	Model 3	Model 4	Model 5	Parsimonious
Social resources						
Goes out for meals (No)						
Irregularly			1.706***	2.041***	1.882***	2.018***
Regularly			2.584***	3.136***	2.809***	3.041***
Organises home meetings (No)						
Irregularly			1.217*	1.104	1.071	1.082
Regularly			1.674***	1.631**	1.607**	1.545**
Uses the internet (No)						
Irregularly			1.103	0.993	0.916	
Regularly			1.065	1.017	0.952	
Goes to the cinema/theatre (No)						
Irregularly			1.238*	1.464*	1.609**	1.486**
Regularly			1.361	1.264	1.452	1.381
Decision-making						
Decision-maker for contraception use (Mainly respondent)						
Mainly husband, partner				1.649	1.658	1.662
Joint decision				1.780***	1.781***	1.773***
Other				1.001	1.058	1.014
Person who usually decides on the respondent's health care (Herself)						
With partner				1.516***	1.390**	1.512***
Partner or someone else				0.820	0.733	0.837
Couple power dynamics						
Partner's educational attainment (No education)						
Incomplete/Complete primary					1.501	
Incomplete/Complete secondary					1.662	
Higher than secondary					2.214	
Partner has social security (No)						
Yes					0.811	
Partner's mother tongue (Turkish)						
Other					1.033	
Observations	5484	5480	5480	1912	1755	1912
Pseudo R-squared	0.029	0.0348	0.0594	0.0921	0.0860	0.0889

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 293–318

The age U-shape in Europe: The protective role of partnership

***Andrew E. Clark*[1], *Hippolyte d'Albis*[1,*] *and Angela Greulich*[2]**

Abstract

In this study, we ask whether the U-shaped relationship between life satisfaction and age is flatter for individuals who are partnered. An analysis of cross-sectional EU-SILC data indicates that the decline in life satisfaction from the teens to the fifties is almost four times larger for non-partnered than for partnered individuals, whose life satisfaction essentially follows a slight downward trajectory with age. However, the same analysis applied to three panel datasets (BHPS, SOEP and HILDA) reveals a U-shape for both groups, albeit somewhat flatter for the partnered than for the non-partnered individuals. We suggest that the difference between the cross-sectional and the panel results reflects compositional effects: i.e., there is a significant shift of the relatively dissatisfied out of marriage in mid-life. These compositional effects tend to flatten the U-shape in age for the partnered individuals in the cross-sectional data.

Keywords: life satisfaction; life cycle; partnership; marriage

1 Introduction

In the literature on the relationship between subjective well-being and age, one of the recurrent themes has been the shape of this relationship. While some contributions on this this topic have suggested that the relationship is essentially flat (for example, Kassenböhmer and Haisken-DeNew 2012), many have concluded that it is U-shaped (recent examples using data from a variety of different sources can be found in Cheng et al. 2017; Blanchflower 2020a and 2020b). This literature is now voluminous, and useful summaries appear in three chapters on the subject

[1]Paris School of Economics – CNRS, Paris, France

[2]Sciences Po – OSC, Paris, France and Institut Universitaire de France, Paris, France

[*]Correspondence to: Hippolyte d'Albis, hdalbis@psemail.eu

DOI: 10.1553/populationyearbook2021.res3.1

in the recent book edited by Mariano Rojas (Blanchflower and Oswald 2019; Clark 2019; Helliwell et al. 2019).

While there is probably majority agreement in the literature that the relationship is U-shaped, there is much more debate among scholars about why this is the case, and whether this pattern holds equally for all countries and all groups in society. Here, we contribute to the second of these discussions, and ask whether the U-shaped relationship between life satisfaction and age is flatter for individuals who are partnered. Partnership is obviously a relevant candidate as a variable in explanations of well-being, as being in a partnership requires substantial investments of time and resources and high levels of emotional involvement for the partners (Gove et al. 1990). Moreover, like other characteristics of family life that are important for well-being, being in a partnership is associated with greater financial security and less loneliness. Earlier work has found significant differences in well-being according to partnership status, although the size of these differences has been shown to vary across countries (Verbakel 2012). Moreover, an analysis of World Values Survey/European Values Survey data (Mikucka 2016) has indicated that the size of these differences has declined among men since the 1980s.

We are not the first to examine the question of whether being in a partnership flattens the U-shaped relationship between subjective well-being and age. Helliwell et al. (2019) introduced an interaction between marriage/cohabitation and age in subjective well-being regressions drawn from both the US Gallup Healthways data and the UK Annual Population Survey. They found a substantially flatter U-shaped relationship between well-being and age for partnered individuals based on their analysis of both of these cross-sectional datasets, which was only partly explained by control variables (sex, education, labour force status, log income and the number of children). A flatter U-shape for married people was also found in an analysis of the Gallup World Poll data by Jebb et al. (2020). The authors imposed a quadratic relationship between age and the Cantril ladder well-being question, though they argued that the difference between the two curves was not large.

In the current analysis, we use a very large-scale harmonised dataset covering 32 European countries, the EU-SILC, to first show that the U-shape holds in general in Europe. This is a very standard observation in the empirical literature. Moreover, as in Helliwell et al. (2019), we demonstrate that this U-shape is far flatter for partnered individuals, and that the age U-shape is mostly concentrated among those individuals who are not partnered. This pattern is not found for all of the subjective evaluations in the EU-SILC data. The protective role of being in a partnership is found to be particularly important for people's overall life satisfaction, their satisfaction with their financial situation and their satisfaction with their personal relationships, but not at all for their job satisfaction or their satisfaction with time use. Thus, being partnered seems to protect people against the typical decline in well-being associated with middle age in at least certain domains of subjective well-being.

This first result is, however, less clear in the panel data. Fixed-effect analyses using the BHPS (UK), the SOEP (Germany) and the HILDA (Australia) reveal a U-shape in age for both partnered and non-partnered individuals, although the U-shape for the partnered individuals was again found to be somewhat flatter. We consider a number of explanations for this gap between the cross-sectional and the panel results. In particular, we do not believe that this discrepancy reflects the different countries in the cross-sectional and panel analyses, or differential mortality (or attrition in general) between the partnered and the non-partnered individuals. However, we do find some evidence of compositional effects that tend to flatten the U-shape in age for the partnered individuals in the cross-sectional data. Specifically, we observe that those individuals who experience a break-up in their forties and fifties are noticeably less satisfied with their lives than those who experience a break-up at younger or older ages. This composition or shift-share effect will mechanically increase the average well-being among those individuals who remain partnered in their forties and fifties, and will tend to produce a less U-shaped age profile.

The remainder of the paper is organised as follows. In Section 2, we describe the European data that we use and the hypotheses that we test, followed by the main cross-sectional results. In Section 3, we present our analysis of panel data in selected countries. Section 4 concludes.

2 Life satisfaction, age and partnership: cross-sectional analysis of EU-SILC data

Our main analysis of subjective well-being by age and gender is based on data from the EU Statistics on Income and Living Conditions (EU-SILC) survey conducted by Eurostat, which provides harmonised information on hundreds of thousands of respondents in more than 30 European countries. This annual survey is the reference source for comparative statistics on income distribution and social inclusion in the European Union (EU). The EU-SILC provides (micro-level) information on both households and individuals on topics such as income and tax, material deprivation, housing conditions, employment, childcare, health and education.

Additional specific topics are covered in *ad hoc* annual modules, which complement the core variables that appear in every EU-SILC wave by providing supplementary information that highlights different aspects of social inclusion. Examples of the topics covered in these modules are the intergenerational transmission of poverty (2005 and 2011), social participation (2006 and 2015), indebtedness (2008) and the intra-household sharing of resources (2010). The EU-SILC 2013 *ad hoc* module provided information on well-being for 350,000 individuals in 32 different European countries.[1]

[1] Austria, Belgium, Bulgaria, Switzerland, Cyprus, the Czech Republic, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Croatia, Hungary, Ireland, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Serbia, Sweden, Slovenia, Slovakia and the United Kingdom.

EU-SILC is a panel dataset with a four-year rotation (so that 25% of the sample are replaced each year). The 2018 EU-SILC wave repeated parts of the 2013 well-being module, although the gap between the two waves was large enough that none of the 2013 respondents remained in the 2018 sample. However, in this analysis, we carry out a cross-sectional analysis of the 2013 data only, as the 2018 module on well-being did not include all of the well-being variables that we will consider below.

Among the aims of the 2013 well-being module was to place greater emphasis on individuals' feelings of well-being, which are known to be highly correlated with levels of social inclusion and labour market integration. The module covered subjects like satisfaction with different life domains, the emotional state of mind, social deprivation, and the option to ask for help if necessary. All questions were posed at the individual level (to household members aged 16 and older).

Survey design can have an impact on individual replies. Thus, the placement of the well-being module in the survey questionnaire, the response scales used and how respondents perceive the questions in different countries can influence the responses. As the order of the questions within the module and the position of the module questionnaire in the survey were mostly standardised across countries, Eurostat considers the contribution of these parameters to the overall context–related bias to be minimal. For further details on assessments of the implementation of the 2013 EU-SILC module on well-being, see Eurostat (2012*a*).

The main dependent variable in our analysis is "overall life satisfaction", with replies on a scale from zero (not at all satisfied) to 10 (completely satisfied).[2] The life satisfaction variable is intended to capture the respondent's broad, reflective appraisal of his/her life as a whole. The term "life" is intended to be understood here as all of the areas of a person's life at a particular point in time (Diener 2006). The variable captures the respondent's opinion/feeling about how s/he is feeling "these days", rather than specifying an explicit longer or shorter time period. Thus, the intent is not to measure the individual's current emotional state (or affect), but for him/her to make a reflective judgement about his/her satisfaction with his/her life overall (Eurostat 2012*b*).[3]

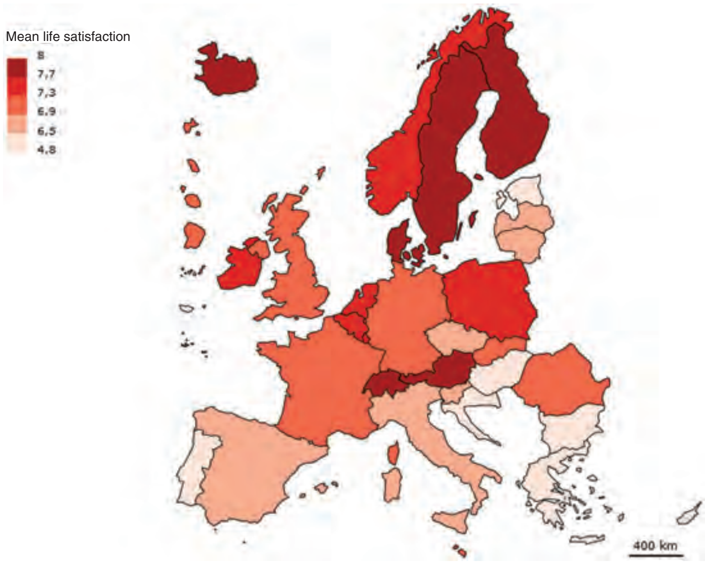
The weighted average overall life satisfaction score in the 2013 EU-SILC, covering all 32 European countries and considering respondents up to age 80, is seven. Table A.1 in the Appendix presents the descriptive statistics of the life satisfaction scores and the percentage of the respondents who are partnered by the different age groups we will analyse.

The average overall life satisfaction score appears in the middle of the ranking of the more specific domain satisfaction variables: satisfaction is lower for "trust in the political system", "trust in the legal system", "trust in others", "satisfaction

2 The validity of life satisfaction as a measure of well-being is discussed in Clark (2016).

3 However, (recent) mood and life satisfaction may be strongly correlated with each other, and be similarly correlated with a number of explanatory variables: see Clark (2016).

Figure 1:
Overall life satisfaction in 32 European countries



Source: EU-SILC 2013.

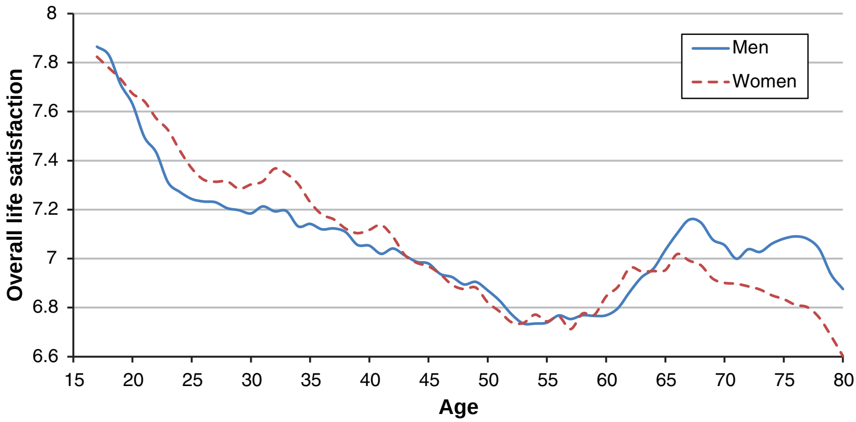
with the financial situation", "trust in the police" and "satisfaction with time use"; satisfaction is higher for "satisfaction with recreational or green areas", "satisfaction with the living environment", "job satisfaction", "satisfaction with commuting times", "meaning of life", "satisfaction with accommodation" and "satisfaction with personal relationships". It should, however, be emphasised that the "overall life satisfaction" score comes from a question in its own right in the SILC questionnaire, and is not calculated as a summary measure of the various domain satisfaction scores.

The cross-country distribution of life satisfaction in our analysis is similar to that found in other surveys (see, for example, the distribution of the answers to the Cantril ladder question in the Gallup World Poll, which is presented in the annual World Happiness Report). Overall life satisfaction is generally shown to be higher in Northern Europe and lower in Southern Europe, as illustrated in Figure 1's heatmap.

The lowest life satisfaction score on the 0–10 scale is found in Bulgaria (4.8), followed by in Serbia (4.9) and Portugal (6.1); and the highest score is found in Switzerland (8.0), followed by in Finland and Denmark (both 7.9).

The relationship of life satisfaction to age and sex is depicted in Figure 2. As throughout the empirical analysis, here we only consider respondents up to age 80. Across all 32 countries in the study sample, women are found to be less satisfied at higher ages and more satisfied at younger ages than men.

Figure 2:
Overall life satisfaction by age and sex



Note: The figures here are weighted averages over 32 European countries. The points refer to three-year (with respect to age) moving averages.
Source: EU-SILC 2013.

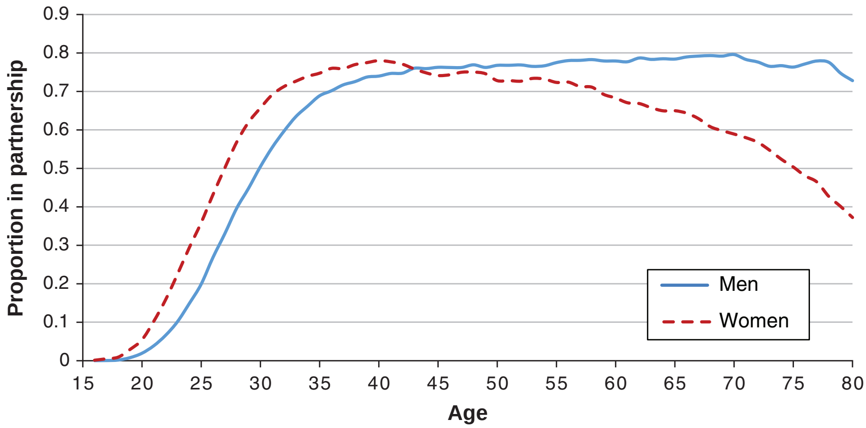
However, we find no systematic gender gap in overall life satisfaction within couples, independent of the man's age, the woman's age and the size and direction of the age difference within the couple. The distribution of the absolute difference in life satisfaction between the two partners is normally distributed around zero, with 40% of couples reporting zero difference.[4] It may be the case that the gender gap in life satisfaction at younger and older ages shown in Figure 2 reflects the respondents' partnership status. Indeed, we find that women are more likely to be partnered at younger ages and less likely to be partnered at older ages, as illustrated in Figure 3.

To find out whether the gender difference in life satisfaction disappears once we control for partnership status, we estimate now, in a first step, life satisfaction as a function of age, sex and partner status. We control for country fixed effects in order to focus on within-country variations only. The results presented in the following are based on OLS regressions.

The regression results appear in Table 1, with first a stripped-down model controlling only for age (as 10-year age group dummies) and country dummies; then a second specification with an additional control for sex; and, finally, a third specification controlling for both sex and partner status. An individual is considered as "partnered" when s/he is observed to be cohabiting with a partner at the time of

4 Guven et al. (2012) showed using SOEP data for Germany that life satisfaction gaps between partners predict future separation, but only when the woman is less satisfied than the man.

Figure 3:
Partnership status by age and sex



Note: The figures here are weighted averages over 32 European countries. The points refer to three-year (with respect to age) moving averages.
Source: EU-SILC 2013.

the survey. Note, however, that the results are robust to considering only those who are married as being "partnered". As the omitted age category in these regressions is 40–49, the estimated coefficients on the dummy variables for the other 10-year age categories should be read as relative to this age group.

Column 1 indicates that life satisfaction is highest for teenagers and lowest for those in their fifties, and again for those in their seventies.[5] Column 2 then controls for sex, which barely changes the estimated age coefficients. The difference in life satisfaction between men and women appears to be insignificant once we control for age. Finally, column 3 shows that there is a small but significant gender difference, with women being somewhat more satisfied than men once we control for age and partnership.[6] Partnership itself turns out to be highly correlated with overall life

[5] The finding that the U-shape bottoms out around the fifties has often been reported in the literature. Based on BHPS data, Clark (2019) found that life satisfaction reaches its lowest point at age 42 in cross-sectional analysis, and between ages 40–50 in panel regressions. In an analysis of four waves of the World Values Survey, Blanchflower and Oswald (2008) found that the average minimum level is reached at age 43. Graham and Ruiz Pozuelo (2017) reported that the minimum level is reached between the ages of 40 and 60 in 44 out of the 46 countries they analysed using Gallup World Poll data. A very thorough recent analysis of the U-shape using many different datasets appeared in Blanchflower (2020a), who again found that the U-shape bottoms out most often in the forties or fifties.

[6] The finding that women report higher life satisfaction than men is common in the literature: see Nolen-Hoeksema and Rusting (1999).

Table 1:
Life satisfaction by age, controlling for sex and partnership status

	Overall life satisfaction		
Age category:			
Ages 15–19	0.924***	0.924***	1.430***
	(44.86)	(44.84)	(67.81)
Ages 20–29	0.419***	0.419***	0.742***
	(33.03)	(33.01)	(56.99)
Ages 30–39	0.253***	0.253***	0.286***
	(21.61)	(21.61)	(24.71)
Ages 40–49	*Ref.*	*Ref.*	*Ref.*
Ages 50–59	−0.225***	−0.225***	−0.217***
	(−20.81)	(−20.81)	(−20.36)
Ages 60–69	−0.0644***	−0.0645***	−0.0373***
	(−5.88)	(−5.88)	(−3.44)
Ages 70–80	−0.229***	−0.229***	−0.0836***
	(−21.10)	(−21.09)	(−7.71)
Sex:			
Male	/	*Ref.*	*Ref.*
Female	/	−0.00494	0.0493***
		(−0.75)	(7.57)
Partnership status:			
Non-partnered	/	/	*Ref.*
Partnered	/	/	0.670***
			(92.52)
Constant	7.838***	7.840***	7.301***
	(371.15)	(366.08)	(332.44)
Country FE	Yes	Yes	Yes
N	370074	370074	370074
Adj. *R*-sq	0.153	0.153	0.172

Note: These are OLS regressions with country-fixed effects. Ages 15–80. *T*-statistics appear in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Source: EU-SILC 2013.

satisfaction, with partnered individuals being significantly more satisfied than non-partnered individuals. The effect size here is large, at two-thirds of a life satisfaction point, or almost one-third of a standard deviation (see the figures in the last column of the first panel of Appendix Table A.1). A comparison of columns 1 and 3 of Table 1 shows that controlling for partnership deepens the U-shape, which again suggests that partnership plays a role in the age-satisfaction relationship.

To see whether the age pattern of life satisfaction differs significantly between the partnered and the non-partnered respondents, we now run separate regressions

Table 2:
Life satisfaction by partner status

	Overall life satisfaction				
	Partnered	**Partnered**	**Non-partn.**	**Non-partn.**	**All**
Age category:					
Ages 15–19	0.257	0.245	1.624***	1.628***	1.690***
	(1.40)	(1.33)	(62.62)	(62.73)	(47.42)
Ages 20–29	0.271***	0.265***	1.076***	1.082***	1.065***
	(13.59)	(13.27)	(52.20)	(52.36)	(37.18)
Ages 30–39	0.255***	0.253***	0.392***	0.396***	0.465***
	(19.56)	(19.43)	(16.64)	(16.80)	(14.23)
Ages 40–49	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Ages 50–59	−0.235***	−0.234***	−0.148***	−0.150***	−0.184***
	(−19.79)	(−19.69)	(−6.55)	(−6.63)	(−5.47)
Ages 60–69	−0.104***	−0.101***	0.164***	0.158***	0.129***
	(−8.58)	(−8.31)	(7.32)	(7.06)	(3.69)
Ages 70–80	−0.203***	−0.196***	0.163***	0.154***	0.279***
	(−15.61)	(−15.02)	(8.12)	(7.60)	(8.41)
Sex:					
Male	/	*Ref.*	/	*Ref.*	*Ref.*
Female	/	0.0554***	/	0.0486***	0.0728*
		(7.01)		(4.15)	(2.32)
Partnership status:					
Non-partnered	/	/	/	/	*Ref.*
Partnered	/	/	/	/	0.953***
					(35.07)
Interaction: AgeCat#PartnershipStatus#Sex					
Ages 15–19,					−0.0981*
non-partnered, female					(−1.99)
Ages 15–19,					−1.052**
partnered, male					(−2.68)
Ages 15–19,					−1.600***
partnered, female					(−7.11)
Ages 20–29,					0.0767
non-partnered, female					(1.96)
Ages 20–29,					−0.851***
partnered, male					(−18.98)
Ages 20–29,					−0.801***
partnered, female					(−13.27)
Ages 30–39,					−0.113*
non-partnered, female					(−2.51)

Continued

Table 2:
Continued

	Overall life satisfaction				
	Partnered	**Partnered**	**Non-partn.**	**Non-partn.**	**All**
Interaction: AgeCat#PartnershipStatus#Sex					
Ages 30–39, partnered, male					−0.240***
					(−6.22)
Ages 30–39, partnered, female					−0.232***
					(−3.90)
Ages 40–49, partnered, female					−0.0411
					(−1.14)
Ages 50–59, non-partnered, female					0.0562
					(1.29)
Ages 50–59, partnered, male					−0.0851*
					(−2.23)
Ages 50–59, partnered, female					−0.0667
					(−1.12)
Ages 60–69, non-partnered, female					0.0444
					(1.00)
Ages 60–69, partnered, male					−0.252***
					(−6.38)
Ages 60–69, partnered, female					−0.253***
					(−4.16)
Ages 70–80, non-partnered, female					−0.148***
					(−3.61)
Ages 70–80, partnered, male					−0.446***
					(−11.73)
Ages 70–80, partnered, female					−0.558***
					(−9.25)
Constant	8.046***	8.016***	7.179***	7.151***	7.097***
	(320.35)	(314.81)	(192.65)	(188.88)	(231.12)
Country FE	Yes	Yes	Yes	Yes	Yes
N	232684	232684	137390	137390	370074
Adj. *R*-sq	0.172	0.172	0.158	0.158	0.175

Note: These are OLS regressions with country-fixed effects. Ages 15–80. T-statistics appear in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Source: EU-SILC 2013.

for the two groups, and we interact age, sex and partnership status. Table 2 shows the results, with the first two columns referring to the partnered individuals and columns 3 and 4 referring to the non-partnered individuals. Columns 1 and 3 control for age and country-fixed effects only, while columns 2 and 4 also include sex. Finally, column 5 shows the estimation results when we analyse the whole sample together and introduce a full set of interactions between age, sex and partnership.

The results are very different by partnership status. Columns 1 and 2 indicate that for the partnered individuals, the life satisfaction differences between most of the age groups are relatively small, with the U-shape of life satisfaction over age being rather flat. By contrast, for the non-partnered individuals, the effect size of the age coefficients is much larger, indicating a profound U-shaped relationship with life satisfaction, with those individuals in their fifties being the least satisfied. A comparison of columns 2 and 4 shows that the decline in life satisfaction from the teens to the fifties is almost four times larger for the non-partnered individuals than for the partnered individuals.[7] Finally, women are again found to report slightly higher life satisfaction than men in both partnership groups.

Column 5 of Table 2 shows the estimated coefficients based on a specification with interaction terms between the following three categorical variables: "age category", "partnership status" and "sex". The interaction coefficients are difficult to interpret on their own. They can, however, be used to calculate the estimated levels of "overall life satisfaction' for each possible combination of age category, sex and partnership status. For example, the estimated level of life satisfaction for a non-partnered man aged 40 to 49 (the reference categories) is 7.097, which corresponds to the intercept in column 5. The estimated level of life satisfaction for a non-partnered woman aged 40 to 49 is 7.1698, which is obtained by adding the estimated coefficient of "female' (0.0728) to the intercept. The estimated life satisfaction level of a partnered women aged 15 to 19 is 8.2128, which is calculated as follows: 7.097 (constant) + 0.0728 (estimated coefficient on "female") + 1.69 (estimated coefficient on age category "15–19") + 0.953 (estimated coefficient on "partnered") − 1.6 (estimated coefficient on the interaction "age 15–19, partnered, female").[8]

Figure 4 illustrates these estimated life satisfaction scores for each possible combination of age category, sex and partner status using the numbers from column 5 of Table 2. While life satisfaction is clearly U-shaped for the non-partnered individuals, the pattern is relatively flat for the partnered individuals. The U-shape is the most profound for the non-partnered men, who are the least satisfied at ages 50 to 59, and are somewhat more satisfied than the non-partnered women at ages 70–80. Note that the confidence intervals are large for the partnered women and men at ages 15 to 19, as few of the respondents cohabit with a partner before the age of 20.

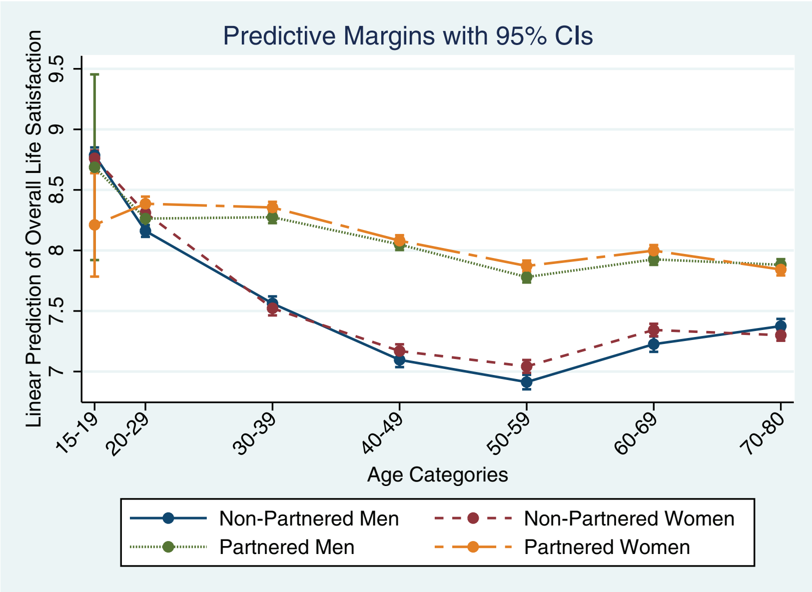
Figure 4 further shows that within each age and partnership category, the gender difference in life satisfaction is small. Independent of sex, the main difference in life satisfaction found among the groups studied is between the partnered and the non-partnered individuals, especially in mid-life. Thus, based on the raw data, the

---

[7] Partnership is then not only associated with higher mean life satisfaction, but also with lower variance, which might be thought to be valuable in its own right (see Clark et al. 2016).

[8] Note that all of our estimations include country fixed effects, so that the estimated coefficients refer to within-country differences in life satisfaction.

Figure 4:
Illustration of the predictive margins



Note: These are calculated from the OLS regression with country-fixed effects in column 5 of Table 2; ages 15–80.
Source: EU-SILC 2013.

gender difference in life satisfaction at younger and at older ages, as illustrated in Figure 2, is due to differences in partnership status. Until their early forties, women are happier than men because they are more likely than men to be partnered at these ages. From their sixties onwards, men are happier than women because they are more likely than women to be partnered at these older ages.

The results above come from a very stripped-down specification. However, they do not change in nature when we add more variables to look for potential confounders in the relationship between age, partnership and life satisfaction. We consider country of birth (Native, European, and other), the degree of urbanisation, living in a relatively rich country (as measured by GDP per capita in PPP terms), education, and both individual and household income. Within each education and income group, the U-shaped pattern of life satisfaction remains much sharper for the non-partnered than the partnered individuals.

We also consider different measures of subjective well-being, replacing “life satisfaction” with “meaning of life”, “satisfaction with financial situation”, “satisfaction with personal relationships” and the “ability to make ends meet”. We find a flatter U-shape for the partnered than for the non-partnered individuals for all four of these variables. In this case, the protection derived from being in a partnership may be partly financial. By contrast, we find no significant differences by age group, sex or partnership status for job satisfaction.[9] Finally, for satisfaction with time use, we find a U-shape in age, but with a minimum level (at age 30 to 39) that is lower than the figure for overall life satisfaction shown in Table 2. The highest level of satisfaction with time use is observed at ages 60 to 69. Few significant differences in satisfaction with time use between men and women or between partnered and non-partnered individuals are found.

We might wonder whether our results for being in a partnership reflect the presence of children in the household. The EU-SILC only provides information on the number and the ages of children who are living in the respondent's household. Moreover, having children at home becomes increasingly selective as the respondents grow older, when empty-nesting becomes more prevalent. We therefore controlled for children by limiting the sample to individuals aged 16 to 44.[10] We find that overall life satisfaction does not differ significantly by age for men and women with children, whereas it drops significantly from age 16 to age 44 for both childless men and childless women. However, within the group of partnered individuals, we detect no significant differences in life satisfaction by age between childless individuals and those with children. Thus, the presence of children seems to play a less important role in overall life satisfaction than partnership status, at least for adults up to age 44.[11]

Finally, we carry out a country-by-country analysis. In each of the 32 European countries in our sample, the pattern of overall life satisfaction over age is found to be relatively flat for the partnered individuals and U-shaped for the non-partnered individuals, independent of their sex.[12]

Thus, our cross-sectional analysis based on the EU-SILC data suggests that the U-shaped pattern of life satisfaction by age holds for the non-partnered individuals, but largely disappears for the partnered individuals. It therefore appears that gender differences in overall life satisfaction are mainly caused by age-specific gender differences in the likelihood of being partnered. These results confirm the role of partnership in well-being, as, for example, reported by Verbakel (2012), whose analysis of the European Values Study showed that married people have the highest level of well-being, followed by those who are cohabiting, dating or single; and, finally, by those who are widowed or divorced. Our results presented above suggest

---

[9] Helliwell et al. (2019) have suggested that the flatter U-shape for the married individuals may come about because "*together spouses can better shoulder the extra demands that may exist in mid-life when career and other demands coincide*". Our failure to confirm that this is the case for job or time-use satisfaction seems at odds with this hypothesis.

[10] The Eurobarometer data used by Blanchflower and Clark (2020) also only include information on children living at home. They too cut the sample of parents at age 45.

[11] Only when we focus on satisfaction with time use do we find that partnered childless men and women report significantly higher satisfaction than partnered men and women with children.

[12] We also note that overall life satisfaction is lower at higher ages in Eastern and Southern Europe than in Western and Northern Europe, conditional on partnership status and sex.

that being partnered also matters for the dynamics of well-being across age groups. To help us understand why this is the case, we can turn to the abundant empirical literature showing that being in a partnership provides a person not only with additional resources, but also with social and emotional support, and with health benefits. First, it has been shown that the economic benefits of being partnered, and, in particular, of being married, increase with income if income is taxed in a progressive way. The observation that income tends to be hump-shaped with age may explain why the protective effect on well-being of being in a partnership is particularly strong for middle-aged people. Second, as noted above, numerous studies have suggested that being married is associated with better psychological health. However, there is no strong evidence that this effect varies with age. In an analysis of data from the Danish Twin Registry, Kohler et al. (2005) found a slight increase in this effect with age for men, but no significant change in this effect for women. Finally, marriage has been shown to be associated with better health and longevity (see Wilson and Oswald 2005, for an extensive survey). For instance, Ben-Shlomo et al. (1993) found that married men have a lower risk of dying from cardiovascular disease and lung cancer, and Helmer et al. (1999) reported that never-married people have a significantly higher risk of developing Alzheimer's disease. As these pathologies tend to occur in the second part of life, these findings may help to explain why the protective role of partnership is attenuated for younger people.

However, the results of a cross-sectional analysis will not necessarily correspond to the findings of an analysis of panel data if there is selection into partnership. It is not only possible that those individuals who are more satisfied select into partnerships (as in Stutzer and Frey 2006); it also is quite likely that people who are not satisfied with their relationships (and therefore have lower overall life satisfaction scores) select out of partnerships.[13] Once these individuals find a new relationship that makes them happy, they will re-enter our sample of cohabiting partners. Both of these sources of endogeneity make it more difficult to conclude that being in a partnership protects against the U-shaped pattern, as the people who are partnered (or are non-partnered) are not the same people at different ages.

To shed light on the causal impact of being in a partnership on life satisfaction from a life-cycle perspective, we need to add panel data to our cross-sectional analysis. If being in a partnership indeed protects against the U-shaped pattern, we should find that the age profile of the partnered individuals in the panel data (i.e., within the same person) is flatter than that of the non-partnered individuals. The EU-SILC data does not help us here, given its four-year revolving design (and with information on life satisfaction only being available in the 2013 and 2018 survey waves). Thus, to examine this question, we turn to panel data from three of

[13] See the lags and leads analysis of life satisfaction around the time of marital separation using Australian, British and German panel data in Figure 5.3 of Clark et al. (2018).

the longest-running panel datasets: the HILDA (Australia), the BHPS (the United Kingdom) and the SOEP (Germany).[14]

3 Life satisfaction, age and partnership in panel data

Our analysis of panel data uses the same kinds of specifications as the cross-sectional analysis above. The dependent variable, life satisfaction, is on a 0–10 scale in the HILDA and the SOEP, and a 1–7 scale in the BHPS.[15] We standardise the life satisfaction scores, so that our correlations are in terms of standard deviations. We consider those individuals aged 16–80, and split up age into dummies for the categories 15–19, 20–29, 30–39, 40–49, 50–59, 60–69 and 70–80. Partnered is a dummy for marriage or cohabitation. There is a heated discussion in the literature regarding the controls that should be introduced when analysing the relationship between subjective well-being and age.[16] Here, as in our analysis of the EU-SILC data above, we take a minimalist approach, and introduce no other controls. All of our regressions include an individual fixed effect (which subsumes the effect of sex, plus any other variable that is time-invariant).

The regression to be estimated is

$$\mathrm{LS}_{it,m} = \boldsymbol{\beta}'\mathbf{Age}_{it,m} + \mu_i + \varepsilon_{it} \tag{1}$$

where $\mathrm{LS}_{it,m}$ is the life satisfaction of individual i at time t in marital status m, $\mathbf{Age}_{it,m}$ is a set of age dummies, and μ_i is the individual fixed effect. The presence of the latter implies that the age coefficients, $\boldsymbol{\beta}$, are identified by the same individual as s/he moves from one age category to the next in the panel data.

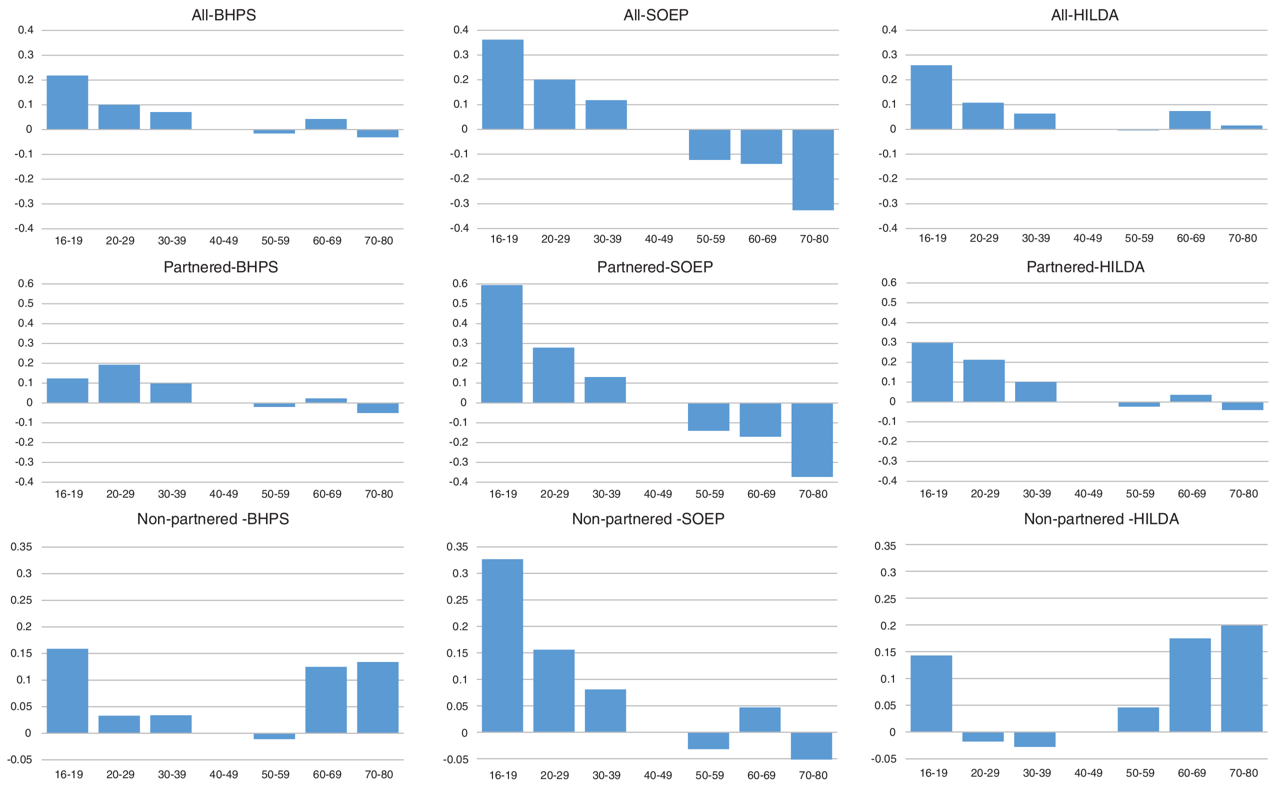
The estimated coefficients from these regressions on our three panel datasets are depicted in Figure 5. The three columns refer to the BHPS, the SOEP and the HILDA, respectively. Within each column, the first graph shows the estimated age coefficients from a regression that includes the whole sample, while the second and the third graphs depict those for the partnered and the non-partnered individuals separately. The Y-axes are on the same scale within each row. It is immediately obvious that the largest age coefficients are found in the SOEP.

[14] We have data from 2001–2016 for the HILDA, from 1984–2016 for the SOEP, and from 1996–2000 and 2000–2008 for the BHPS (these are the only BHPS waves that include life satisfaction information; the survey stopped in 2008).

[15] Appendix Table A.1 lists the descriptive statistics for life satisfaction and the percentage partnered, by age, in these three panel surveys.

[16] The discussion here revolves around confounders and mediators, with the former affecting both the dependent variable and the right-hand side variables, and the mediators describing the pathways via which the relationship from the right-hand side variable to the dependent variable pass. This distinction is less obvious when the explanatory variable is exogenous (such as age, sex or ethnicity). Adding controls helps to explain how the relationship works, but not whether there is a relationship in the first place. This debate is ongoing: see Blanchflower (2020a).

Figure 5:
Life satisfaction, age and partnership in panel data



Note: These are OLS regressions with individual fixed effects. Ages 15–80.
Source: BHPS, SOEP and HILDA.

Looking at the first row, we can see that the estimated age coefficients suggest that there is something of a U-shaped relationship in both the BHPS and the HILDA. By contrast, the panel results for the SOEP produce a monotonically negative relationship between life satisfaction and age.[17]

The most important element of Figure 5 in the context of our analysis here is the comparison of the shapes of the estimated coefficients in the partnered and non-partnered graphs in the second and third rows. A broad characterisation of the results is that the age profile is more U-shaped for the non-partnered than for the partnered individuals, as in the EU-SILC data. However, this distinction is somewhat less apparent in the SOEP, and, in general, there is still a noticeable age profile in life satisfaction for the partnered individuals (whereas the age profile in life satisfaction for the partnered individuals in the cross-sectional EU-SILC data was largely flat).[18]

Thus, there is only partial agreement between our pooled EU-SILC results presented in Section 2 and the results from the three panel datasets in Figure 5: i.e., in the panel data, the partnered individuals continue to drop down the first arm of the U-shape at least as much as the non-partnered individuals. Below, we consider a number of potential explanations for the differences between the pooled and panel results.

3.1 Different countries

There are 32 different countries in the EU-SILC, and only two of these countries appear in our panel analysis above. However, when we consider the EU-SILC results for the UK and Germany on their own, we see that they are the same as those for all of the 32 countries covered in Section 2 above.

---

[17] We can speculate as to why the German results are so different. Note that this cannot be a cohort effect, as the estimated coefficients come from within-individual regressions. Even so, it may be the case that people who grew up during World War II are less equipped to deal with ageing, and that their age-life satisfaction profile is therefore different. We use SOEP waves from 1984 to 2016. Thus, the average person in the 50–59 age group in this dataset was born in 1945. The birth years for those aged 60–69 and 70–80 were 1935 and 1925.

[18] The year coverage of the three panels is not the same. All of our panel regressions include year dummies that will pick up any common macro effects that may be correlated with life satisfaction and marital status. It is, however, possible that the relationship between partnership and life satisfaction itself has changed over time (and within each individual, as these are all panel regressions). To check whether this is the case, we re-estimate the equations behind Figure 5 only from 2000 onwards, and then only for the overlap years for the three panels, from 2001 (the start of the HILDA) to 2008 (the end of the BHPS). Reassuringly, we find that the data shapes are similar to those in Figure 5.

3.2 Data comparability

Alternatively, there could be some design issues in the BHPS, the SOEP and the HILDA that make the results of an analysis of life satisfaction and age based on the data from these surveys not comparable to the results of an analysis of 2013 cross-sectional data from the EU-SILC. However, these three panel datasets have been very widely analysed in the literature, and when we look at them in a cross-sectional manner, we find exactly the same kind of U-shaped patterns as those displayed in Figure 4 above: i.e., there is a U-shape in age for both the partnered and the non-partnered individuals.[19]

3.3 Mortality

The U-shape between life satisfaction and age is (of course) estimated using data on the living. Thus, one potential explanation for the upward-sloping arm of the U-shape is that the dissatisfied die off, leaving an increasingly selected sample of the (satisfied) living.[20] As such, if mortality and subjective well-being were unrelated, the well-being profile of the older individuals would be skewed downwards.

Can mortality explain the different well-being age profiles for the partnered and the non-partnered individuals in the cross-sectional analyses? For this to be the case, we would first need to observe excess mortality among the partnered individuals (relative to the non-partnered individuals), and then that this excess mortality was especially pronounced in middle age. Neither of these possibilities appear likely.

First, the UK death rates suggest that there are “not enough” deaths to explain the U-shape: i.e., the cumulative death rate is only 11 per 1000 from age 16 to age 55, and is 63 per 1000 from age 16 to age 75. Even taking the extreme position of assigning a life satisfaction value of zero to these missing people reduces average life satisfaction by less than 0.1 of a point at age 55, and by around 0.5 of a point at age 75; these figures are to be compared to the actual observed life satisfaction gap of -1.2 points at both ages.

Second, there would have to be excess mortality among the partnered individuals. However, the existing evidence (see Kravdal et al. 2018) suggests that there is a negative relationship between marriage and mortality: protecting the married from mortality means that the non-partnered individuals are more selected, such that their life satisfaction scores are “too high”—which is the opposite of what we find in Figure 4.

A related point is attrition, whereby individuals drop out of the panel for whatever reason. The likelihood of dropping out is known to be related to subjective well-being, and it is indeed the case that in all three of our panels, lower life satisfaction

19 This cross-sectional U-shape minimises at age 40–49 in the BHPS and the HILDA, and at 50–59 in the SOEP.

20 A well-known contribution here is Danner et al. (2001).

at time t predicts dropping out of the panel at time $t + 1$. However, we find no significant difference in this satisfaction-attrition relationship by partnership status, or by partnership status by age group (mid-life, which we set at ages 35–55, versus the rest). Thus, selective attrition does not seem to explain our panel results.

3.4 Selection into marriage

Finally, we consider the role of selection into and out of a partnership, and whether this composition effect can explain the cross-sectional gap in life satisfaction between the partnered and the non-partnered individuals shown in Figure 4.

First, we find that the percentage of people who are partnered changes sharply with age. In the EU-SILC data, the percentage of people who are partnered is only 10% in the early twenties, but is 80% by age 40. This figure tails off at older ages. One way of squaring the (flat) cross-sectional and (more U-shaped) panel results is selection into and out of marriage: each individual follows his/her own U-shape as s/he ages, but there are composition effects of different well-being types in and out of a partnership at different ages.

Following our results above, this would mean either that those individuals who select into a partnership in their thirties and forties are more satisfied types (in terms of Equation (1), that their average value of μ_i is higher) than those who become partnered in their twenties or sixties, and/or that those individuals who select out of marriage in their thirties or forties are unhappier types. Alternatively, we could allow for heterogeneity in the relationship between life satisfaction and being in a partnership, with those individuals who become partnered in their thirties and forties experiencing a greater life satisfaction boost from it than those who marry earlier or later.

Figures A.1 and A.2 in the Appendix investigate both of the above possibilities. In Figure A.1, the left-hand side shows the level of life satisfaction of those individuals who move from non-partnership to partnership status (this is the life satisfaction score they report in the year before they make this move), while the right-hand side shows the rise in life satisfaction reported by individuals as they move from being non-partnered in wave $t - 1$ to being partnered at wave t. The three rows show these numbers for the BHPS, the SOEP and the HILDA.

To explain the difference between the pooled and the panel results, we would need those individuals who move from non-partnership to partnership status in their thirties and forties to be more satisfied types than those who make this transition earlier or later (on the left-hand side of the figure), or to enjoy a greater satisfaction boost from doing so (on the right-hand side of the figure): thus, we should be looking for hump-shapes in age.

Figure A.1 shows the results for the NP-P movers. In the BHPS and the SOEP data, there is some evidence of a hump-shape in life satisfaction levels (although this occurs in the twenties and thirties, rather than in the thirties and forties). When we look at the change in life satisfaction shown on the right-hand side, two points stick out. The first is that there is a large decline in life satisfaction associated with teen partnership in the UK and Australia, but a large rise in Germany. The second is that the increase in life satisfaction is larger for those individuals who become partnered in their fifties, except in Australia. This will flatten the cross-sectional age profile.

Figure A.2 shows the results of the same exercise for those individuals who leave a partnership. There is no column for ages 16–19 here, as very few people separate before age 20. As noted above, we are looking for a U-shaped relationship. There is no evidence of such a relationship in the change in life satisfaction around the time of transition, as shown on the right-hand side. However, the levels of life satisfaction shown on the left-hand side do display a distinct U-shape: those individuals who move out of a partnership in their forties and fifties are notably less satisfied than those who leave at other ages. Thus, the flatter age profile for the partnered individuals in the cross-sectional data may be partly explained by a composition effect, with the unhappy individuals moving from partnered to non-partnered status at this age, which mechanically raises the average life satisfaction score of those who remain in the partnered group.

4 Conclusion

While the finding in the literature that there is a U-shaped relationship between life satisfaction and age is robust, it has not been fully explored. This relationship has been found for multiple countries and datasets (Blanchflower 2020a and 2020b), and often continues to hold even when large numbers of controls are considered, ranging from observable mid-life burdens to aspirations. Thus, there is still much to be learned about the reasons why the U-shaped pattern exists. In this article, we have analysed the possible role of a rather intuitive factor—being in a partnership—and examined the question of whether being partnered could help protect individuals against the mid-life decline in satisfaction. In general, we expected to find that the U-shape would be deeper for those individuals who lacked certain physical or emotional resources.

Our findings were mixed, and depended strongly on the type of data analysed. The life satisfaction of the partnered individuals was shown to be rather flat with age in the cross-sectional data we analysed, but displayed a more U-shaped pattern in the three panel datasets we considered. We believe this difference partly reflects a composition effect: i.e., that those who leave partnerships in their forties and fifties are noticeably less satisfied with their lives than those who break up at younger or older ages.

To study how key demographic, economic and psychological variables change over the life-cycle, reliable longitudinal datasets are needed. As we have seen above, imputing life-cycle changes from cross-sectional data can be hazardous. While developing and supporting panel datasets will require considerable investments by public funding agencies, longitudinal data are essential to research on the evolution of societies, which are composed of individuals who change over time. Having access to panel datasets will also facilitate inter-disciplinary work, with well-being arguably being a good example of a topic that could be analysed jointly by researchers in a number of different fields.

Acknowledgments

We gratefully acknowledge financial support from the Agence Nationale de la Recherche within the JPI MYBL framework (Award No. ANR-16-MYBL-0001-02) and EUR grant ANR-17-EURE-0001. We thank Conchita D'Ambrosio, Andrew Oswald, Nick Powdthavee and seminar participants at ISQOLS (Granada), the Age Well Accounts workshop in Vienna and the Demographic Aspects of Human Wellbeing Conference for their helpful comments. Two referees made a number of very useful points that helped us make our analysis clearer. We are very grateful to Anthony Lepinteur for his help and advice with the panel data analysis.

References

Ben-Shlomo, Y., G. D. Smith and M. Shipley 1993. Magnitude and causes of mortality differences between married and unmarried men. *Journal of Epidemiology and Community Health* 47: 200–205.

Blanchflower, D. G. 2020a. Is happiness U-shaped everywhere? Age and subjective well-being in 145 countries. *Journal of Population Economics* 34: 575–624. https://doi.org/10.1007/s00148-020-00797-z

Blanchflower, D. G. 2020b. Unhappiness and age. *Journal of Economic Behavior and Organization* 176: 461–488. https://doi.org/10.1016/j.jebo.2020.04.022

Blanchflower, D. G. and A. E. Clark 2020. Children, unhappiness and family finances. *Journal of Population Economics* 34: 625–653. https://doi.org/10.1007/s00148-020-00798-y

Blanchflower, D. G. and A. J. Oswald 2008. Is wellbeing U-shaped over the life cycle? *Social Science & Medicine* 66: 1733–1749. https://doi.org/10.1016/j.socscimed.2008.01.030

Blanchflower, D. G. and A. J. Oswald 2019. Do humans suffer a psychological low in midlife? Two approaches (with and without controls) in seven data sets. In *The economics of happiness: How the Easterlin paradox transformed our understanding of well-being and progress*, ed M. Rojas, 439–453. New York: Springer. https://doi.org/10.1007/978-3-030-15835-4_19

Cheng, T., A. J. Oswald and N. Powdthavee 2017. Longitudinal evidence for a midlife nadir in human well-being: Results from four data sets. *Economic Journal* 127: 126–142. https://doi.org/10.1111/ecoj.12256

Clark, A. E. 2016. SWB as a measure of individual well-being. In *Oxford handbook of well-being and public policy*, eds M. Adler and M. Fleurbaey. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199325818.013.17

Clark, A. E. 2019. Born to be mild? Cohort effects don't (fully) explain why well-being is U-shaped in age. In *The economics of happiness: How the Easterlin paradox transformed our understanding of well-being and progress*, ed M. Rojas, 387–408. New York: Springer. https://doi.org/10.1007/978-3-030-15835-4_17

Clark, A. E., S. Flèche, R. Layard, N. Powdthavee and G. Ward 2018. *The Origins of happiness: The science of well-being over the life-course*. Princeton NJ: Princeton University Press.

Clark, A. E., S. Flèche and C. Senik 2016. Economic growth evens-out happiness: Evidence from six surveys. *Review of Income and Wealth* 62: 405–419. https://doi.org/10.1111/roiw.12190

Danner, D., D. Snowdon and W. Friesen 2001. Positive emotions in early life and longevity: Findings from the nun study. *Journal of Personality and Social Psychology* 80: 804–813. https://psycnet.apa.org/doi/10.1037/0022-3514.80.5.804

Diener, E. 2006. Guidelines for national indicators of subjective well-being and ill-being. *Journal of Happiness Studies* 7: 397–404. https://psycnet.apa.org/doi/10.1007/s10902-006-9000-y

Eurostat 2012*a*. EU-SILC module on well-being. Assessment of the implementation.

Eurostat 2012*b*. EU-SILC 2013 module on well-being - description of SILC secondary target variables. Version 5 March 2012.

Gove, W. R., C. B. Style and M. Hughes 1990. The effect of marriage on the well-being of adults. A theoretical analysis. *Journal of Family Issues* 11: 4–35. https://doi.org/10.1177%2F019251390011001002

Graham, C. and J. Ruiz Pozuelo 2017. Happiness, stress, and age: How the U-curve varies across people and places. *Journal of Population Economics* 30: 225–264. https://doi.org/10.1007/s00148-016-0611-2

Guven, C., C. Senik and H. Stichnoth 2012. You can't be happier than your wife. Happiness gaps and divorce. *Journal of Economic Behavior & Organization* 82: 110–130. https://doi.org/10.1016/j.jebo.2012.01.003

Helmer, C., D. Damon and L. Letenneur 1999. Marital status and risk of Alzheimer's disease. *Neurology* 53: 1953–1958. https://doi.org/10.1212/WNL.53.9.1953

Helliwell, J., H. Huang, M. Norton and S. Wang 2019. Happiness at different ages: The social context matters. In *The economics of happiness: How the Easterlin paradox transformed our understanding of well-being and progress*, ed M. Rojas, 455–481. New York: Springer. https://doi.org/10.1007/978-3-030-15835-4_20

Jebb, A., M. Morrison, L. Tay and E. Diener 2020. Subjective well-being around the world: Trends and predictors across the life span. *Psychological Science* 31: 293–305. https://doi.org/10.1177%2F0956797619898826

Kassenböhmer, S. C. and J. Haisken-DeNew 2012. Heresy or enlightenment? The well-being age U-shape effect is flat. *Economics Letters* 117: 235–238. https://doi.org/10.1016/j.econlet.2012.05.013

Kohler, H., J. Behrman and A. Skytthe 2005. Partner + children = happiness? The effects of partnerships and fertility on well-being. *Population and Development Review* 31(3): 407–445. https://doi.org/10.1111/j.1728-4457.2005.00078.x

Kravdal, O., E. Grundy and K. Keenan 2018. The increasing mortality advantage of the married: The role played by education. *Demographic Research* 38: 471–512. https://doi.org/10.4054/DemRes.2018.38.20

Mikucka, M. 2016. The life satisfaction advantage of being married and gender specialization. *Journal of Marriage and Family* 78: 759–779. https://doi.org/10.1111/jomf.12290

Nolen-Hoeksema, S. and C. L. Rusting 1999. Gender differences in well-being. In *Well-being: The foundations of hedonic psychology*, eds D. Kahneman, E. Diener and N. Schwartz. New York: Russell Sage Foundation.

Stutzer, A. and B. S. Frey 2006. Does marriage make people happy, or do happy people get married? *Journal of Socio-Economics* 35: 326–347. https://doi.org/10.1016/j.socec.2005.11.043

Verbakel, E. 2012. Subjective well-being by partnership status and its dependence on the normative climate. *European Journal of Population/Revue Européenne de Démographie* 28: 205–232. https://doi.org/10.1007/s10680-012-9257-2

Wilson, C. M. and A. J. Oswald 2005. How does marriage affect physical and psychological health? A survey of the longitudinal evidence. IZA Discussion Papers 1619, Institute of Labor Economics (IZA). http://ftp.iza.org/dp1619.pdf

Appendix

Table A.1:
Life satisfaction, partnership and age: Summary statistics

	15–19	20–29	30–39	40–49	50–59	60–69	70–80	All
EU-SILC								
N	*10 795*	*39 256*	*51 365*	*65 139*	*69 543*	*65 418*	*68 410*	*369 926*
Life satisfaction (0–10)	7.80	7.37	7.20	6.98	6.77	6.96	6.84	7.02 (SD 2.06)
% Partnered	.008	.31	.69	.73	.71	.68	.51	.60
BHPS								
N	*9 869*	*24 928*	*29 725*	*27 593*	*23 683*	*18 087*	*14 937*	*148 822*
Life satisfaction (1–7)	5.33	5.18	5.12	5.01	5.15	5.48	5.58	5.22 (SD 1.28)
% Partnered	.004	.21	.61	.69	.74	.72	.55	.54
SOEP								
N	*25 982*	*91 993*	*113 673*	*121 915*	*96 842*	*76 424*	*52 106*	*578 935*
Life satisfaction (0–10)	7.52	7.25	7.19	7.03	6.88	7.09	7.05	7.10 (SD 1.79)
% Partnered	.009	.25	.68	.75	.77	.77	.64	.62
HILDA								
N	*17 302*	*41 847*	*40 145*	*42 466*	*36 424*	*26 886*	*18 692*	*223 762*
Life satisfaction (0–10)	8.17	7.87	7.75	7.66	7.78	8.14	8.34	7.89 (SD 1.48)
% Partnered	.004	.19	.57	.63	.64	.67	.61	.49

Figure A.1:
The life satisfaction of individuals who move into a partnership

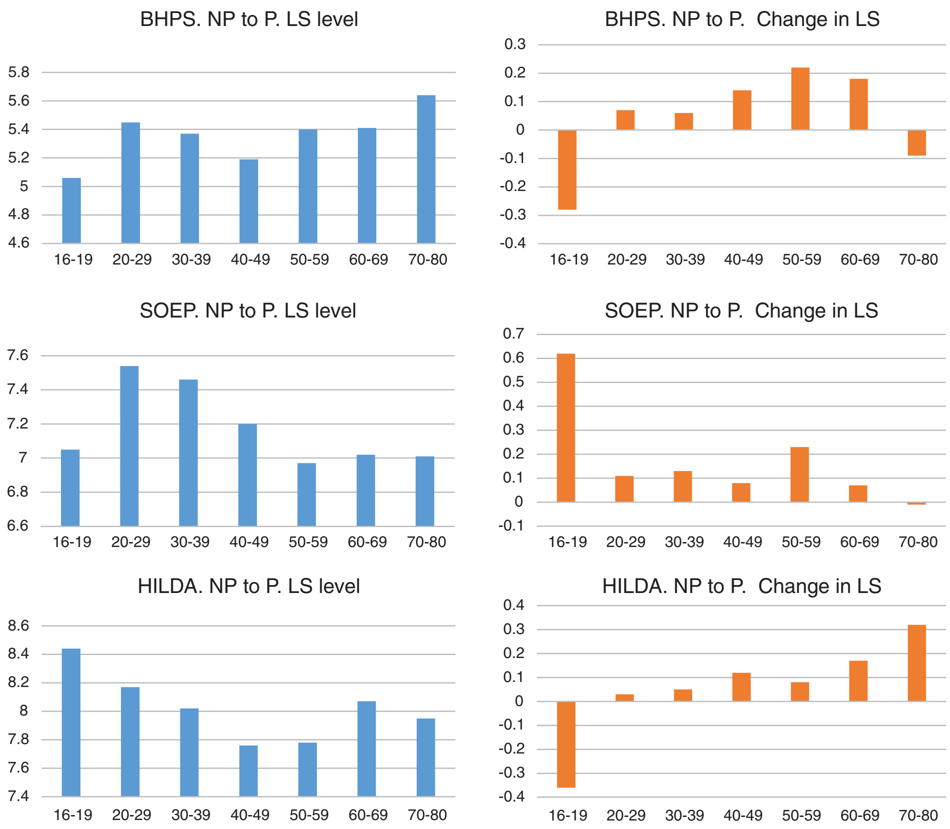
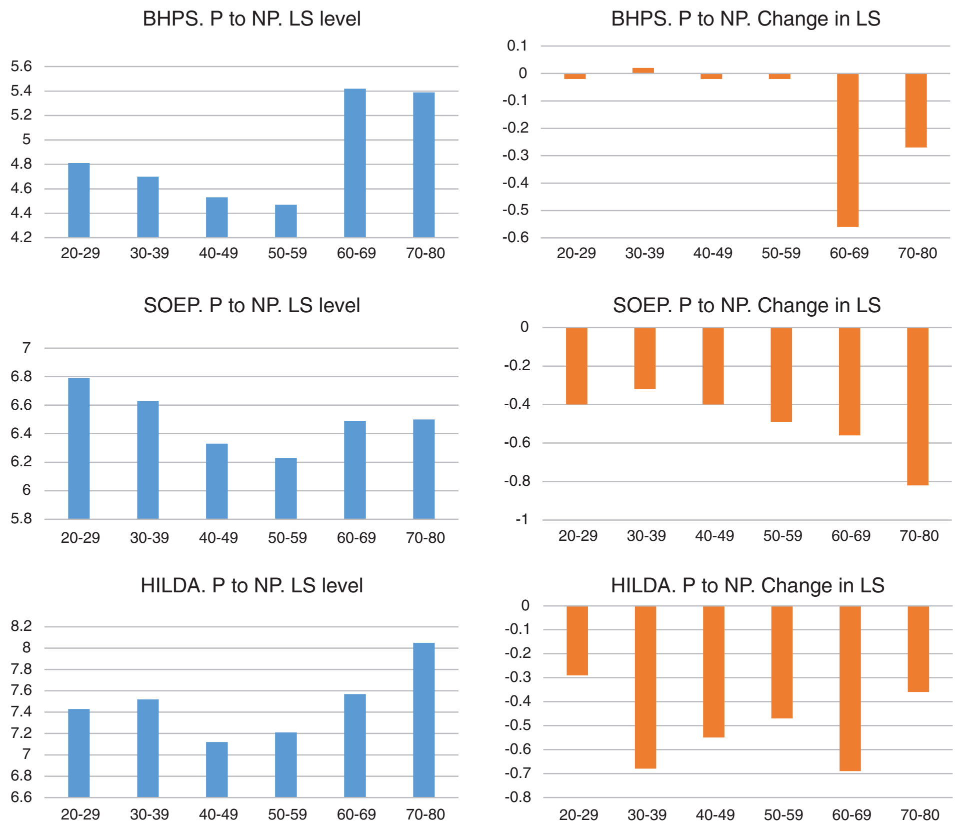


Figure A.2:
The life satisfaction of individuals who move out of a partnership



Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 319–353

Singles in the city: Happily ever after?

***Bernhard Riederer*[1,2,*]*, Nina-Sophie Fritsch*[3,4] *and Lena Seewann*[4]**

Abstract

More people than ever are living in cities, and in these cities, more and more people are living alone. Using the example of Vienna, this paper investigates the subjective well-being of single households in the city. Previous research has identified positive and negative aspects of living alone (e.g., increased freedom vs. missing social embeddedness). We compare single households with other household types using data from the Viennese Quality of Life Survey (1995–2018). In our analysis, we consider overall life satisfaction as well as selected dimensions of subjective well-being (i.e., housing, financial situation, main activity, family, social contacts, leisure time). Our findings show that the subjective well-being of single households in Vienna is high and quite stable over time. While single households are found to have lower life satisfaction than two-adult households, this result is mainly explained by singles reporting lower satisfaction with family life. Compared to households with children, singles are more satisfied with their financial situation, leisure time and housing, which helps to offset the negative consequences of missing family ties (in particular with regard to single parents).

Keywords: singles; city; Vienna; subjective well-being; comparison of household types

1 Introduction

In recent decades, Western societies have undergone a huge transformation in family and household structures, which has been partly caused by fundamental

[1]Vienna Institute of Demography (OeAW), Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Austria

[2]Institute for Sociology, University of Vienna, Austria

[3]Institute for Sociology and Social Research, Vienna University of Economics and Business, Austria

[4]Department of Sociology, University of Potsdam, Germany

[*]Correspondence to: Bernhard Riederer, bernhard.riederer@oeaw.ac.at and bernhard.riederer@univie.ac.at

DOI: 10.1553/populationyearbook2021.res3.2

changes in norms and values (Hettlage 2000; Lesthaeghe 2010; Van de Kaa 1987). Individualisation and the deinstitutionalisation of the nuclear family (Bengtson 2001; Cherlin 2004) have led to a rising variety of household arrangements and a proliferation of alternative living arrangements (Buzar et al. 2005; Friedlander et al. 1999; Kuijsten 1995). In particular, the share of people living alone has increased sharply in Western industrialised countries (Chandler et al. 2004). In response to these developments, social scientific research has shown that so-called 'singles' have gained in societal relevance (Hradil 1995; Poortman and Liefbroer 2010). The term 'single' is usually defined in reference to a person's living arrangement and/or relationship status (see Hertel et al. 2007). In the present article, however, we mainly refer to single households (i.e., individuals living alone). Recent research has described the new trend towards living alone as a personal lifestyle that people choose, rather than as the result of a lack of alternatives (Hertel et al. 2007). This new way of life has become increasingly common in part because of more liberal attitudes towards people living alone, and in part because people have an increasing economic ability to do so (Ogden and Hall 2000). Mass urbanisation has also provided the conditions for the rise of single households, because it has led to the emergence of booming subcultures of singles, and of an infrastructure that supports their lifestyles (Klinenberg 2012). Today, single households seem to be predominantly a phenomenon of urban metropolitan regions, where people who live alone cluster together in specific areas (Bertram 1993, p. 4; Klinenberg 2012). Accordingly, the increase in single households has been most pronounced in global cities like Paris, London and New York (Buzar et al. 2005; Hall and Ogden 2003; Hall et al. 1997). Thus, 'living alone' is mainly a feature of urban life (Bennett and Dixon 2006; Chandler et al. 2004).

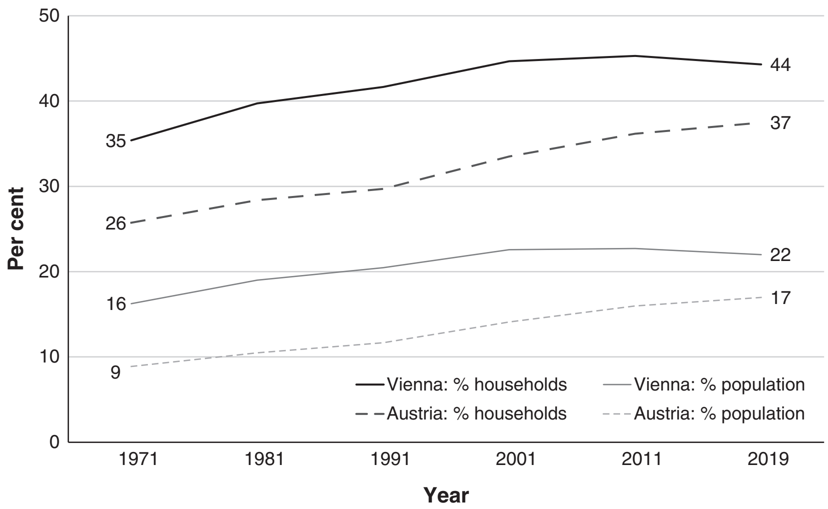
Against this background, the present paper addresses the subjective well-being of *single households in the city*. Previous research on quality of life has repeatedly shown that family status is among the most relevant predictors of subjective well-being (Frey and Stutzer 2002; Haller and Hadler 2006; Layard 2005; Riederer 2018). Within this research tradition, a plethora of studies have analysed the extent to which marital status or living arrangements influence life satisfaction or happiness (e.g. Blekesaune 2018; Chang 2013; Mastekaasa 1994; Soons and Kalmijn 2009; Vanassche et al. 2013). However, there is little quantitative research on subjective well-being that has focused explicitly on the *well-being of those who live alone* (exceptions are, for example, Luhmann and Hawkley 2016; Soons and Liefbroer 2008). Broadening our perspective, we assume that life satisfaction and happiness are affected not just by family ties, but also by developments in many other areas of life. While quality of life research indicates that not being in an intimate relationship is associated with lower well-being among people who are living alone, research on social change has emphasised that having personal freedom (including gains in independence and leisure time) enables single people to develop new lifestyles, and thus positively influences their levels of well-being. Hence, the question of whether the subjective well-being of single households is lower overall is anything but trivial. The answer to this question would likely be interesting not only for quality of life researchers and family scientists, but also for demographers and urbanologists, as many cities are increasingly affected by both the prevalence of smaller households and issues of urban quality of life (Brown et al. 2016; Hogan et al. 2016; Krekel et al. 2016; Marans 2012). Against this background, we aim to answer the following research questions: *Are people who live alone more or less satisfied with their lives than others? Which specific domains of subjective well-being are relevant to explain the revealed differences in life satisfaction between household types? And, which factors contribute to the differences in well-being domains between household types?*

Our comparative analysis of the well-being of single households contributes to the state of research in various ways. Going beyond previous studies, we include in our analysis different domains of subjective well-being (i.e., housing, financial situation, main activity, family, social contacts, leisure time) in order to disentangle the broad concept of quality of life, rather than examining a concept that is operationalised through a single question (Vanassche et al. 2013). This approach allows us to determine which 'components' of well-being contribute the most to the differences observed between single households and other household types. Moreover, our study adds to research on quality of life in metropolitan regions. Although there is a large body of existing research in this area,[1] there has been, to our knowledge, no systematic study of the *subjective quality of life* of *singles in the urban context*, as most of this previous research focused on specific groups and topics (e.g., the elderly in modern smart, digital, e-participating cities). Using the example of Vienna, the present study aims to help fill this research gap. To do so, we analyse a unique dataset, the *Viennese Quality of Life Survey* (total N = 41,920), which covers almost a quarter of a century (1995–2018). Our analyses are of descriptive nature due to the cross-sectional character of the data on the individual level. Nevertheless, we perform a series of regression and decomposition analyses in order to shed light on how specific life domains influence the differences in overall life satisfaction between household types, and to explore the reasons why these differences exist.

The city of Vienna is an interesting case to study. *First*, the composition of Viennese households clearly demonstrates that singles are an urban phenomenon. Of the private households in Vienna, more than 400,000, or 44 per cent, are single households. Figure 1 illustrates the significant growth of single households since 1971 in Vienna, and in Austria as a whole. Today, 22 per cent of Viennese residents and 17 per cent of Austrian residents live in single households (Statistics Austria 2020). *Second*, compared to other European cities, Vienna has a similar or even

[1] Studies on quality of life in metropolitan areas can be largely assigned to three main thematic areas: (a) studies emphasising changes in population composition, i.e., demographic rejuvenation and the ageing of cities (Bouzarovski et al. 2010, p. 214); (b) quality of urban life studies that refer to perceptions of the residential environment, e.g., satisfaction with housing or leisure activities (Marans 2012; Marans and Stimson 2011; Weziak-Bialowolska 2016) and their interrelationship with overall quality of life (Ge and Hokao 2006; Okulicz-Kozaryn 2013); and (c) quality of life analyses that focus on objective indicators such as climate characteristics, the availability of public services or the unemployment rate of cities (Ballas 2013; Kahn 1995; Zenker et al. 2013).

Figure 1:
The rise of single households in Vienna and Austria (1971–2019)



Notes: The graphs show the share of single households (i.e. one-person households) in total households (darker grey) and the share of persons living in single households in total population (lighter grey) for Vienna (continuous lines) and Austria (dotted lines). The figure for persons living in single households in Vienna in 2019 refers to 2017.
Source: Statistics Austria (2020).

larger share of the population who are living alone (e.g., 22 per cent in Vienna, 21 per cent in Brussels, 19 per cent in Stockholm, 18 per cent in Rome, 17 per cent in Paris)[2] (Eurostat 2019).

2 Theoretical concepts and hypotheses

According to the literature, well-being is a relatively stable attribute that reflects the extent to which individuals experience positive affect and have favourable views of themselves and their lives (Kamp Dush and Amato 2005, p. 608). Some researchers have differentiated between an emotional component (positive and negative affect) and a cognitive component (life satisfaction as an evaluative judgement) of well-being. In our paper, we focus on the cognitive component of subjective well-being (i.e., satisfaction). We also rely on theoretical concepts of separate *domains* of subjective well-being. A glance at the literature reveals that a wide range of concepts to capture well-being have been proposed (Allardt 1976; Campbell et al. 1976; Diener 2009; Diener and Suh 1997; Tov and Diener 2007; Veenhoven 2000,

2 The data in the text refer to 2015 to 2017. Outliers are Madrid (12) or Berlin (31 per cent in 2012). The figures refer to the population living in private households. Institutional households are excluded.

2007). Whereas Lindenberg, for instance, distinguished between material and social components (Ormel et al. 1999),[3] Diener and Fujita (1995) referred to a related concept of material, personal and social resources that fulfil people's needs, and, in turn, enhance their well-being.[4] Based on these concepts, we differentiate between *material*, *social* and *cross-sectional domains of well-being*. *Material* well-being refers to satisfaction with one's present *housing* situation and *financial* situation. *Social* well-being refers to satisfaction with one's *family* ties and *social contacts*. Finally, the *cross-sectional* domains refer to satisfaction with one's current *main activity (e.g., employment)* and *leisure time activities*. The latter two domains are both related to material as well as social well-being. Employment, for instance, is a main source of financial as well as social gratification; whereas leisure time activities depend on the person's time resources, social contacts and financial means. The concept of separate well-being domains can help to explain the differences between singles and other household types by pointing out how the situations of household types differ in specific life domains. In the following, we summarise the existing literature on the differences between various kinds of households in these domains, and present our resulting predictions in Table 1.

2.1 Differences between singles and other household types regarding material well-being

In terms of financial well-being, singles are less likely than other households to experience financial pressure due to commitments to a partner and/or children, and are more likely to feel free to spend money on themselves. In addition, because singles tend to have more (time) resources than other households to devote to their careers, they are more likely to achieve better-paid positions, and they may advance more quickly (Klinenberg 2012). Previous research has shown that in recent decades, the share of single-adult households in the middle-income segment has increased across OECD countries (Salvatori and Manfredi 2019, p. 34). By contrast, households with dependent children often have more difficulties making ends meet, and single-parent families in particular face a higher risk of falling below the poverty line in all EU member states, and in most OECD countries (Brady and Burroway 2012; Maldonado and Nieuwenhuis 2015; Riederer et al. 2017). For Austria, Zartler et al. (2011) found that compared to the total population, single-parent families are at least twice as likely to experience poverty.[5] However,

[3] Material components refer to factors such as health or wealth. Social components include friendship or love.

[4] Material resources are external possessions (e.g., income, educational attainment and employment); personal resources are traits that enable people to cope with stress; and social resources are elements such as social integration and social support that help people achieve their valued outcomes.

[5] Indications for increased labour market risks for single parent families have also been provided by Fritsch et al. (2019), Teitzer et al. (2014), Verwiebe and Fritsch (2011) and Verwiebe et al. (2013).

Table 1:
Predictions from theoretical perspectives and empirical findings on differences in well-being domains between household types

	Single households	Two adults without children	Two adults with children	Single parents
Material well-being				
Financial resources	+	++	−	−−
Housing (incl. living environment)	+	++	−−	−−
Social well-being				
Romantic partnership and family	−	++	++	+
Friends and colleagues	++	+	−	−
Cross-sectional domains				
Leisure (personal autonomy)	+	++	−	−−
Employment (career)	++	+	−	−−

Source: Own illustration.
Note: + signifies advantages/positive effects on well-being, − signifies disadvantages/negative effects on well-being in the respective domain.

it has also been reported that two-adult households without children tend to have considerable financial advantages over single-person households (Verwiebe 2014, p. 297). Across Europe, 'the major share of upper middle-income and upper-income class households comprises non-retired adults without children' (Vaughan-Whitehead et al. 2016, p. 41). The presence of two earners enables the household to benefit from the pooling of financial resources and economies of scale. Financial gains are highest for childless dual-earner households who are not burdened with the need to spend on children (cf. Table 1).

Financial assets are also important for achieving high-quality housing arrangements. Couples without children (who can pool resources), but also people who live alone, are likely to have advantages in the housing market. Both household types are free to move and to change neighbourhoods if the benefits of doing so exceed the costs. As households with children are more dependent on the infrastructure of the neighbourhood (e.g., schools, health facilities), they are less flexible. Thus, the well-being of households with children may be negatively affected by the pressure to relocate (Ballas 2013, p. 44; Bratt 2002, p. 22). While households with children generally require more space, larger accommodations are often unaffordable, or are very distant from city centres (Ballas 2013, p. 45; Brown et al. 2016, p. 117). Furthermore, in Austria, homeowners are not necessarily more satisfied with their housing conditions than tenants (Elsinga and Hoekstra 2005, p. 422f). Overall, the literature suggests that singles have a slight advantage over other households in terms of housing satisfaction (Vera-Toscano and Ateca-Amestoy 2008, p. 262).

Thus, we expect to find that singles have higher levels of material well-being than households with children, and single-parent households in particular; and that households made up of two adults without children who benefit from the pooling of their resources have the highest levels of material well-being (cf. Table 1).

2.2 Differences between singles and other household types in terms of social well-being

Previous research has shown that the well-being of different household types tends to differ considerably because of their varying levels of social embeddedness and their family ties. Single households are characterised by the absence of a permanently present second human being in the home. This absence often means more than the lack of a physical presence, as it can also mean the lack of emotional support. Although not everyone who lives alone is a 'single' in the sense of lacking an intimate relationship (Hertel et al. 2007), it may be assumed that the majority of individuals who live alone are not in a stable, long-term romantic relationship.[6] In all cultures, families have been the fundamental building blocks of social life. Thus, the lack of an intimate relationship is often detrimental to an individual's subjective well-being (Klinenberg 2012). Helliwell (2003), for example, found that single people are, on average, less happy than (married) couples, especially if they had a previous partnership that ended (Ballas and Dorling 2007; Dolan et al. 2008). Similar findings have been reported for the U.S. by Kamp Dush and Amato (2005). They found that married couples reported the highest levels of subjective well-being, followed (in order) by individuals in cohabiting relationships, individuals in steady dating relationships, and individuals who dated infrequently or not at all. The most commonly accepted explanation for the link between well-being and family life is that being part of a family provides people with social support and social integration (Diener et al. 2000; Lee et al. 1991; Mastekaasa 1994; Shapiro and Keyes 2008). Studies indicated that individuals who are embedded in networks of supportive, helpful and loving or caring others tend to have better emotional health and higher levels of well-being. Against this background, we assume that singles and single parents are generally less satisfied with their family life than couples with or without children. However, we also expect to find that single households are less satisfied with their family life than single parents, because having children offers parents the opportunity to form stable, long-lasting and particularly strong emotional bonds (e.g., Huinink 1995; Morgan and King 2001) (cf. Table 1).

[6] In the present paper, we conceptualise singles as individuals who live alone (Chandler et al. 2004; Hall and Ogden 2003; Hertel et al. 2007; Ogden and Hall 2000). Therefore, we cannot a priori preclude the possibility that they have a partner. Nevertheless, in our dataset, less than 30 per cent of individuals living alone were in a relationship, compared to 70 to 80 per cent of individuals in two-adult households (Appendix, Table A.2).

Nevertheless, we should also consider the possibility that social well-being is less determined by an individual's family ties than ever before. Both psychological need theories (Baumeister and Leary 1995) as well as rational choice sociology (Nieboer and Lindenberg 2002) discuss the relevance of 'substitution effects'. According to these theories, a basic aim (e.g., belongingness) can be achieved by different means. Therefore, specific social relations and emotional bonds (e.g., within the family) may be substituted by others (e.g., friendship networks, close colleagues), and these kinds of substitution may go beyond the borders of different areas of private life. Hochschild (2001), for instance, observed that positive experiences in the workplace can compensate for negative experiences at home. Thus, in sum, being single does not necessarily mean that the individual lacks social relationships or contact with others; instead, it could mean that the person has more free time for fostering social contacts beyond the family of origin (Chandler et al. 2004). This leads us to assume that singles have more freedom to invest in non-family contacts than couples, and, in particular, couples with children (cf. Table 1). In contrast to single parents, people who live alone may be better able to compensate for their (potential) disadvantages in terms of their satisfaction with family life by forming social contacts above and beyond the confines of a nuclear family.

2.3 Differences between singles and other household types regarding cross-sectional domains of well-being

Leisure time and employment are life domains that are related to both financial and social well-being. Studies have, for instance, shown that having the time and the financial assets to connect with local communities and friends is a strong predictor of positive well-being (Ballas 2013, p. 44). Having social networks and friendships, but also engaging in solitary forms of leisure (such as educational, relaxational or physiological leisure activities) can enhance people's satisfaction with their leisure time, and, thus, their overall well-being (e.g., Heintzman 2020). Since singles are able to schedule their time independently, and have no immediate child care obligations, they likely are better able to arrange their leisure time around their personal preferences (Ateca-Amestoy et al. 2008; Lee and Bhargava 2004). Therefore, we expect to observe that singles are more satisfied with their leisure time than single parents and couples with children. At the same time, compared to couples, singles often feel greater pressure to actively keep in touch with their social networks (Klinenberg 2012) in order to avoid falling into a 'loner status' (Lawton and Cohen 1974, p. 198), which is characterised by a lack of interaction with both family and peers. Therefore, we expect to find that couples have the advantage of enjoying shared leisure time, which could lead to higher levels of overall well-being (cf. Table 1).

Satisfaction with one's employment is another important facet of subjective well-being. As a main activity, professional work gives life additional meaning, increases self-confidence and offers the opportunity for social and material gratification. In addition, Western societies generally value work orientation, and place considerable importance on employment (Fritsch et al. 2019). Thus, employment is considered a positive factor for life satisfaction (Frey and Stutzer 2002; Layard 2005; Schoon et al. 2005). In contrast, unemployment has been shown to have (long-lasting) negative effects on well-being, even if no income is lost (Berry and Okuliz-Kozaryn 2009, p. 120). Singles and single-parent households are likely to suffer more from the negative consequences of unemployment, since they are both financially self-reliant and socially dependent on contacts outside their own household. However, as singles are able to schedule their time independently and have no immediate child care obligations, they are able to put more emphasis on their careers than coupled individuals, or, in particular, adults with children (Klinenberg 2012). In addition, as singles are socially more dependent on contacts outside of their own household, their social contacts with colleagues may become crucial. Finally, the presence of a partner or of children in the household increases the likelihood of experiencing work-family conflicts, which can lead to lower satisfaction with work and reduced subjective well-being (e.g., Byron 2005; Matthews et al. 2006; Matysiak et al. 2016). Therefore, we assume that single households are, overall, more satisfied with their employment situations than other household types, and particularly compared to households with children (cf. Table 1).

2.4 Considerations regarding the overall well-being of single households

As we mentioned in the introduction, the question of whether the subjective well-being of single households is lower overall is anything but trivial. Our theoretical discussion clearly suggests that living alone has positive as well as negative consequences. Nevertheless, previous research has shown that singles tend to have lower subjective well-being than other types of households (e.g., Blekesaune 2018; Chang 2013; Mastekaasa 1994; Soons and Kalmijn 2009; Vanassche et al. 2013). If the overall life satisfaction of single households is shown to be lower than that of other households, this would, in turn, mean that family life contributes heavily to well-being, and that having advantages in other life domains would not fully compensate for this specific disadvantage.

However, according to our overview (Table 1), the family domain is the only one in which singles are clearly disadvantaged. Compared to each of the other household types we consider, the subjective well-being of single households is expected to be higher in some life domains, but lower in at least one. For instance, we assume that, on average, single households have higher levels of cross-sectional and material well-being than households with co-resident children. Couples without children seem to be best off, whereas single parents appear to face the most difficulties. It follows that we expect to find that the well-being of single households is lower than that of some household types, but is also higher than that of others.

2.5 Implicit assumptions: Domain satisfactions as mediators and the influence of background factors

In the theoretical discussion, we identified a series of factors that may be responsible for the differences in domain satisfactions between single households and other household types. We further assume that these differences in domain satisfactions result in varying levels of overall well-being between different household types. For instance, we expect to find that for singles, their relationship status (i.e., the absence of an intimate partner) leads to lower satisfaction with family life, which, in turn, lowers their overall well-being. Our assumptions include two additional hypotheses. First, we assume throughout the paper that domain satisfactions mediate the association between household type and overall well-being. Second, we expect to observe that two types of background variables are at work: namely, confounders as well as mediators. The mediating role of life domain satisfactions follows directly from our theoretical starting point, and the assumption that a person's satisfaction with his/her individual *life situation* is an overall appraisal that is shaped by his/her levels of satisfaction in different life domains (cf. Diener 2009). The issue of how background factors can function as confounders or mediators requires a brief explanation. Different household types may have different characteristics that lead to advantages in some life domains, and disadvantages in others, which can, in turn, have positive and negative effects on overall life satisfaction. These characteristics may differ between households for at least two reasons. First, the household type may influence a characteristic that is relevant for an individual's well-being (i.e., mediating factors). Second, a background factor may affect both the household type and an individual's subjective well-being (i.e., confounding factors).[7]

Mediators comprise the equivalised household income and housing characteristics (e.g., the square footage each person has available or the living environment). These characteristics are affected by the type of household an individual lives in. As we outlined above, singles generally have less income than two-adult households (Verwiebe 2014, p. 297), but they also have more freedom to choose their living environment than households with children (e.g., Vera-Toscano and Ateca-Amestoy 2008). In turn, people's income and living environment are strong predictors of their well-being, and the lack of both in poor households negatively affects their well-being (Ballas 2013; Easterlin 2005). Age, gender, subjective health and immigration background are important *confounders*. Healthy women from Austrian families who are under age 25 or are over age 55 are the most likely to be living alone in Vienna (Statistik Journal Wien 2016, p. 22). At the same time, studies have shown that women experience higher levels of well-being than men when living

[7] The authors acknowledge that the covariates presented below likely form even more complex multilateral relationships. The presented relationships are a simplified model of these interrelations, and other combinations should be further explored.

alone (Ballas 2013; Hejj 1997); and that among singles, age typically forms a u-shaped relationship with well-being (Ballas 2013; Brown et al. 2016; Clark and Oswald 2006; Fasang et al. 2016). People who are young and healthy might also find it easier to compensate for the lack of family ties with large networks of colleagues and friends (Chandler et al. 2004). Therefore, we expect to find that these factors not only increase the chances of living alone, but increase well-being per se.[8]

In sum, we hypothesise that confounders, mediators and life domain satisfactions explain large shares of the differences in levels of subjective well-being between single households and other household types. By using this approach, we seek to emphasise the complexity and the layered nature of the differences in well-being between household types. However, in our paper, we do not give the background factors centre stage. Instead, we investigate how domain satisfactions contribute to the differences in overall well-being between singles and other households.

3 Data and methods

3.1 Data and variables

Our analysis is based on a unique dataset that resulted from research collaborations undertaken by the University of Vienna and the City of Vienna. Between 1995 and 2018, data on living arrangements, living conditions, family issues, personal relationships, labour market behaviour, quality of life, health and happiness were collected in five cross-sectional surveys. The surveys were conducted by the IFES (www.ifes.at) in 1995 (N = 8,066), 2003 (N = 8,300), 2008 (N = 8,704), 2013 (N = 8,400) and 2018 (N = 8,450). The methods of data collection comprise face-to-face interviews (1995–2008), computer-assisted telephone interviews (2003–2018) and computer-assisted web interviews (2018). The telephone interviews were based on random digit dialing (random last digit) to include mobile phone numbers. Most of the interviews were conducted in the German language. In addition, some immigrants from Turkey and the former Yugoslavia were interviewed in their mother tongue (300 interviews in 2003, 2008 and 2013; 350 in 2018).[9] Random samples have been disproportionally stratified by districts to allow for regional analyses. The data are representative of the residential population of Vienna (aged

---

[8] The classification of some covariates as mediators or confounders is debateable. In principle, a variable like employment status, for instance, could also be both, as it affects the household type, and it is affected by the household type (particularly from a longitudinal perspective). The measure of equivalent household income is, however, directly dependent on household type and household size.

[9] In 2003 and 2008, face-to-face interviews were only used for interviews in foreign languages. In 2018, web interviews accounted for 23 per cent of all interviews.

15 years and older), and cover 11,015 respondents living in single households (between 1,785 and 2,535 respondents in each wave). The IFES prepared (a) design weights that account for the variation in the selection probabilities of households and districts; (b) post-stratification weights that adjust for age by gender, education, district by gender and age and type of housing; and, in 2018, (c) weights that adjust the results for mode effects. We use a combined weight for all analyses presented here.

Based on the literature, we distinguish between dimensions of well-being using satisfaction with different life domains (Campbell et al. 1976; Diener and Fujita 1995; Ormel et al. 1999). The main variables of interest are the respondents' satisfaction with their housing (apartment or house) and their satisfaction with their financial situation as indicators of *material well-being*; the respondents' satisfaction with their family situation and with their social contacts (friends, acquaintances, etc.) as indicators of *social well-being;* and the respondents' satisfaction with their main activity (professional work, education or training, household labour, etc.) and with the time they have for leisure activities (e.g., cultural activities, sport activities, meeting friends)[10] as indicators for *cross-sectional life domains* that affect and/or are being affected by material and social aspects of life. Finally, we also consider the respondents' satisfaction with their individual *life situation* as a whole, while assuming that life satisfaction is an overall appraisal that is shaped by people's satisfaction in different life domains (cf. Diener 2009). Respondents could rate their satisfaction in different domains or with life as a whole on five-point scales, ranging from one (very satisfied) to five (not satisfied at all). In the results section, we present the shares of respondents who were very satisfied or satisfied (values one and two) because these numbers are intuitively understood by readers (and sensitivity analyses with the original rating scales lead to the same conclusions). Descriptive analyses using the whole range of the scale (values from one to five) are shown in Table A.1 in the appendix.

In the following, we label background factors assumed to affect both household type and at least one domain of subjective well-being as confounders; and we label variables assumed to be affected by household type, and to affect subjective well-being, as mediators. Variables assumed to be *confounders* comprise partnership status measured by three variables (a) partner (romantic relationship: no/yes),[11] (b) married (no/yes), (c) divorced or widowed (no/yes), gender (male/female), age (15 to 29, 30 to 44, 45 to 59, 60 to 99), subjective health (very good (1), good (2), moderate (3), worse (four or five on a five-point rating scale)), employment

[10] The wording regarding satisfaction with leisure time varied slightly across survey waves: in 2018, leisure activities explicitly included cultural activities, whereas in 1995, no examples of leisure time activities were mentioned. In 2003, 2008 and 2013, satisfaction with cultural activities was measured using a separate question.

[11] Unfortunately, both the number and the wording of questions referring to the partner changed several times between survey waves. Some of these changes were unavoidable, as, for instance, the option of a registered partnership did not exist in 1995.

status (employed, unemployed, other) and immigrant background (respondent or at least one of his/her parents not born in Austria). The variables assumed to have a mediating function include equivalised household income (quartiles), square meters per person in household (up to 25 m^2, 26 to 35 m^2, 36 to 45 m^2, 46 to 70 m^2, more than 70 m^2) and the quality of the living environment (four categories ranging from 'no impairments' to 'some to major impairments'). The measure for the quality of the living environment is based on an index summarising subjective assessments of the quality of (a) air, (b) water supply, (c) road cleaning and (d) waste disposal. Descriptive statistics regarding confounders and mediators are shown in the appendix (Table A.2).

3.2 Analysing strategy

Our analytical strategy comprises three distinct steps. *First*, we describe the development of the subjective well-being of individuals living in single households over time, and compare it to the average subjective well-being of individuals in other selected living arrangements: namely, two-person households (with two adults), two adults living with at least one child (under 15 years old) and single-parent families (households with one adult and at least one child under 15 years old). This approach will allow us to compare the differences in the structure of well-being across household types. *Second*, we analyse the role of material, social and cross-sectional domains of well-being in the differences in overall life satisfaction between household types. To do so, we run a series of regression models employing hierarchical model building with a stepwise inclusion of domain satisfaction indicators. In binomial logistic regressions, we estimate average marginal effects, as these are the most likely to be comparable across different models (Mood 2010). The KHB method (Karlson et al. 2012) is used to examine whether differences in satisfaction with life domains account for the observed differences in overall life satisfaction between single households and other household types. This procedure enables us to draw conclusions about the mediating function of domain satisfactions. In addition, we use decomposition analysis to assess *how much* the differences in domain satisfactions contribute to the differences in overall life satisfaction between single households and other household types. Thus, these analyses also allow us to draw conclusions about substitution and compensation effects (as proposed by psychological need theories and rational choice authors). As our calculations are based on logistic models, we employ the method proposed by Fairlie (2005) and Jann (2006). *Third*, we examine the role of background factors (i.e., confounders and mediators) in differences between household types in domain satisfactions in further regression models. This approach allows us to assess the robustness of our assumptions described in Section 2.5. Finally, we perform complementary analyses that enable us to explore whether differences in confounders and mediators also contribute to observed differences in life satisfaction between single households and other household types (beyond their meaning for life domain satisfactions).

4 Results

4.1 Differences in subjective well-being across household types, 1995–2018

Table 2 shows the development of the subjective well-being of single households, and compares individuals living in single households with other selected household types, as well as with the total population. Our analysis generated a number of important findings. Between 1995 and 2018, the overall life satisfaction levels of individuals living in single households remained quite stable. In each survey year, 74 to 80 per cent reported that they were satisfied with their life. In line with this general result, satisfaction with life domains was remarkably stable at a rather high level: i.e., at least 79 per cent of respondents in each survey year said they were satisfied with their social contacts, 75 per cent indicated they were satisfied with their leisure activities and 73 per cent said they were satisfied with their family

Table 2:
Differences in the subjective well-being of single households and other selected living arrangements 1995–2018 (% very satisfied or satisfied)

Subjective well-being (SWB)		Single households	Two adults without children	Two adults with children below age 15	Single parents	Total population
		%	Diff.	Diff.	Diff.	Diff.
Overall life satisfaction	1995	79	+10***	+5***	−5*	+3
	2003	79	+11***	+7***	−3	+4
	2008	79	+9***	+7***	−5	+4
	2013	80	+9***	+7***	−1	+3
	2018	74[b]	+11***	+7***	−4	+5
Domains of material WB						
Housing	1995	78	+1	−15***	−14***	−5
	2003	80	+1	−6***	−7**	−1
	2008	79	+5***	+1	−6*	+2
	2013	81[c]	+2*	−2	−3	0
	2018	78	+3*	−4**	−16***	−1
Finances	2003	62	+12***	−1	−13***	+2
	2008	65[c]	+9***	−1	−20***	+2
	2013	65[c]	+7***	+1	−12***	+3
	2018	60	+11***	1	−12***	+4

Continued

Table 2:
Continued

Subjective well-being (SWB)		Single households	Two adults without children	Two adults with children below age 15	Single parents	Total population
		%	Diff.	Diff.	Diff.	Diff.
Domains of social WB						
Family	1995	73	+22***	+19***	−2	+11
	2003	75	+18***	+17***	0	+9
	2008	76[c]	+17***	+17***	−2	+8
	2013	78[a]	+16***	+15***	−5	+7
	2018	73	+16***	+15***	−1	+8
Social contacts	1995	87	+3***	+4***	+4*	+2
	2003	86	+6***	+4**	+2	+2
	2008	84[b]	+6***	+3**	+3	+3
	2013	87	+3***	+2*	−1	+1
	2018	79[a]	+7***	−2	−1	+1
Cross-sectional domains						
Main activity	2003	80	+5***	0	−8***	+1
	2008	80	+6***	+1	−5*	+2
	2013	80	+8***	+6***	−1	+3
	2018	74[c]	+7***	+3*	−2	+3
Leisure time	1995	77	+7***	−8***	−11***	−1
	2003	76	+6***	−3*	−3	+1
	2008	77	+4**	−6***	−11***	0
	2013	78	+5***	−2*	−4*	+1
	2018	75	0	−24***	−16***	−7

Note: Differences have been measured in relation to single households. For instance, the difference of '+10' for two adults without children with regard to life satisfaction in 1995 indicates that the share of satisfied persons was 10 percentage points higher among respondents from two-adult households than among respondents living in single households (89 per cent versus 79 per cent). Tests on differences between the respective multi-person households and single households: $^{***}p \leq .001$, $^{**}p \leq .01$, $^{*}p \leq .05$. Tests on differences between single households' satisfaction in the earliest available survey year (i.e., 1995 or 2003) and in the respective later survey year: $^{a}p \leq .001$, $^{b}p \leq .01$, $^{c}p \leq .05$.
Source: Viennese Quality of Life Surveys (1995–2018); N = 41,920; own weighted calculations.

life. Up to 80 per cent reported that they were satisfied with their housing situation and with their main activity. However, the respondents reported feeling somewhat less satisfied with their financial situation (only 60 to 65 per cent said they were satisfied in each survey year). Overall, our results suggest that over the last two decades, the subjective well-being of single households in Vienna has, on average, been high. *Compared to the total population*, the well-being of single households hardly deviated from the average satisfaction levels in Vienna with regard to separate life domains, or to life in general. The level of life satisfaction was only slightly lower among single households (the share of people who were satisfied was three to five percentage points lower). However, there was one notable exception that is in line with our theoretical considerations (see section 2). While the share of people who were satisfied with their family life was at least 81 per cent among the total population of Vienna, this share was between seven and 11 percentage points lower among single households.

When we compare single households with selected other household types, we find the following: *two-adult households without children* were generally more satisfied than individuals who were living alone (average differences around 10 percentage points), in particular with regard to their financial situation (e.g., +11 percentage points in 2018) and their family life (+16). Likewise, *two-adult households with children* were more satisfied with their family life (e.g. +15 percentage points in 2018), but were less satisfied with the time they had for leisure activities than individuals who were living in single households. The decline in satisfaction with leisure activities (1995: −8 percentage points, 2018: −24 percentage points) in the two-adult households with children was especially remarkable.[12] Finally, we observe that *single parents* were worse off in terms of their satisfaction with their housing, finances and leisure time. Overall, our findings demonstrate that there were household types that were more or less satisfied than single households, and that there were household types that were more satisfied in some dimensions of life, and were less satisfied in others. Nevertheless, it appears that the presence of an adult cohabitant had a strong positive effect on life satisfaction.

4.2 The role of domain satisfactions in the differences in life satisfaction between single households and other household types

Based on our regression models, Table 3 presents results on the relevance of material, social and cross-sectional domains of subjective well-being for differences in life satisfaction between households. Model M1 shows the observed differences between the single households and the other household types. The comparison of coefficients in Model M1 and Model M2 shows that the inclusion of *material well-being* reduces the differences in life satisfaction between single households and two-adult households (2003 and 2018), and between single households and single-parent households (2003 and 2018). For the latter, material well-being is absolutely crucial. In 2018, the difference between single parents and those living

[12] One reason may be that the questionnaire in 2018 explicitly mentioned that leisure time activities include cultural activities, sport and meeting friends; while the 1995 questionnaire referred to the organisation of leisure time.

Table 3:
Differences in overall life satisfaction (% very satisfied or satisfied) between single households and other living arrangements 2003 and 2018 (AMEs)

Overall life satisfaction	M1	M2	M3	M4	M5
Year 2003					
Single households					
Two adults without children	.10***	.08***b	.04***a	.08***d	.04***a
Two adults with children below age 15	.05***	.06***	−.01a	.06***	.02*
Single parents	−.04(*)	.01a	−.04*	.00b	.00b
Other households	.04**	.04***	.00a	.04**	.01
Cragg & Uhler's R^2	2.7	23.1	24.2	27.5	42.3
Year 2018					
Single households					
Two adults without children	.11***	.07***c	.05***a	.10***	.04***a
Two adults with children below age 15	.07***	.07***	.03*c	.09***	.05***
Single parents	−.07*	.01a	−.05(*)	−.02b	.01a
Other households	.01	.01	.00	.03*	.01
Cragg & Uhler's R^2	2.5	21.6	32.7	28.7	49.6
Domains of material WB		incl.			incl.
Domains of social WB			incl.		incl.
Cross-sectional domains				incl.	incl.
Differences between models		lr chi^2	lr chi^2	lr chi^2	lr chi^2
Tests vs. M1: Year 2003	ref.	993***	1051***	1221***	2056***
Year 2018	ref.	1630***	1739***	1486***	2906***

Notes: The table gives average marginal effects (AMEs). Life domains satisfaction measures refer to housing and financial situation (domains of material well-being), family life and social contacts (domains of social well-being), and main activity and leisure time (cross-sectional domains). Tests referring to coefficients: (a) Tests on differences compared to single households: $^{***}p \leq .001$, $^{**}p \leq .01$, $^{*}p \leq .05$, $^{(*)}p \leq .06$; (b) tests on differences compared to M1: $^{a}p \leq .001$, $^{b}p \leq .01$, $^{c}p \leq .05$, $^{d}p \leq .06$.
Source: Viennese Quality of Life Surveys (2003, 2018); $N_{2003} = 7.866$; $N_{2018} = 8.051$; own unweighted calculations.

alone was −.07 in Model M1, but was +.01 in Model M2. This means that the observed share of respondents satisfied with their life was seven percentage points lower among single parents than among single households in 2018. If, however, the two groups had not differed in terms of their material well-being, the difference would have completely disappeared. Material well-being was not as relevant for the life satisfaction differences between single households and two-adult households without children. In 2018, the observed difference in the share of respondents who were satisfied with their life decreased from eleven (M1) to seven percentage points (M2). Comparing the coefficients in Model M1 and Model M3 demonstrates that *social well-being* was highly relevant for differences in life satisfaction between single households and two-adult households (with and without children). The differences between single households and respondents from two-adult households without children were reduced by six percentage points in both 2003 and 2018, and the differences between single households and respondents from two-adult households with children disappeared entirely in 2003, and were reduced by four percentage points in 2018.

The results for the *cross-sectional domains of well-being* (M4) were similar to those for material well-being, and were mainly relevant for the differences between single households and single-parent households. Together, the differences in the material, social and cross-sectional domains of subjective well-being accounted completely for the differences between single households and single-parent households, and partly explained the differences between single households and two-adult households (with or without children) in both 2003 and 2018 (M5).

The results of our decomposition analyses (Table 4) reflect the main results of the regression analyses (Table 3, M1 and M5), and provide some additional insights regarding the role of domain satisfactions in overall life satisfaction. Table 4 again shows that the share of respondents who were satisfied with their life was about 10 (2003) to 11 (2018) percentage points lower among single households than among two-adult households without children. By and large, seven of those 10 to 11 percentage points could be explained by differences in domain satisfactions. The most relevant indicator was *satisfaction with family life,* which accounted for about four percentage points. Thus, lower satisfaction with family life was the main driver of lower life satisfaction among single households. For households with two adults and children under age 15, the results were very similar, but were still different. The share of respondents satisfied with their life was about five to seven percentage points lower among single households than it was among households with two adults and children. Again, lower satisfaction with family life accounted for about four percentage points. However, higher levels of satisfaction among single households with their housing, financial situation, main activity (only 2003) and leisure time (only 2018) helped to offset their disadvantage in satisfaction with family life. Thus, there were at least weak *compensation effects* related to higher levels of satisfaction in material and cross-sectional domains of well-being.

Finally, the differences in life satisfaction levels between single households and single parents were completely explained by differences in domain satisfactions. Regression analyses (Table 3) already showed that material and cross-sectional domains of well-being were decisive for the higher life satisfaction levels among single households. Decomposition analyses (Table 4) now reveal that the respondents' *satisfaction with their financial situation* contributed the most to their differences in life satisfaction. However, the higher levels of satisfaction among single households with their housing, main activity and leisure time also contributed to the difference

Table 4:
Decomposition analyses to explain differences in overall life satisfaction between single households and other household types by differences in dimensions of well-being (%)

Overall life satisfaction	(1) Single households vs. two adults without children		(2) Single households vs. two adults with children below age 15		(3) Single households vs. single parents	
Share satisfied (%)	**2003**	**2018**	**2003**	**2018**	**2003**	**2018**
Single households	80.52	75.56	80.52	75.56	80.52	75.56
Other household type	90.83	86.64	85.91	82.56	76.35	68.49
Difference	−10.31	−11.08	−5.39	−7.00	4.17	7.07
Explained	−6.88	−7.54	−3.18	−2.50	3.91	7.15
Domains of material WB						
Housing	.14***	−.09***	.73***	.75***	.92***	2.20***
Finances	−.79***	−1.21***	.93***	.38***	2.47***	2.64***
Domains of social WB						
Family	−4.01***	−3.88***	−4.30***	−4.03***	−.97***	−.13*
Social contacts	−.95***	−1.47***	−.84***	−.20***	−.52***	.48***
Cross-sectional domains						
Main activity	−.50***	−.91***	.59***	−.24***	1.71***	1.06***
Leisure time	−.79***	.01*	−.28***	.78***	.23***	.91***
$N_{\text{single households}}$	2274	2205	2274	2205	2274	2205
$N_{\text{other household type}}$	2126	2290	1810	1141	427	238

Source: Viennese Quality of Life Surveys (2003, 2018); own unweighted calculations.

(in particular in 2018). Overall, our findings clearly demonstrate that the lower levels of social well-being (mainly due to lower satisfaction with the family situation) among single households was only partly compensated for by their higher levels of satisfaction in material and cross-sectional domains of well-being. Compared to single-parent households and two-adult households with children, the levels of life satisfaction among single households were positively affected by their higher levels of satisfaction with their housing, financial situation and (in particular in 2018) leisure time. The highest levels of life satisfaction were, however, experienced by two-adult households without children. These households enjoyed high levels of satisfaction with their family life, and had no disadvantages in the material or cross-sectional domains of well-being.

4.3 Background factors accounting for the differences in domain satisfaction between single households and other household types

Above, we showed that differences in domain satisfactions were (partly) responsible for differences in life satisfaction. In the next step, we want to disentangle the background factors that are related to differences in the material, social and cross-sectional domains of well-being. The detailed results of Table 5 are in line with our conceptual assumptions. For instance, we found that a higher income increased all sub-dimensions of well-being, except satisfaction with leisure time. Being unemployed was associated with reduced satisfaction with one's main activity and finances. Having a partner was linked to improved satisfaction with one's family life and social contacts. Being married was associated with higher satisfaction with one's family life, leisure time, housing and main activity. Being in bad health and having a low-quality living environment both negatively affected all sub-dimensions of subjective well-being. As single households differed from other household types in one or more of these characteristics (see Table A.2), they explained a huge part of the differences in life domain satisfaction observed between household types.

The results of regression models shown in Table 5 reveal that background factors (i.e., confounders and mediators) entirely accounted for the differences between single households and single parents in their levels of satisfaction with their housing, financial situation and main activity. These material and cross-sectional domains of satisfaction were affected by income, square meters of living space per person and employment status; i.e., factors in which single households and single-parent households clearly differed (Table A.2). Additionally, background factors partly explained the differences between single households and single-parent households in terms of their satisfaction with their leisure time. Most importantly, however, the differences in these characteristics between single households and respondents from two-adult households without children explained significant parts of the differences in their social well-being (i.e., satisfaction with family life and social contacts; cf. Table 5).

4.4 Life satisfaction of households being shaped by domain satisfactions and background factors

Background factors play a role in life satisfaction beyond their significance for domain satisfactions. Taken together, the differences in background factors and domain satisfaction *completely* explained the observed differences in life satisfaction between single households and other household types (Model A5 in Table A.3). The inclusion of confounding variables like relationship status, gender, age, health, employment and immigrant background reduced the differences between single households and two-adult households (2003 and 2018), respondents living with a second adult and children (2018), and single-parent households (2003 and 2018; cf.

Table 5:
Differences in dimensions of subjective well-being (% very satisfied or satisfied) between single households and other household types 2018 (AMEs)

Subjective well-being:	Domains of material WB		Domains of social WB		Cross-sectional domains	
Share of those satisfied with	**Housing**	**Finances**	**Family**	**Social contacts**	**Main activity**	**Leisure time**
Models without background factors						
Single households						
Two adults without children	.04***	.11***	.15***	.06***	.07***	−.01
Two adults with children below age 15	−.04*	.01	.15***	−.02	.03*	−.22***
Single parents	−.15***	−.15***	−.03	−.03	−.04	−.19***
Other households	−.04***	.02	.05***	−.01	.00	−.09***
Models with background factors						
Single households						
Two adults without children below age 15	.06***	.03(*)a	.10***a	.02a	.01a	−.03
Two adults with children	.09***a	.05*c	.13***	−.01	.04(*)	−.13***a
Single parents	.06*a	.04a	.04a	.03a	.06*a	−.07*a
Other households	.10***a	.09***a	.07***	.02c	.04**c	−.04*a
Confounders						
Relationship status:						
Partner (0/1)	−.02	.04**	.06***	.05***	.02	.00
Married (0/1)	.06***	.01	.03*	.02	.04**	.04*
Divorced/widowed (0/1)	−.03*	−.02	−.01	.01	.03*	.03
Gender: female (0/1)	−.01	−.02(*)	.00	.03***	−.01	−.02
Age: 15 to 29 yrs.	.00	.02	.01	.03*	.01	.01
30 to 44 yrs.	ref.	ref.	ref.	ref.	ref.	ref.
45 to 59 yrs.	.03*	.07***	.06***	.05***	.03*	.08***
60 to 99 yrs.	.07***	.17***	.14***	.12***	.14***	.24***
Subjective health: very good	ref.	ref.	ref.	ref.	ref.	ref.
Good	−.04***	−.09***	−.07***	−.09***	−.08***	−.09***
Moderate	−.09***	−.19***	−.16***	−.19***	−.20***	−.17***
Worse	−.13***	−.24***	−.20***	−.2***	−.26***	−.23***
Employment: unemployed	.00	−.11***	−.02	−.02	−.18***	.02
In employment	−.02	−.02	−.01	−.01	.01	−.10***
Others	ref.	ref.	ref.	ref.	ref.	ref.
Immigrant background (0/1)	−.02*	.00	−.04***	−.02(*)	.00	−.02
Mediators						
Equivalised household						
Income: lower 25%	−.06***	−.24***	−.03*	−.03*	−.06***	−.02
26 to 50%	−.04***	−.14***	−.02	−.02	−.02	−.01
51 to 75%	ref.	ref.	ref.	ref.	ref.	ref.
76 to 100%	.03*	.16***	.04*	.05**	.06***	−.03

Continued

Table 5:
Continued

Subjective well-being:	**Domains of material WB**		**Domains of social WB**		**Cross-sectional domains**	
Share of those satisfied with	**Housing**	**Finances**	**Family**	**Social contacts**	**Main activity**	**Leisure time**
Square meters per person:						
Up to 25 m^2	−.10***	.01	.00	−.01	−.01	−.02
26 to 35 m^2	ref.	ref.	ref.	ref.	ref.	ref.
36 to 45 m^2	.05***	.03*	−.02	.00	.01	.00
46 to 70 m^2	.10***	.05***	.01	.01	.02	.03
More than 70 m^2	.16***	.07***	.00	.03	.02	.00
Living environment:						
No impairments	ref.	ref.	ref.	ref.	ref.	ref.
Almost no impairments	−.04***	−.02(*)	−.03***	−.03*	−.02	−.06***
Rarely impairments	−.13***	−.08***	−.07***	−.07***	−.07***	−.11***
Some to major impairments	−.26***	−.17***	−.12***	−.14***	−.13***	−.18***
Cragg & Uhler's R^2	24.8	28.3	20.2	18.4	21.7	18.1

Note: The table gives average marginal effects (AME). ***$p \le .001$, **$p \le .01$, *$p \le .05$, (*)$p \le .06$. Tests on differences compared to the model without confounders and mediators: $^a p \le .001$, $^b p \le .01$, $^c p \le .05$, $^d p \le .06$. The quality of the living environment refers to the absence of problems with quality of air, water supply, road cleaning and waste disposal.
Source: Viennese Quality of Life Survey (2018); N = 8.051; own unweighted calculations.

Model A2 in Table A.3). The mediators – equivalised household income, square meters per person and quality of the living environment – were mainly relevant for the differences in life satisfaction between single households and single-parent households. On average, the single-parent households had lower household income and less living space per person, and were more likely to be living in an area with lower environmental quality. Our results even suggest that single parents would have been more satisfied with their life than single households in 2018 if there had been no differences in these characteristics between the two groups (Models A3 and A4).

Finally, our findings clearly support the hypothesis that domain satisfactions function as mediators. Table 5 shows associations between household types and domain satisfactions. Moreover, Table 4 displays the influence of higher levels of satisfaction with social, material and cross-sectional domains of well-being on life satisfaction. The results of KHB tests support the mediating assumption (Tables 3 and A.3). However, when we look at the background factors, we see that being divorced or widowed (only in 2003), being aged 60+ (only in 2018), health, employment status, household income and the quality of the living environment (all 2003 and 2018) had effects on life satisfaction that were only partly mediated by life domain satisfactions (results from Model A5 in Table A.3).

5 Discussion

Individualisation and urbanisation have led to substantial changes in household structures (Bengtson 2001; Hall et al. 1997). Living alone has become a common phenomenon, especially in growing metropolitan regions. Nevertheless, research on the subjective well-being of urban single households is still sparse. The present paper sought to address this research gap by examining the dynamics in the quality of life of single households in Vienna over a quarter of a century (1995–2018). Based on a unique dataset, we obtained a number of important findings. *First*, we showed that four in five individuals living in single households were satisfied with their lives. Moreover, this high level of average satisfaction among single households was very stable over time. Thus, we found that singles were (almost) as satisfied with life as the average population of Vienna. Assessing whether single households were worse off than other household types was not straightforward. Apparently, the comparison standard matters: the life satisfaction of two-adult households (with and without children) was significantly higher than that of single households, whereas the life satisfaction of single parents was not. *Second*, our findings demonstrated that satisfaction with family life was considerably lower in single households than in the average population. Decomposition analyses showed that differences in satisfaction with family life were the main reason for the lower levels of life satisfaction among single households than among two-adult households with and without children. Thus, our findings support previous research that has attributed the lower subjective well-being observed among singles to missing social support, a lack of social integration and an unsatisfied feeling of belonging (Diener et al. 2000; Lee et al. 1991; Mastekaasa 1994; Shapiro and Keyes 2008).

Third, we found that among singles, their higher levels of satisfaction in other life domains partly compensated for their lower satisfaction with family life. Our results showed, for instance, that for single households, their higher levels of satisfaction with their housing, financial situation, main activity and leisure time helped to offset their disadvantages relative to two-adult households with children. These findings imply that living alone has advantages and disadvantages (although two-adult households without children seem to have the highest levels of subjective well-being). These results are also in line with theories that refer to substitution effects (Baumeister and Leary 1995; Nieboer and Lindenberg 2002). *Fourth*, our findings showed that the lower satisfaction with family life found among singles was not always decisive. Single parents who reported a comparable level of satisfaction with family life had, on average, lower levels of life satisfaction than single households because of their lower levels of material well-being. The factor that was most relevant for the finding that single-parent households had lower life satisfaction than single households was that these respondents were less satisfied with their financial situation. In addition, single-parent households were less satisfied with their housing and leisure time. These results are not surprising given that single parents are known to face many difficulties (Brady and Burroway 2012; Damaske et al. 2017; Riederer et al. 2017). Nevertheless, together with the evidence for compensation effects, these findings go beyond the usual focus on family issues in the academic discussion of subjective well-being among singles. Thus, these results enrich our understanding of the differences in well-being between household types.

In a nutshell, our results confirmed that a variety of factors accounted for the differences we observed between household types in levels of life satisfaction. These differences could be completely explained by domain satisfaction and background factors, and the differences in domain satisfaction could also be largely explained by the latter. These findings suggest that the household type per se was not decisive for subjective well-being. More important were factors that led to living alone (e.g., not having a partner) or factors that may have followed from being a single parent (e.g., economic hardship). There were always reasons for the advantages or disadvantages in well-being we found for a specific household type. While our study has taken a first step towards providing a more detailed picture of single households, future studies are needed that focus on other household types, and that analyse them more systematically.

Finally, our study has several limitations. The primary limitation is that we did not distinguish between singles (people without a partner) and one-person households and between two-adult households and couple households.[13] The reason for this issue was that both the number and the wording of questions referring to the partner changed between survey waves several times. Thus, we decided to use broader categories of household types that were comparable over time. In addition, the single households were very heterogeneous in several other respects (gender, age, income etc.). Thus, our next step will be to perform an analysis of differences within the group of single households in Vienna.

References

Allardt, E. 1976. Dimensions of welfare in a comparative Scandinavian study. *Acta Sociologica* 19(3): 227–239. https://doi.org/10.1177/000169937601900302

Ateca-Amestoy, V., R. Serrano-del-Rosal and E. Vera-Toscano 2008. The leisure experience. *The Journal of Socio-Economics* 37(1): 64–78. https://doi.org/10.1016/j.socec.2006.12.025

Ballas, D. 2013. What makes a 'happy city'? *Cities* 32(S1): S39–S50. https://doi.org/10.1016/j.cities.2013.04.009

Ballas, D. and D. Dorling 2007. Measuring the impact of major life events upon happiness. *International Journal of Epidemiology* 36(6): 1244–1252. https://doi.org/10.1093/ije/dym182

[13] Besides couples, two-adult households also comprise shared flats or parent-child households. In 2018, for instance, about 70 to 80 per cent of two-adult households (without/with children) were couple households (cf. Table A.2).

Baumeister, R. F. and M. R. Leary 1995. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin* 117(3): 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Bengtson, V. 2001. Beyond the nuclear family: The increasing importance of multigenerational bonds. *Journal of Marriage and Family* 63(1): 1–16. https://doi.org/10.1111/j.1741-3737.2001.00001.x

Bennett, J. and M. Dixon 2006. *Single person households and social policy: Looking forwards*. York: Joseph Rowntree Foundation.

Berry, B. and A. Okuliz-Kozaryn 2009. Dissatisfaction with city life: A new look at some old questions. *Cities* 26(3): 117–124. https://doi.org/10.1016/j.cities.2009.01.005

Bertram, H. 1993. Die Stadt, das Individuum und das Verschwinden der Familie. Berlin: Humboldt-Universität zu Berlin. https://doi.org/10.18452/1538

Blekesaune, M. 2018. Is cohabitation as good as marriage for people's subjective well-being? *Journal of Happiness Studies* 19(2): 505–520. https://doi.org/10.1007/s10902-016-9834-x

Bouzarovski, S., A. Haase, R. Hall, A. Steinführer, S. Kabisch and P. Ogden 2010. Household structure, migration trends, and residential preferences in inner-city León, Spain: Unpacking the demographies of reurbanization. *Urban Geography* 31(2): 211–235. https://doi.org/10.2747/0272-3638.31.2.211

Brady, D. and R. Burroway 2012. Targeting, universalism, and single-mother poverty: A multilevel analysis across 18 affluent democracies. *Demography* 49(2): 719–746. https://doi.org/10.1007/s13524-012-0094-z

Bratt, R. 2002. Housing and family well-being. *Housing Studies* 17(1): 13–26. https://doi.org/10.1080/02673030120105857

Brown, Z., W. Oueslati and J. Silva 2016. Links between urban structure and life satisfaction in a cross-section of OECD metro areas. *Ecological Economics* 129: 112–121. https://doi.org/10.1016/j.ecolecon.2016.05.004

Buzar, S., P. Ogden and R. Hall 2005. Households matter: The quiet demography of urban transformation. *Progress in Human Geography* 29(4): 413–436. https://doi.org/10.1191/0309132505ph558oa

Byron, K. 2005. A meta-analytic review of work-family conflict and its antecedents. *Journal of Vocational Behavior* 67(2): 169–198. https://doi.org/10.1016/j.jvb.2004.08.009

Campbell, A., P. Converse and W. Rogers 1976. *The quality of american life: Perceptions, evaluations, and satisfactions*. New York: Russell Sage Foundation.

Chandler, J., M. Williams, M. Maconachie, T. Collett and B. Dodgeon 2004. Living alone: Its place in household formation and change. *Sociological Research Online* 9(3): 1–13. https://doi.org/10.5153/sro.971

Chang, W.-C. 2013. Family ties, living arrangement, and marital satisfaction. *Journal of Happiness Studies* 14(1): 215–233. https://doi.org/10.1007/s10902-012-9325-7

Cherlin, A. 2004. The deinstitutionalization of American marriage. *Journal of Marriage and Family* 66(4): 848–861. https://doi.org/10.1111/j.0022-2445.2004.00058.x

Clark, A. and A. Oswald 2006. The curved relationship between subjective well-being and age. Paris: HAL Archive. Retrieved from https://halshs.archives-ouvertes.fr/halshs-00590404/

Damaske, S., J. Bratter and A. Frech 2017. Single mother families and employment, race, and poverty in changing economic times. *Social Science Research* 62: 120–133. https://doi.org/10.1016/j.ssresearch.2016.08.008

Diener, E. 2009. *Assessing well-being*. Netherlands: Springer. https://doi.org/10.1007/978-90-481-2354-4

Diener, E. and Fujita 1995. Resources, personal strivings, and subjective well-being: A nomothetic and idiographic approach. *Journal of Personality and Social Psychology* 68(5): 926–935. https://doi.org/10.1037/0022-3514.68.5.926

Diener, E., C. Gohm, E. Suh and S. Oishi 2000. Similarity of the relations between marital status and subjective well-being across cultures. *Journal of Cross-Cultural Psychology* 31(4): 419–436. https://doi.org/10.1177/0022022100031004001

Diener, E. and E. M. Suh 1997. Measuring quality of life: Economic, social, and subjective indicators. *Social Indicators Research* 40: 189–216. https://doi.org/10.1023/A:1006859511756

Dolan, P., T. Peasgood and M. White 2008. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology* 29(1): 94–122. https://doi.org/10.1016/j.joep.2007.09.001

Easterlin, R. 2005. A puzzle for adaptive theory. *Journal of Economic Behavior and Organization* 56(4): 513–521. https://doi.org/10.1016/j.jebo.2004.03.003

Elsinga, M. and J. Hoekstra 2005. Homeownership and housing satisfaction. *Journal of Housing and the Built Environment* 20: 401–424. https://doi.org/10.1007/s10901-005-9023-4

Eurostat. 2019. Living conditions - cities and greater cities. Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=urb_clivcon&lang=en, accessed on 23 May 2019

Fairlie, R. W. 2005. An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement* 30(4): 305–316. https://doi.org/10.3233/JEM-2005-0259

Fasang, A. E., J. Huinink and M. Pollmann-Schult 2016. Aktuelle Entwicklungen in der deutschen Familiensoziologie: Theorien, Daten, Methoden. *Zeitschrift für Familienforschung* 28(1): 112–143. https://doi.org/10.3224/zff.v28i1.22923

Frey, B. and A. Stutzer 2002. *Happiness and economics: How the economy and institutions affect human well-being*. Oxford: Princeton University Press.

Friedlander, D., B. Okun and S. Segal 1999. The demographic transition then and now: processes, perspectives and analyses. *Journal of Family History* 24(4): 493–533. https://doi.org/10.1177/036319909902400406

Fritsch, N.-S., R. Verwiebe and C. Liebhart 2019. Arbeit und Berufe in Österreich: Veränderte Einstellungsmuster im Kontext eines tief greifenden Strukturwandels. In *Sozialstruktur und Wertewandel in Österreich*, eds J. Bacher, A. Grausgruber, M. Haller, F. Höllinger, D. Prandner and R. Verwiebe, 333–385. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-658-21081-6_15

Fritsch, N.-S., R. Verwiebe and B. Liedl 2019. Declining gender differences in low-wage employment in Germany, Austria and Switzerland. *Comparative Sociology* 18(4): 449–488. https://doi.org/10.1163/15691330-12341507

Ge, J. and K. Hokao 2006. Research on residential lifestyles in Japanese cities from the viewpoints of residential preference, residential choice and residential satisfaction. *Landscape and Urban Planing* 78(3): 165–178. https://doi.org/10.1016/j.landurbplan.2005.07.004

Hall, R. and P. Ogden 2003. The rise of living alone in Inner London: Trends among the population of working age. *Environment and Planning* 35(5): 871–888. https://doi.org/10.1068/a3549

Hall, R., P. Ogden and C. Hill 1997. The pattern and structure of one-person households in England and Wales and France. *International Journal of Population Geography* 3(2): 161–181. https://doi.org/10.1002/(SICI)1099-1220(199706)3:2<161::AID-IJPG64>3.0.CO;2-2

Haller, M. and M. Hadler 2006. How social relations and structures can produce happiness and unhappiness: An international comparative analysis. *Social Indicators Research* 75(2): 169–216. https://doi.org/10.1007/s11205-004-6297-y

Heintzman, P. 2020. Empirical research on leisure and spiritual well-being: Conceptualisation, measurement and findings. *Leisure Studies* 39(1): 146–155. https://doi.org/10.1080/02614367.2019.1684981

Hejj, A. 1997. Jung, ledig sucht: die Welt der Singles. *Journal of Family Research* 9(2): 26–47. Retrieved from https://nbn-resolving.org/urn:nbn:de:0168-ssoar-291674

Helliwell, J. 2003. How's life? Combining individual and national variables explain subjective well-being. *Economic Modelling* 20(2): 331–360. https://doi.org/10.1016/S0264-9993(02)00057-3

Hertel, J., A. Schütz, B. M. DePaulo, W. L. Morris and T. S. Stucke 2007. She's single, so what? *Journal of Family Research* 19(2): 139–158. https://doi.org/10.20377/jfr-301

Hettlage, R. 2000. Individualization, pluralization, post-familiarization: dramatic or dramatized upheavals in the modernization process in the family? *Journal of Family Research* 12(1): 72–97. Retrieved from https://nbn-resolving.org/urn:nbn:de:0168-ssoar-291123

Hochschild, A. R. 2001. *The time bind. When work becomes home and home becomes work* 2nd ed., New York: Holt Paperbacks.

Hogan, M., K. Leyden, R. Conway, A. Goldberg, D. Walsh and P. McKenna-Plumley 2016. Happiness and health across the lifespan in five major cities: The impact of place and government performance. *Social Science & Medicine* 162: 168–176. https://doi.org/10.1016/j.socscimed.2016.06.030

Hradil, S. 1995. *Die Single-Gesellschaft*. München: C.H. Beck.

Huinink. 1995. *Warum noch Familie? Zur Attraktivität von Partnerschaft und Elternschaft in unserer Gesellschaft*. Frankfurt/Main: Campus.

Jann, B. 2006. Stata module to generate nonlinear decomposition of binary outcome differentials. *Boston College Department of Economics: Statistical Software Components*. Retrieved from http://ideas.repec.org/c/boc/bocode/s456727.html

Kahn, M. 1995. A revealed preference approach to ranking city quality of life. *Journal of Urban Economics* 38(2): 221–235. https://doi.org/10.1006/juec.1995.1030

Kamp Dush, C. and P. Amato 2005. Consequences of relationship status and quality for subjective well-being. *Journal of Social and Personal Relationships* 22(5): 607–627. https://doi.org/10.1177/0265407505056438

Karlson, K. B., A. Holm and R. Breen 2012. Comparing regression coefficients between same-sample nested models using logit and probit: A new method. *Sociological Methodology* 42(1): 286–313. https://doi.org/10.1177/0081175012444861

Klinenberg, E. 2012. *Going solo. The extraordinary rise and surprising appeal of living alone*. New York: The Penguin Press.

Krekel, C., J. Kolbe and H. Wüstemann 2016. The greener, the happier? The effect of urban land use on residential well-being. *Ecological Economics* 121: 117–127. https://doi.org/10.1016/j.ecolecon.2015.11.005

Kuijsten, A. 1995. Recent trends in household and family structures in Europe: An overview. In *Household demography and household modelling*, eds E. Van Imhoff, A. Kuijsten, P. Hooimeijer and L. Van Wissen, 53–84. New York and London: Plenum Press. https://doi.org/10.1007/978-1-4757-5424-7_3

Lawton, P. and J. Cohen 1974. The generality of housing impact on the well-being of older people. *Journal of Gerontology* 29(2): 194–204. https://doi.org/10.1093/geronj/29.2.194

Layard, R. 2005. *Happiness: Lessons from a new science*. New York, London: Penguin Books.

Lee, G., K. Seccombe and C. Sheehan 1991. Marital status and personal happiness: An anlysis of trend data. *Journal of Marriage and the Family* 53(4): 839–844. https://doi.org/10.2307/352991

Lee, Y. and V. Bhargava 2004. Leisure time: Do married and single individuals spend it differently? *Journal of Family and Consumer Sciences* 32(2): 254–274. https://doi.org/10.1177/1077727X03261631

Lesthaeghe, R. 2010. The unfolding story of the second demographic transition. *Population and Development Review* 36(2): 211–251. https://doi.org/10.1111/j.1728-4457.2010.00328.x

Luhmann, M. and L. Hawkley 2016. Age differences in loneliness from late adolescence to oldest old age. *Developmental Psychology* 52(6): 943–959. https://doi.org/10.1037/dev0000117

Maldonado, L. and R. Nieuwenhuis 2015. Family policies and single parent poverty in 18 OECD countries, 1978-2008. *Community, Work & Family* 18(4): 395–415. https://doi.org/10.1080/13668803.2015.1080661

Marans, R. 2012. Quality of urban life studies: An overview and implications for environment-behaviour research. *Procedia - Social and Behavioral Sciences* 35: 9–22. https://doi.org/10.1016/j.sbspro.2012.02.058

Marans, R. and R. Stimson (eds). 2011. *Investigating quality of urban life*. Dodrecht: Springer. https://doi.org/10.1007/978-94-007-1742-8

Mastekaasa, A. 1994. Marital status, distress, and well-being: An international comparison. *Journal of Comparative Family Studies* 25(2): 183–205. https://doi.org/10.3138/jcfs.25.2.183

Matthews, R. A., R. E. Del Priore, L. K. Acitelli and J. L. Barnes-Farrell 2006. Work-to-relationship conflict: Crossover effects in dual-earner couples. *Journal of Occupational Health Psychology* 11(3): 228–240. https://doi.org/10.1037/1076-8998.11.3.228

Matysiak, A., L. Mencarini and D. Vignoli 2016. Work-family conflict moderates the relationship between childbearing and subjective well-being. *European Journal of Population* 32(3): 355–379. https://doi.org/10.1007/s10680-016-9390-4

Mood, C. 2010. Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review* 26(1): 67–82. https://doi.org/10.1093/esr/jcp006

Morgan, S. P. and R. B. King 2001. Why have children in the 21st century? Biological predisposition, social coercion, rational choice. *European Journal of Population* 17(1): 3–20. https://doi.org/10.1023/A:1010784028474

Nieboer, A. and S. Lindenberg 2002. Substitution, buffers and subjective well-being: A hierarchical approach. In *The universality of subjective well-being indicators*, eds E. Gullone and R. A. Cummins, 175–189. Dordrecht: Kluwer Academic Publishers. https://doi.org//10.1007/978-94-010-0271-4_10

Ogden, P. and R. Hall 2000. Households, reurbanisation and the rise of living alone in the principal french cities, 1975-90. *Urban Studies* 37(2): 367–390. https://doi.org/10.1080/0042098002230

Okulicz-Kozaryn, A. 2013. City life: Rankings (livability) versus perceptions (statisfaction). *Social Indicators Research* 110(2): 433-451. https://doi.org/10.1007/s11205-011-9939-x

Ormel, J., S. Lindenberg, N. Steverink and L. Verbrugge 1999. Subjective well-being and social production functions. *Social Indicators Research* 46(1): 61–90. https://doi.org/10.1023/A:1006907811502

Poortman, A.-R. and A. Liefbroer 2010. Singles' relational attitudes in a time of individualization. *Social Science Research* 39(6): 938–949. https://doi.org/10.1016/j.ssresearch.2010.03.012

Riederer, B. 2018. *Elternschaft und Wohlbefinden*. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-22561-2

Riederer, B., M. Mynarska, M. Winkler-Dworak, T. Fent, B. Rengs and D. Philipov 2017. *Vulnerability and the future of families with children in Europe. Nine questions and corresponding answers.* Vienna: Austrian Academy of Sciences. Retrieved from: https://www.oeaw.ac.at/fileadmin/subsites/Institute/VID/PDF/Publications/Schriftenreihe/VID_Schriftenreihe_23.pdf.

Salvatori, A. and T. Manfredi 2019. *Job polarisation and the middle class: New evidence on the changing relationship between skill levels and household income levels from 18 OECD countries*. Paris: OECD Social, Employment and Migration Working Papers N. 232. https://doi.org/10.1787/1815199X

Schoon, I., L. Hansson and K. Salmela-Aro 2005. Combining work and family life. *European Psychologist* 10(2005): 309–319. https://doi.org/10.1027/1016-9040.10.4.309

Shapiro, A. and C. L. Keyes 2008. Marital status and social well-being: Are the married always better off? *Social Indicators Research* 88(2): 329–346. https://doi.org/10.1007/s11205-007-9194-3

Soons, J. and M. Kalmijn 2009. Is marriage more than cohabitation? Well-being differences in 30 European Countries. *Journal of Marriage and Family* 71(5): 1141–1157. https://doi.org/10.1111/j.1741-3737.2009.00660.x

Soons, J. and A. Liefbroer 2008. Together is better? Effects of relationship status and resources on young adults' well-being. *Journal of Social and Personal Relationships* 25(4): 603–624. https://doi.org/10.1177/0265407508093789

Statistics Austria 2020. Lebensformen 1971–2019. Retrieved from https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/haushalte_familien_lebensformen/lebensformen/index.html, accessed on 21 April 2020.

Statistik Journal Wien 2016. *Entwicklung der Privathaushalte in Wien und den 23 Wiener Gemeindebezirke*. Wien: Statistik Austria.

Teitzer, R., N.-S. Fritsch and R. Verwiebe 2014. Arbeitsmarktflexibilisierung und Niedriglohnbeschäftigung: Deutschland und Österreich im Vergleich. *WSI-Mitteilungen* 67(4): 257–266. https://doi.org/10.5771/0342-300X-2014-4-257

Tov, W. and E. Diener 2007. Culture and subjective well-being. In *Handbook of cultural psychology*, eds S. Kitayama and D. Cohen, 691–713. New York: Guilford Press. https://doi.org/10.1007/978-90-481-2352-0_2

Van de Kaa, D. J. 1987. Europe's second demographic transition. *Population Bulletin* 42(1): 1–59.

Vanassche, S., G. Swicegood and K. Matthijs 2013. Marriage and children as a key to happiness? Cross-national differences in the effects of marital status and children on well-being. *Journal of Happiness Studies* 14(2): 501–524. https://doi.org/10.1007/s10902-012-9340-8

Vaughan-Whitehead, D., R. Vazquez-Alvarez and N. Maitre 2016. Is the world of work behind middle-class erosion? In *Europe's disappearing middle class*, ed D. Vaugan-Whitehead, 1–61. Cheltenham: Edward Elgar Publishing.

Veenhoven, R. 2000. The four qualities of life. *Journal of Happiness Studies* 1(1): 1–39. https://doi.org/10.1023/A:1010072010360

Veenhoven, R. 2007. Quality-of-life research. In *21st century sociology. A reference handbook*, eds C. D. Bryant and D. L. Peck, 54–62. Thousand Oak: Sage.

Vera-Toscano, E. and V. Ateca-Amestoy 2008. The relevance of social interactions on housing satisfaction. *Social Indicators Research* 86: 257–274. https://doi.org/10.1007/s11205-007-9107-5

Verwiebe, R. 2014. Die Auflösung der Mittelschicht und wachsende Armut in Deutschland. *Sozialer Fortschritt* 63(12): 293–300.

Verwiebe, R. and N.-S. Fritsch 2011. Working Poor in Österreich. In *Armut in Österreich*, ed R. Verwiebe, 149–167. Vienna: Braumüller Verlag.

Verwiebe, R., T. Troger, L. Wiesböck, R. Teitzer and N.-S. Fritsch 2013. *GINI Country Report: Growing inequalities and their impacts in Austria*. Amsterdam: AIAS Amsterdam Insitute for Advanced Labour Studies. Retrieved from https://EconPapers.repec.org/RePEc:aia:ginicr:austria

Weziak-Bialowolska, D. 2016. Quality of life in cities - Empirical evidence in comparative European perspective. *Cities* 58: 87–96. https://doi.org/10.1016/j.cities.2016.05.016

Zartler, U., M. Beham, I. Kromer, H. Leitgöb, C. Weber and P. Friedl 2011. *Alleinerziehende in Österreich: Lebensbedingungen und Armutsrisiken*. Wien: ÖGB-Verlag/BMASK. Retrieved from https://nbn-resolving.org/urn:nbn:de:0168-ssoar-412644

Zenker, S., S. Petersen and A. Aholt 2013. The Citizen Satisfaction Index (CSI): Evidence for a four basic factor model in a German sample. *Cities* 31: 156–164. https://doi.org/10.1016/j.cities.2012.02.006

Appendix

Table A.1:
Subjective well-being among single households and the total population 1995–2018 (mean satisfaction and % very satisfied or satisfied)

		Single households		Population	
Subjective well-being (SWB)		**m**	**%**	**m**	**%**
Overall life satisfaction	1995	1.93	78	1.78	83
	2003	1.96	79	1.83	84
	2008	1.94	79	1.79	84
	2013	1.90	80	1.77	85
	2018	2.01	74	1.88	79
Domains of material WB					
Housing	1995	1.93	78	2.07	72
	2003	1.86	80	1.89	79
	2008	1.83	78	1.77	82
	2013	1.81	81	1.79	81
	2018	1.84	78	1.88	76
Finances	2003	2.40	62	2.31	64
	2008	2.24	65	2.16	67
	2013	2.26	65	2.14	68
	2018	2.38	60	2.26	64
Domains of social WB					
Family	1995	1.96	73	1.59	86
	2003	2.00	75	1.68	86
	2008	1.94	76	1.61	87
	2013	1.83	78	1.56	87
	2018	2.03	73	1.75	81
Social contacts	1995	1.65	87	1.55	90
	2003	1.73	86	1.63	89
	2008	1.70	84	1.58	88
	2013	1.61	87	1.55	89
	2018	1.84	79	1.77	81

Continued

Table A.1:
Continued

Subjective well-being (SWB)		Single households m	Single households %	Population m	Population %
Cross-sectional domains					
Main activity	2003	1.91	80	1.86	81
	2008	1.86	80	1.81	82
	2013	1.84	80	1.76	84
	2018	2.00	74	1.95	77
Leisure time	1995	1.93	77	1.91	76
	2003	1.99	76	1.97	77
	2008	1.97	77	1.92	77
	2013	1.87	78	1.88	79
	2018	1.93	75	2.11	68
N (min–max)		1,685–2,533		7,933–8,693	

Note: Satisfaction scales range from 1 (very satisfied) to 5 (not satisfied at all).
Source: Viennese Quality of Life Surveys (1995–2018); *N* = 41,920; own weighted calculations.

Table A.2:
Characteristics of persons in single households and in other household types 2018 (%)

Living arrangement:	Single households	Two adults without children	Two adults with children below age 15	Single parents	Other households	Total sample
Confounders						
Relationship status:						
Partner	28	67	80	40	51	53
Married	3	62	75	15	30	38
Divorced/widowed	40	6	3	39	15	18
Gender: female	58	54	53	76	53	55
Age:						
15 to 29 yrs.	15	15	11	18	48	24
30 to 44 yrs.	22	20	60	51	13	25
45 to 59 yrs.	27	29	27	30	28	28
60 to 99 yrs.	36	36	2	1	11	23

Continued

Table A.2:
Continued

Living arrangement:	Single households	Two adults without children	Two adults with children below age 15	Single parents	Other households	Total sample
Subjective health:						
Very good	25	29	36	31	33	30
Good	45	50	45	40	42	46
Moderate	24	18	17	23	20	20
Worse	6	3	2	7	5	5
Employment:						
Unemployed	7	3	5	13	7	6
In employment	51	57	80	70	51	57
Other	42	40	15	17	42	37
Immigrant background	27	26	42	42	49	35
Mediators						
Equivalised household	28	11	25	49	39	26
Income: lower 25%						
26 to 50%	24	17	30	31	26	24
51 to 75%	26	32	25	13	20	26
76 to 100%	22	39	19	7	15	24
Square meters	1	10	52	38	35	21
Per household						
Member: up to 25 m^2						
26 to 35 m^2	8	27	33	33	32	24
36 to 45 m^2	14	28	9	18	18	18
46 to 70 m^2	42	27	5	8	11	24
More than 70 m^2	35	7	1	2	4	13
Living environment:	36	34	33	34	32	34
No impairments						
Almost no impairments	33	33	33	27	30	32
Rarely impairments	18	20	21	18	20	19
Some to major impairments	13	13	14	21	17	15

Note: The quality of the living environment refers to the absence of problems with quality of air, water supply, road cleaning and waste disposal.
Source: Viennese Quality of Life Survey (2018); N = 8.051; own unweighted calculations.

Table A.3:
Differences in overall life satisfaction (% very satisfied or satisfied) of persons from single households and other household types 2003 and 2018 (AMEs)

Overall life satisfaction	A1 (= M1)	A2	A3	A4	A5
Year 2003					
Single households					
Two adults without children	.10***	.07^{***d}	.09^{***c}	.05^{**b}	.02^{a}
Two adults with children below age 15	.05***	.04*	.09^{***a}	.05**	.01
Single parents	−.04$^{(*)}$	−.02^{d}	.01^{a}	.01^{a}	.00^{c}
Other households	.04**	.03$^{(*)}$	.06***	.03	.00
Cragg & Uhler's R^2	2.7	17.6	11.7	22.3	45.2
Year 2018					
Single households					
Two adults without children	.11***	.04^{***a}	.11***	.05^{***a}	.01^{a}
Two adults with children below age 15	.07***	.03$^{(*)c}$	.13^{***a}	.06***	.02^{d}
Single parents	−.07*	−.01^{a}	.05$^{(*)a}$	.04^{a}	.01^{b}
Other households	.01	.02*	.10^{***a}	.05^{***c}	.01
Cragg & Uhler's R^2	2.5	26.9	19.0	33.5	54.8
Confounders		incl.		incl.	incl.
Mediators			incl.	incl.	incl.
Domains of material WB					incl.
Domains of social WB					incl.
Cross-sectional domains					incl.
Differences between models		lr chi^2	lr chi^2	lr chi^2	lr chi^2
Tests vs. A1: Year 2003	ref.	711***	423***		
Year 2018	ref.	1375***	901***		
Tests vs. A2: Year 2003		ref.		239***	
Year 2018		ref.		414***	
Tests vs. A3: Year 2003			ref.	528***	
Year 2018			ref.	887***	
Tests vs. A4: Year 2003				ref.	1280***
Year 2018				ref.	1513***

Note: The table gives average marginal effects (AMEs). Confounders comprise the following characteristics: gender (male/female), partnership (no/yes), married (no/yes), divorced/widowed (no/yes), age (4 categories), subjective health (4 categories), unemployed (no/yes), in employment (no/yes), and immigrant background (no/yes). Mediators include household income (4 categories), square meters per person in household (5 categories), quality of living environment (index incl. quality of air, water supply, road cleaning, waste disposal; 4 categories). Life domains satisfaction measures refer to housing and financial situation (domains of material well-being), family life and social contacts (domains of social well-being), and main activity and leisure time (cross-sectional domains). Tests referring to coefficients: (a) Tests on differences compared to single households: $^{***}p \leq .001$, $^{**}p \leq .01$, $^{*}p \leq .05$, $^{(*)}p \leq .06$; (b) tests on differences compared to A1: $^{a}p \leq .001$, $^{b}p \leq .01$, $^{c}p \leq .05$, $^{d}p \leq .06$.
Source: Viennese Quality of Life Surveys (2003, 2018); $N_{2003} = 7.866$; $N_{2018} = 8.051$; own unweighted calculations.

Fathers' and mothers' enjoyment of childcare: The role of multitasking

***Marina Zannella*[1,*] *and Alessandra De Rose*[1]**

Abstract

Using data from the latest edition of the Italian Time Use Survey (ITUS, 2013–2014), we analyse 31,309 childcare episodes to investigate the relationship between multitasking (i.e., the combination of childcare with housework tasks) and parents' enjoyment of the time they spent on childcare, with a gender perspective. To this end, we rely on information from the episode enjoyment scores the respondents used to evaluate the degree of (un)pleasantness associated with the different activities they recorded in a daily diary. These episode enjoyment scores are a novelty in the ITUS, and provide a unique measure of the respondents' momentary assessments of their subjective well-being. Our results highlight the existence of a negative relationship between multitasking and parental well-being when spending time on childcare for both mothers and fathers, regardless of the nature of the childcare activity they were performing (i.e., routine or recreational childcare). Our findings add to prior research by shedding new light on the role of multitasking as a relevant contextual characteristic of care that affects the well-being of fathers, as well as of mothers.

Keywords: time use; childcare; well-being; parents

1 Introduction

The rise of dual-earner couples together with the emergence of more time-demanding jobs (e.g., non-standard working hours) in recent decades are among the factors that have contributed to the time squeeze reported by individuals and by families. Nevertheless, existing studies have shown that in industrialised countries, parents are devoting more time to caring for their children today than they did several decades ago (Gauthier et al. 2004; García-Mainar et al. 2011; Faircloth

[1]Department of Methods and Models for Economics, Territory and Finance, Sapienza University of Rome, Italy

[*]Correspondence to: Marina Zanella, marina.zannella@uniroma1.it

DOI: 10.1553/populationyearbook2021.res3.3

2014). The upward trend in the time spent on childcare has been accompanied by a cultural shift towards more time-intensive and child-centred parenting (Hays 1996; Craig et al. 2014). Busier parents have adapted to "the changing rhythms of family life" by becoming increasingly likely to engage in multitasking to ensure they are spending adequate amounts of time with their children (Bianchi et al. 2006). Prior research has suggested that multitasking is a source of gender inequality, as mothers are more likely than fathers to multitask by doing housework and childcare simultaneously. Thus, it appears that this trend has further widened the gender gap in time use patterns. Moreover, for mothers, multitasking at home is more likely to be associated with negative emotional experiences and work-family conflicts (Offer and Schneider 2011). The rise of the intensive parenting culture has contributed to mothers feeling increased pressure to spend time with their children, given their traditional role as the primary caregiver. The significant changes in the context and the meaning of parenting that have taken place in recent decades have led scholars to pay increasing attention to the well-being of parents. A number of studies have shown that parents have lower general levels of subjective well-being than childless people (e.g., Aassve et al. 2012). Nelson and colleagues (2014) have suggested that there is a highly complex relationship between parenthood and well-being. Specifically, they have argued that parents are more likely than non-parents to report negative experiences (e.g., financial problems, sleep disturbances, troubled marriages), but that parents also experience more positive emotions, such as happiness and the feeling that their life has greater meaning. Drawing on data from the American Time Use Survey (ATUS), Negraia and Augustine (2020) compared the well-being of parents and non-parents in a large set of time use activities, and found support for the claim that parental experiences are a "mixed bag".[1] Fewer studies have analysed the relationship between spending time on childcare and well-being among parents. Connelly and Kimmel (2015) provided evidence that both American mothers and fathers enjoy the time they spend on childcare. Their findings also indicated that fathers enjoy their parenting time more than mothers, with the latter reporting higher levels of stress and tiredness. Similarly, Musick and colleagues (2016), using ATUS data, found that while parents consistently report greater subjective well-being when engaged in activities with their children than without, mothers report feeling less happiness and more stress and fatigue than fathers when spending time with their children. According to McDonnell and colleagues (2019), who used ATUS data to examine mothers' and fathers' momentary affect during childcare activities, such imbalances are partly explained by the gendered nature of parenting; i.e., by the considerable differences in the amounts of time as well as in the characteristics of the time mothers and fathers spend with their children.

[1] The authors show that parents of minor children experience higher levels of positive emotions (happiness, meaning) and less sadness than non-parents in their daily activities, but also more negative emotions (stress and fatigue).

To date, most studies on parents' well-being when spending time on childcare have focused on the US (e.g., Offer 2014; Connelly and Kimmel 2015; Musick et al. 2016; Meier et al. 2018), while European research on this topic has been rarer, partly due to data availability.[2] This article investigates the subjective experience of time spent on childcare among Italian fathers and mothers. Our focus is on two aspects of parenting that have been shown to be considerably gendered: the type of childcare activity performed and the simultaneous performance of childcare and housework (i.e., multitasking). We also take into consideration other dimensions of the "care context" (McDonnell et al. 2019) by controlling our results for a number of relevant characteristics of childcare that have been found to differ considerably between mothers and fathers, such as the presence of other adults, the amount of time spent on childcare during the day, the duration of the activity and the time of day when the activity is performed. We aim to answer the following research questions: Do momentary assessments of well-being vary according to the type of childcare activity being performed by mothers and fathers? Does performing childcare and housework tasks simultaneously affect parents' subjective experience of the time they spend with their children? Do men and women differ in their multitasking behaviours? Drawing on prior research that showed that performing multiple unpaid tasks at the same time is associated with a feeling of being under time pressure (Craig and Brown 2017), and that multitasking at home is related to negative emotions and psychological distress for mothers only (Offer and Schneider 2011), we expect to find that multitasking when performing childcare negatively affects the subjective well-being of mothers, but not of fathers. Furthermore, we assume that the negative effect of multitasking will be more evident when parents are engaged in physical care and supervising activities for two reasons. First, as such activities have been conceptualised as forms of routine unpaid care work given their repetitive nature and the relatively low degree of interaction involved (e.g., Craig 2006; Craig and Powell 2011), they are likely to be less enjoyable for parents than engaging in other activities with their children, such as recreational activities. Second, there is ample evidence of a gendered division of childcare tasks, with women taking on the largest share of the routine work burden (e.g., Fuligni and Brooks-Gunn 2004; Craig and Mullan 2011). Therefore, we assume that combining routine care work with other unpaid routine tasks is likely to result in a less enjoyable experience for mothers.

Our study contributes to the literature in the field by evaluating how multitasking affects mothers' and fathers' experiences of the time they spent on childcare, while distinguishing between routine and recreational activities. To this end, the paper is organised as follows. Section 2 presents the theoretical background of the paper, with a focus on the relationship between multitasking and well-being. Section 3 describes the data and methods used, while Section 4 presents the

[2] Time use surveys are less frequent, and information on the assessment of momentary well-being is available for a few European countries only.

results. More specifically, Section 4.1 reports the descriptive findings for detailed childcare activities (i.e., routine care, recreational care, educational care and travel-related care) in order to provide an initial broad picture of the differences between mothers and fathers in the amounts and the nature of the time they spent with their children. Then, Section 4.2 presents the model results on the relationship between multitasking and the subjective well-being of mothers and fathers. Due to the small sample size of multitasking episodes recorded by parents while performing educational and travel-related childcare tasks, only the results for routine and recreational activities (in addition to the results related to the overall childcare episodes) are presented. Section 4.3 provides a brief discussion of the sensitive analyses we performed to check the robustness of the models' results. Finally, Section 5 concludes.

2 Background

In recent decades, the rhythm of family life in industrialised countries has changed, as women's labour market participation has risen significantly, and double-earner households have become increasingly common (Bianchi et al. 2006). The observation that the new role of women in the labour market would inevitably reduce the amount of time they had available to devote to unpaid family work led to concerns being raised that these changes were occurring at the expense of children, with mothers taking time away from their childcare duties to meet their job responsibilities. However, scholars have since demystified the *bad working mother* myth by providing evidence that contemporary employed mothers are devoting at least as much time to their children as the "Golden Age" housewives (Bianchi and Milkie 2010; Liss et al. 2013). Moreover, even as mothers have continued to spend a significant portion of their time caring for their children, the amount of time fathers are spending with their children has been increasing since the 1960s in the US (Bianchi 2000; Wang and Bianchi 2009; Hofferth and Lee 2015) and in Europe (Goldscheider et al. 2014). In their comparative study of 16 industrialised countries, Gauthier and colleagues (2004) showed that the upward trend in the time parents are spending on childcare can be mostly explained by their increasing time investments in activities involving a high degree of interaction with their children, such as playing. Higher parental time investments have also been shown to have positive effects on the cognitive development of children (Del Boca et al. 2014), and, ultimately, on the production of human capital in many industrialised countries (Vargha and Donehower 2019). However, this upward trend in the amount of time parents are devoting to childcare is likely to result in increasing pressures on parents. The contemporary growth of dual-earner couples has contributed to increases in parents' workloads, and, thus, to time squeezes for individuals and families (Buddelmeyer et al. 2018). Time-use studies conducted in different European countries have shown that dual-earner couples with young children are especially likely to be rushed (e.g., Zannella et al. 2019). While the amount of time men spend on childcare has been increasing, the gender gap in unpaid care work persists in many highly industrialised countries (Altintas and Sullivan 2016). These findings suggest that many women today work a second shift at home in addition to the paid job they perform in the formal sector (Hochschild 1989; Bianchi and Milkie 2010; Neilson and Stanfors 2014). As Mattingly and Sayer (2006: 1) noted: "*persistent inequality in gendered time-use patterns is paralleled by gendered experiences of time pressure*". In other words, compared to men, women not only experience more time scarcity, they are also more likely to report feelings of time pressure and other forms of psychological distress (Nelson et al. 2014).

2.1 Multitasking and well-being

Parents have responded to growing time pressures by becoming increasingly likely to engage in multitasking; that is, to performing different activities at the same time (Bianchi et al. 2006). Based on time use data for the UK, Sullivan and Gershuny (2013; 2018) have consistently shown that women are more likely than men to report multitasking time use episodes, especially when carrying out childcare or housework tasks. Nevertheless, the authors did not find support for the assumption that there is a relationship between multitasking and being rushed. By contrast, drawing on Australian time use data, Craig and Brown (2017) found evidence of a positive association between multitasking by performing multiple unpaid work tasks simultaneously and feelings of time pressure among mothers. Thus, their results suggest that gender differences in multitasking behaviour are related not only to the quantity of time spent performing parallel activities, but also to the different levels of well-being parents experience while engaged in these activities.

A recent branch of the literature on the time use patterns of mothers and fathers has focused on the relationship between engaging in parenting activities and parents' subjective well-being. These studies have found that spending time with their children has positive effects on the well-being of parents; and that fathers tend to enjoy spending time on childcare more than mothers, with the latter being more likely to report feeling less happiness and more stress and fatigue when engaged in childcare tasks (e.g., Musick et al. 2016). McDonnell et al. (2019) have highlighted the role played by the contextual characteristics of care in determining gender imbalances in subjective experiences of childcare. However, in their study based on ATUS data, the authors overlooked the role of multitasking in their investigation of how the "care context" influences mothers' and fathers' momentary affect during childcare activities. Indeed, very little is known about the relationship between multitasking and measures of the momentary well-being of parents. In one of the few empirical studies on the topic, Offer and Schneider (2011) used data from the 500 Family Study on American dual-earner middle-class parents collected via the experience sampling method to investigate how multitasking influenced mothers' experiences of emotional stress and negative affect. However, in their analysis of the relationship between multitasking and momentary well-being, the authors focused on where the simultaneous activities were performed (e.g., at home, at work), without taking into consideration the specific nature (or typology) of the activities that were carried out simultaneously. In a more recent study, Dunatchik and Speight (2020) used data from the UK Time Use Survey to investigate whether for parents, simultaneously performing two unpaid work tasks or combining an unpaid work task with a secondary leisure activity was associated with higher or lower levels of enjoyment of the time spent on unpaid work. Their results indicated that the parents enjoyed performing housework and childcare tasks more when they combined these activities with a secondary leisure activity. However, the results for the parents' perceptions of combining multiple unpaid work activities were mixed. Indeed, the study found that for mothers, multitasking by combining a primary childcare task with a secondary housework activity was associated with less enjoyment in most instances, while the opposite was the case for fathers.

This article expands on previous time use literature by investigating parental well-being while performing childcare, with a focus on multitasking, defined as combining childcare with housework. This approach is motivated by existing evidence that multitasking by performing multiple unpaid work tasks simultaneously is the "most intense" type of multitasking (Sayer 2007a). I.e., it is associated with the feeling of being under time pressure (e.g., Craig and Brown 2017) and entails more work for mothers, which further enlarges the gender gap in unpaid work (see, among others, Craig 2007; Sayer 2007b). Our definition of multitasking includes all possible combinations of childcare with housework, regardless of whether childcare was recorded by the respondents as the primary or the secondary activity. Indeed, including secondary childcare activities is inherently relevant for the analysis of the subjective experience of parenting time for two main reasons. First, including only childcare as a primary activity would underestimate the amount of time spent by parents e, and especially mothers, on childcare. Second, the amount of time spent by parents on secondary childcare is likely to affect their subjective experiences of parenting time (see Craig 2006).

3 Data and methods

The study builds on data from the Italian Time Use Survey (ITUS) for the year 2013/4 to evaluate fathers' and mothers' subjective well-being during the time they spent caring for their children. ITUS includes three data files: the individual file, the daily diary and the weekly diary. The daily diary consisted of time data collected through the diary technique: the respondents recorded their time use during the previous 24 hours in their own words. The time diaries were randomly distributed across the days of the week to all household members aged three years and older. The respondents were asked to fill in the daily diary every 10 minutes by reporting the following information: a description of the main activity that was being carried out, the possible presence of a parallel secondary activity, the location where the activity was being performed, the level of enjoyment the respondent was experiencing, and whether the respondent was alone or with other people (e.g., partner, own children, other family members, friends). The individual file includes information about the socio-demographic characteristics of the individual respondent, the characteristics of the household, background information and information on the well-being of the respondent in different life domains. Our target population consisted of 6,582 individuals aged 25 to 54 years who were parents of dependent children aged up to 14 years.[3] We relied on information from the episode enjoyment scores used by the respondents to evaluate the degree of (un)pleasantness associated with the different activities recorded in the daily diary to analyse the well-being of mothers and fathers when they were spending time on childcare. The scores ranged from −3 (very unpleasant) to 3 (very pleasant), and reflected answers to the following question: *How pleasant was the moment?* As Kahneman and Krueger (2006, p. 18) have observed: "[...] one could legitimately question whether one should give a cardinal interpretation to the numeric values attached to individuals' responses about their life satisfaction or emotional states [...]". For this reason, we decided to build a dichotomous indicator of momentary well-being (MW) based on the actual distribution of the enjoyment scores. For each childcare episode analysed, the MW indicator was assigned a value of one if its enjoyment score was in the lowest 25th percentile of the scores recorded by the respondents for the corresponding activity; and of zero otherwise.[4,5] The sample distribution of the enjoyment scores by the type of childcare activity performed is shown in Table A.1 in the appendix. The 25th percentile on which our indicator is based is equal to one (i.e., "somewhat pleasant") for all the activities considered.[6] Thus, our indicator distinguished between fully pleasant (i.e., very pleasant and pleasant) and more nuanced experiences, including childcare episodes rated as fairly pleasant, neutral or unpleasant.

Our main aim is to assess the subjective well-being of the parents when they were spending time on childcare. We are particularly interested in understanding whether and, if so, to what extent mothers' and fathers' enjoyment of childcare time varied depending on the type of childcare activity they were performing, and

---

[3] The choice of this specific age cut for parents is motivated by the very small numbers of fathers and mothers at younger (1.1%) and older ages (3.1%) in our sample. Indeed, the selected age group includes 95.8% of the sampled parents who reported living with dependent children.

[4] The 25th percentile is calculated on the complete sample of mothers and fathers.

[5] As shown in Section 4.3, in order to perform a sensitivity analysis, we repeated the analysis using the episode enjoyment scores as the dependent variable in a linear regression model. We found that even the cardinal interpretation of the data on subjective momentary well-being leads to the same path of results. Thus, our findings confirmed those reported by Musick et al. (2016), who first applied the linear regression approach, and then generated the dichotomous indicator of well-being for their own sensitivity analysis.

[6] Unfortunately, we had to exclude travelling and teaching from the analysis due to the small number of multitasking episodes involving these activities that were reported by parents (and especially fathers) (see Table 3).

on whether they were doing housework as a parallel activity (i.e., multitasking). For this purpose, we use logistic regression to analyse the parents' probability of experiencing a childcare episode that was not (fully) pleasant (MW indicator = 1). We run separate models for the subsamples of men and women for each of the following activities considered: performing overall childcare, providing routine care and providing recreational care (resulting in six models in total). Our main explicative variable is multitasking. As we explained in Section 2.1, our definition of multitasking includes all possible combinations of childcare tasks with housework, regardless of whether childcare was recorded by the respondent as a primary or as a secondary activity. The control variables include the following: individual characteristics (age, education, employment status, life satisfaction, single parent/living in a couple), household characteristics (number of children, age of the youngest child, the number of daily hours spent on childcare), characteristics of the episode (i.e., the episode's starting hour, the presence of at least one other adult, the duration of the episode) as well as characteristics of the diary entry (day of the week, survey quarter). In order to check the robustness of our results, we replicated the analysis by using the episode enjoyment scores as a dependent variable in a linear regression model (Section 4.3).

4 Results

4.1 Descriptive results

Our subsample of parents recorded 31,309 childcare episodes, the great majority of which (more than 70%) were performed by mothers. Thus, the numbers of childcare episodes reported by men and women differed considerably, as did the distributions of the episodes across different types of childcare activities (Table 1).

Table 1:
Composition of childcare episodes by detailed activities

	Men		**Women**	
	N	%	*N*	%
Overall childcare	8,974	100.0	22,335	100.0
Routine care	2,386	26.6	8,631	38.6
Educational care	212	2.4	743	3.3
Recreational care	4,300	47.9	8,436	37.8
Travel-related care	2,072	23.1	4,525	20.3

Source: Authors' calculations based on ITUS data (2013/4).

Of the total childcare episodes reported by women, 38.6% involved routine care tasks (i.e., providing physical care and supervising children). By contrast, of the total childcare episodes reported by men, 26.6% were routine care activities. Conversely, of the childcare episodes reported by women, 37.8% involved recreational activities (e.g., playing with the child, talking to the child, reading to the child), compared to 47.9% of the childcare episodes reported by men. Women reported spending more time (3.3% of episodes) on educational childcare (e.g., helping the child doing the homework, correcting homework, etc.) than men (2.4% of episodes), while the opposite pattern emerged for episodes related to accompanying children and other travel-related childcare.

Information on the average hours per week mothers and fathers devoted to both primary and secondary childcare is shown in Table 2. Although the focus of this paper is on the number and the nature of the childcare episodes recorded by each parent in the time diary, such information is relevant to obtain a broader understanding of the gendered nature of the time spent on childcare. On average, mothers spent about 22 h and fathers spent about 13 h per week caring for their children. In line with the results of other studies, we found that the inclusion of secondary childcare activities increased the gender gap in unpaid work (e.g., Craig 2006), with mothers spending twice as much time as fathers on secondary childcare. Most of the time spent on secondary childcare by both men and women involved recreational activities (more than 3 h per week for mothers and almost 2 h per week for fathers).

Table 2:
Mean week hours devoted to childcare activities by gender (weighted sample estimates)

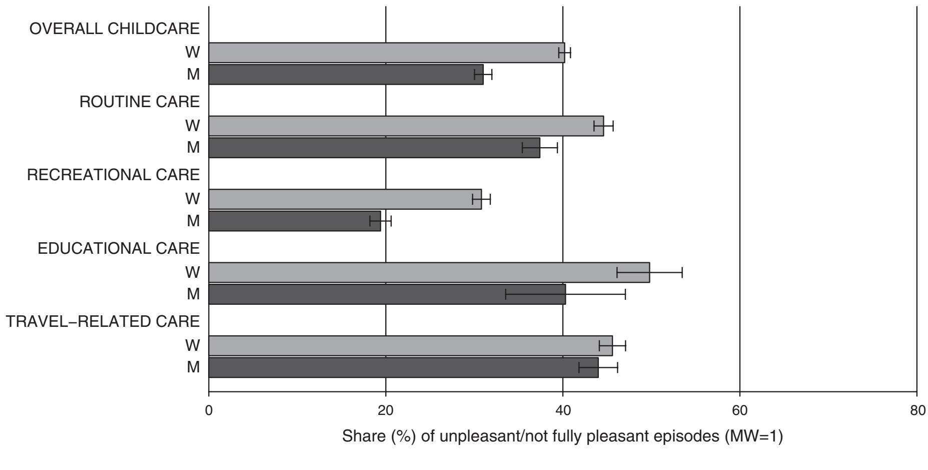
	Overall childcare		**Routine care**		**Educational care**		**Recreational care**		**Travel-related care**	
	Mean	**Std err**	**Mean**	**Std err**	**Mean**	**Std err**	**Mean**	**Std err**	**Mean**	**Std err**
					Men					
Main	10.4	0.21	2.8	0.11	0.7	0.06	4.4	0.13	2.6	0.13
Secondary	2.2	0.10	0.4	0.05	0.0	0.00	1.8	0.09	0.0	0.01
Total	**12.6**	**0.23**	**3.2**	**0.12**	**0.7**	**0.06**	**6.2**	**0.15**	**2.6**	**0.13**
					Women					
Main	17.1	0.23	7.6	0.17	1.7	0.07	3.4	0.10	4.4	0.12
Secondary	4.6	0.14	0.8	0.06	0.0	0.00	3.6	0.12	0.2	0.02
Total	**21.7**	**0.27**	**8.4**	**0.19**	**1.7**	**0.07**	**7.0**	**0.16**	**4.6**	**0.12**

Source: Authors' calculations based on ITUS data (2013/4).

In the individual ITUS questionnaire, parents were asked to assess their level of satisfaction with the quantity of time they were spending with their children. Around 67% of mothers and 44.4% of fathers stated that they were satisfied. The majority of men (54%), but less than one-third (31%) of women, said that they felt they were spending too little time with their children. By contrast, only a very small proportion of the parents said they felt they were spending too much time with their children (2.2% of mothers and 1.5% of fathers).

In line with the results of previous studies (e.g., Offer 2014; Connelly and Kimmel 2015; Musick et al. 2016), we found that the parents in our sample reported enjoying the time they spent on childcare: 81.5% out of 31,309 total childcare episodes were rated as positive, 15.2% were rated as neutral, and only 3.3% were rated as negative (see Table A.1). Nevertheless, compared to fathers, mothers reported having fewer very pleasant moments, as well as more negative and neutral childcare episodes. The sample distribution of childcare episodes according to the MW indicator, gender and type of childcare activity is shown in Figure 1. As we mentioned in Section 3, the 25th percentile on which our indicator is based was positive – that is, was equal to one (i.e., "somewhat pleasant") – for all of the activities we considered; meaning that our indicator of momentary well-being distinguished between fully pleasant and less pleasant or unpleasant childcare episodes. In total, a larger share of childcare episodes was rated as not pleasant/not fully pleasant by women (40.2%) than by men (31.0%). Gender disparities were found for all of the activities considered (although

Figure 1:
Sample distribution of unpleasant or not fully pleasant (MW = 1) episodes by gender and type of childcare activity



Source: Authors' calculations based on ITUS data (2013/4).
Note: The MW indicator is based on time use enjoyment scores. The indicator is equal to one when the episode was assessed as unpleasant or not fully pleasant by the respondent.

Table 3:
Share (%) of multitasking episodes on total episodes by gender and detailed childcare activity

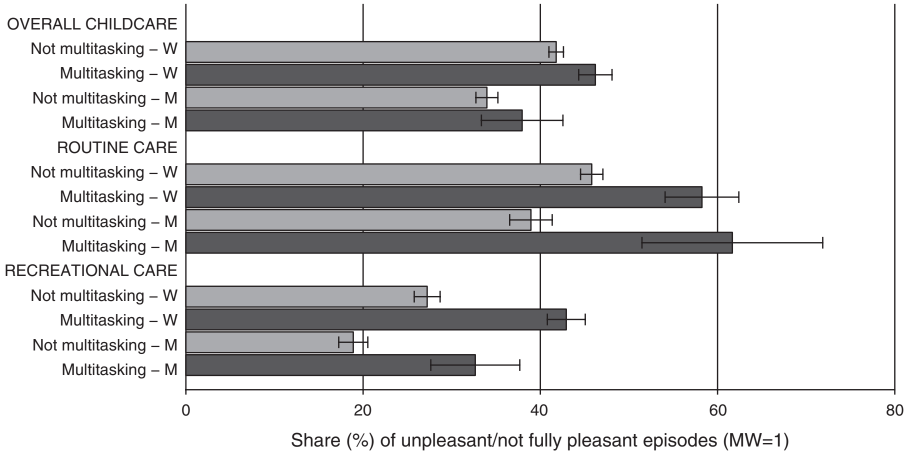
Multitasking	Overall childcare	Routine care	Educational care	Recreational care	Travel-related care
			Men		
Yes	5.4	3.9	0.0	8.2	0.2
Not	94.6	96.1	100	91.8	99.8
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
			Women		
Yes	12.7	6.5	6.1	25.3	2.1
Not	87.3	93.5	93.9	74.7	97.9
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: Authors' calculations based on ITUS data (2013/4).

these differences were not statistically significant for travel-related episodes). For both mothers and fathers, recreational childcare was the activity for which the lowest share of unpleasant childcare episodes was reported. However, this was also the activity for which the largest gender differences were detected (with the MW indicator being equal to one in 30.8% of the episodes reported by mothers and in 19.4% of the episodes reported by fathers).

Table 3 shows the distribution of multitasking childcare episodes by gender and detailed activities. In line with prior research on multitasking (e.g., Craig 2006; Craig and Brown 2017), we found that the share of total childcare episodes in which the respondent reported also doing housework was larger for mothers (12.7%) than for fathers (5.4%). The largest shares of multitasking episodes reported by both mothers (25.3%) and fathers (8.2%) involved recreational activities. By contrast, almost none of the fathers reported engaging in multitasking episodes involving educational or travel-related childcare. Thus, we were unable to further analyse these two activities. Figure 2 displays sample estimates of the shares of childcare episodes rated as unpleasant/not fully pleasant by gender and by whether the episodes involved multitasking while providing overall, routine or recreational childcare. Multitasking episodes were rated as considerably less pleasant by parents than episodes in which childcare was the only activity they were performing. This finding held for both genders and for all the three categories considered, although the confidence intervals indicate that the result was not statistically significant for overall childcare performed by men. In the next section, the relationship between multitasking and parental well-being while spending time on childcare is further investigated in a multivariate setting.

Figure 2:
Sample distribution of unpleasant or not fully pleasant (MW = 1) episodes by gender, type of childcare activity and multitasking (i.e., doing housework as a parallel activity)



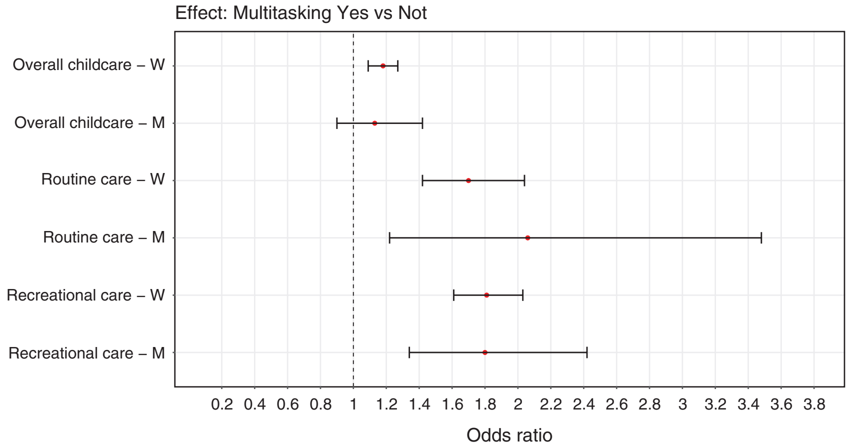
Source: Authors' calculations based on ITUS data (2013/4).
Note: The MW indicator is based on time use enjoyment scores. The indicator is equal to one when the episode was assessed as unpleasant or as not fully pleasant by the respondent.

4.2 Model results

The results of logistic models on the relationship between multitasking and momentary well-being are presented in Figure 3. Performing routine care together with housework was negatively associated with levels of subjective well-being for both mothers and fathers. Contrary to our expectations, we found that multitasking negatively affected the subjective perception of providing routine care for fathers as well as for mothers, even though fathers generally take on a considerably smaller share of the unpaid routine work burden than mothers. Similarly, multitasking significantly reduced the well-being of both mothers and fathers when they were playing or interacting with their children. Indeed, the negative effect of multitasking on the well-being of mothers did not seem to change much depending on the type (i.e., routine or non-routine) of childcare tasks they were performing. In terms of the time spent on overall childcare, multitasking negatively affected the well-being of both parents, but the result was statistically significant only for mothers.[7]

[7] This can likely be at least partly explained by the small sample size of the multitasking episodes reported by fathers involving childcare activities other than routine and recreational care (see Table 3).

Figure 3:
Odds ratio of experiencing an unpleasant childcare episode depending on the presence of multitasking



Source: Authors' calculations based on ITUS data (2013/4).

In summary, our results show that for the parents in our sample, there was a negative relationship between multitasking and their enjoyment of the time they spent on childcare. This result held for both genders, and regardless of the nature of the activity performed (i.e., whether it was routine or recreational care). Thus, whereas prior studies showed a negative effect of multitasking on individual well-being for mothers only (Offer and Schneider 2011; Craig and Brown 2017; Dunatchik and Speight 2020), our results indicate that multitasking in the domestic sphere negatively affects the momentary well-being of fathers as well.

As we mentioned in Section 3, the results relating to the effect of multitasking on mothers' and fathers' subjective perceptions of the time they spent on childcare were obtained by controlling for a large number of characteristics of the respondent, the household and the episode. A brief discussion of the relative coefficients obtained from the models reported in the appendix (Tables A.2–A.7) sheds further light on this phenomenon. A common result is related to the presence of another adult while spending time caring for children: both mothers and fathers reported that sharing the time spent on childcare with another adult increased their enjoyment of the activity (although the coefficient was not statistically significant for men performing routine care). Indeed, previous research has shown that for parents, spending solo time with their children is associated with greater responsibilities, while sharing childcare with other adults lightens the pressure on them (Craig 2006), and, thus, relieves parenting "role overload" (see Nomaguchi and Milkie 2020). Another recurring result was related to the duration of the episode: relatively short (less than 40 minutes) and relatively long episodes (more than 60 minutes) were found to be negatively associated with parental well-being (except for men during routine care). A similar result was found in relation to the average amount of time spent on childcare during the day, except for routine care provided by mothers, for which longer hours devoted to the activity were associated with higher probabilities of experiencing unpleasant moments. This result seems to corroborate the findings of prior research on the gendered nature of parenting time, which showed that spending more cumulative time on childcare during the day and bearing most of the responsibility for routine care tend to have detrimental effects on the well-being of mothers (McDonnell et al. 2019). Looking at the household characteristics, we observed a negative relationship between the number of children and the momentary well-being of parents. For overall and routine care, having pre-school children was negatively associated with the probability of reporting unpleasant episodes. This result is consistent with evidence from existing studies that analysed parental well-being according to the age of the child, and found that parents had higher levels of well-being when caring for very young children, despite the intense time demands associated with caring for them (e.g., Meier et al. 2018). Being a single parent was not shown to be statistically significantly related to the degree of enjoyment parents experienced while spending time on childcare. When we looked at the characteristics of the respondents, we observed that for all parents and for all of the activities considered, being dissatisfied with life in general was associated with higher probabilities of experiencing less enjoyment of the time spent on childcare. The results for men's employment status pointed in a similar direction: unemployment was positively associated with the probability of reporting less enjoyment during overall and recreational childcare activities.[8] Indeed, prior research has shown that there is a strong negative relationship between unemployment and life satisfaction, and that people who are unemployed are more likely than people who are employed to report that they have less enjoyment of the activities they engage in (Hoang and Knabe 2020). All in all, we found that parents' enjoyment of routine care was more weakly related to the individual characteristics of fathers and mothers than parents' enjoyment of recreational care.

4.3 Sensitivity/robustness analysis

We checked the robustness of our results by developing a sensitivity analysis to evaluate whether our results changed depending on the measure used to estimate mothers' and fathers' enjoyment of their parenting time. We used the episode enjoyment scores as the dependent variable, and modelled it as a continuous variable using linear regression (see also Musick et al. 2016; Meier et al. 2018; Negraia and Augustine 2020). We adopted the same analytical strategy that we used for the MW indicator, and ran separate models by gender and by type of activity (resulting in

[8] A similar direction of the relationship was also observed for women in the models for routine care; however, the coefficients were not statistically significant.

a total of six models). The independent variables were also the same. The results from the regression analysis confirm our finding that for all of the activities we considered, performing childcare and housework at the same time was associated with lower levels of subjective well-being for both mothers and fathers, although the effect was not significant for men in relation to the time they spent on overall childcare due to the small number of multitasking episodes they reported while performing educational and travel-related childcare (Table A.8). Overall, the results for the effects of multitasking on the subjective experiences of parenting time are in line with those based on the dichotomous indicator.

5 Concluding remarks

In this article, we used data from the latest time use survey available for Italy (year 2013/4) to study Italian mothers' and fathers' enjoyment of the time they spent on childcare. To do so, we took advantage of novel information included in the latest version of ITUS: the enjoyment scores used by the respondents to evaluate the degree of (un)pleasantness associated with episodes of the different activities recorded in a daily diary. Unlike the ATUS and other time use surveys, in which respondents were asked to report on five emotions (i.e., meaning, happiness, sadness, tiredness, stress and pain) for a number of sampled activities, the ITUS respondents were asked to rate their general level of enjoyment for each time use episode experienced during the day. This approach has both advantages and disadvantages. On one hand, it does not permit to distinguish between the different emotions that determine the respondents' levels of momentary well-being. For instance, existing studies have shown that although mothers and fathers report similarly high levels of happiness and meaning when caring for their children, mothers report higher levels of tiredness and stress when engaged in almost every childcare activity (e.g., Musick et al. 2016). On the other hand, this approach allows for a more in-depth analysis of the link between well-being and the use of time for detailed activities.

Our main goal was to assess the relationship between multitasking and parental well-being. To this end, we created a marker of multitasking to indicate whether childcare was performed as a single task or in combination with other unpaid work activities. Thus, we tested the association between multitasking and parents' momentary assessments of well-being across different types of childcare activities, with a gender perspective. Our results provide further support for the findings of prior research showing that both mothers and fathers positively assess the time they spend on childcare, but that mothers are more likely than fathers to report having less pleasant moments (e.g., Connelly and Kimmel 2015). In line with previous studies (e.g., Sayer 2007b; Offer and Schneider 2011), we found that mothers were more likely than fathers to report that they were performing childcare and housework simultaneously. Thus, by engaging in multitasking, women were doing more work than men, and the gender gap in unpaid work increased (Craig 2007). Our model results demonstrated that there was a negative relationship between multitasking and momentary well-being for both mothers and fathers, regardless of the kind of activity they were performing (whether it was routine or recreational childcare). This finding adds insights to prior research on parental well-being by providing new evidence that multitasking is an important contextual characteristic of care that affects the subjective well-being of fathers, as well as of mothers. Indeed contemporary parents who are time squeezed are likely to feel compelled to spend as much time as possible with their children, even if this means "compressing" time by performing multiple tasks simultaneously (Craig and Brown 2016). Parenthood entails greater unpaid work requirements, not only in terms of the time needed for childcare, but for domestic work as well (Craig and Bittman 2008). From this perspective, multitasking at home may been seen as one of the strategies parents adopt to cope with this time pressure.

References

Aassve, A., A. Goisis and M. Sironi 2012. Happiness and childbearing across Europe. *Social Indicators Research* 108(1): 65–86. https://doi.org/10.1007/s11205-011-9866-x

Altintas, E. and O. Sullivan 2016. Fifty years of change updated: Cross-national gender convergence in housework. *Demographic Research* 35: 455–470. https://doi.org/10.4054/DemRes.2016.35.16

Bianchi, S. M. 2000. Maternal employment and time with children: Dramatic change or surprising continuity? *Demography* 37(4): 401–414. https://doi.org/10.1353/dem.2000.0001

Bianchi, S. M. and M. A. Milkie 2010. Work and family research in the first decade of the 21st century. *Journal of Marriage and Family 72*(3): 705–725. https://doi.org/10.1111/j.1741-3737.2010.00726.x

Bianchi, S. M., J. P. Robinson and M. A. Milkie 2006. *The changing rhythms of American family life*. Russell Sage Foundation.

Buddelmeyer, H., D. S. Hamermesh and M. Wooden 2018. The stress cost of children on moms and dads. *European Economic Review* 109(C): 148–161. https://doi.org/10.1016/j.euroecorev.2016.12.012

Craig, L. 2006. Does father care mean fathers share? A comparison of how mothers and fathers in intact families spend time with children, *Gender & Society* 20(2): 259–281. https://doi.org/10.1177/0891243205285212

Craig, L. 2007. Is there really a 'second shift', and if so, who does it? A time-diary investigation. *Feminist Review* 86(1): 149–170. https://doi.org/10.1057/palgrave.fr.9400339

Craig, L. and M. Bittman 2008. The incremental time costs of children: An analysis of children's impact on adult time use in Australia. *Feminist Economics* 14(2): 57–85. https://doi.org/10.1080/13545700701880999

Craig, L. and J. E. Brown 2016. The multitasking parent: Time penalties, dimensions, and gender differences. In *The economics of multitasking*, eds C. M. Kalenkoski and G. Foster, 33–59. New York: Palgrave Macmillan. https://doi.org/10.1057/9781137381446_4

Craig, L. and J. E. Brown 2017. Feeling rushed: Gendered time quality, work hours, nonstandard work schedules, and spousal crossover. *Journal of Marriage and Family* 79(1): 225–242. https://doi.org/10.1111/jomf.12320

Craig, L. and K. Mullan 2011. How mothers and fathers share childcare: A cross-national time-use comparison. *American Sociological Review* 76(6): 834–861. https://doi.org/10.1177%2F0003122411427673

Craig, L. and A. Powell 2011. Non-standard work schedules, work-family balance and the gendered division of childcare. *Work, Employment and Society 25*(2): 274–291. https://doi.org/10.1177/0950017011398894

Craig, L., A. Powell and C. Smyth 2014. Towards intensive parenting? Changes in the composition and determinants of mothers' and fathers' time with children 1992–2006. *The British Journal of Sociology* 65(3): 555–579. https://doi.org/10.1111/1468-4446.12035

Connelly, R. and J. Kimmel 2015. If you're happy and you know it: How do mothers and fathers in the US really feel about caring for their children? *Feminist Economics* 21(1): 1–34. https://doi.org/10.1080/13545701.2014.970210

Del Boca, D., C. Flinn and M. Wiswall 2014. Household choices and child development. *The Review of Economic Studies* 81(1): 137–185. https://doi.org/10.1093/restud/rdt026

Dunatchik, A. and S. Speight 2020. Re-examining how partner copresence and multitasking affect parents' enjoyment of childcare and housework. *Sociological Science* 7: 268–290. https://doi.org//10.15195/v7.a11

Faircloth, C. 2014. Intensive parenting and the expansion of parenting. In *Parenting culture studies*, eds E. J. Lee, J. Bristow, C. Faircloth and J. Macvarish, 25–50. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan. https://doi.org/10.1057/9781137304612_2

Fuligni, A. S. and J. Brooks-Gunn 2004. Measuring mother and father shared caregiving. In *Conceptualizing and Measuring Father Involvement*, ed R. D. Day, 299–315. Mahwah, NJ: Lawrence Erlbaum.

García-Mainar, I., J. A. Molina and V. M. Montuenga 2011. Gender differences in childcare: Time allocation in five European countries. *Feminist Economics* 17(1): 119–150. https://doi.org/10.1080/13545701.2010.542004

Gauthier, A. H., T. M. Smeeding and F. F. Furstenberg Jr 2004. Are parents investing less time in children? Trends in selected industrialized countries. *Population and Development Review* 30(4): 647–672. https://doi.org/10.1111/j.1728-4457.2004.00036.x

Goldscheider, F., E. Bernhardt and T. Lappegård 2014. Studies of men's involvement in the family—Part 1: Introduction. *Journal of Family Issue* 35(7), Special Issues. https://doi.org/10.1177%2F0192513X14522237

Hays, S. 1996. *The cultural contradictions of motherhood*. New Haven, CT: Yale University Press.

Hoang, T. T. A. and A. Knabe 2020. Time use, unemployment, and well-being: An empirical analysis using British time-use data. *Journal of Happiness Studies*. https://doi.org/10.1007/s10902-020-00320-x

Hochschild, A. R. 1989. *The second shift (working parents and the revolution at home)*. New York: Avon Books.

Hofferth, S. and Y. Lee 2015. Family structure and trends in US fathers' time with children, 2003–2013. *Family Science* 6(1): 318–329. https://doi.org/10.1080/19424620.2015.1082805

Kahneman, D. and A. Krueger 2006. Developments in the measurement of subjective wellbeing. *Journal of Economic Perspectives* 20(1): 3–24. https://doi.org/10.1257/089533006776526030

Liss, M., H. H. Schiffrin and K. M. Rizzo 2013. Maternal guilt and shame: The role of self-discrepancy and fear of negative evaluation. *Journal of Child and Family Studies* 22: 1112–1119. https://doi.org/10.1007/s10826-012-9673-2

Mattingly, M. J. and L. C. Sayer 2006. Under pressure: Gender differences in the relationship between free time and feeling rushed. *Journal of Marriage and the Family* 68(1): 205–221. https://psycnet.apa.org/doi/10.1111/j.1741-3737.2006.00242.x

McDonnell, C., N. Luke and S. E. Short 2019. Happy moms, happier dads: Gendered caregiving and parents' affect. *Journal of Family Issues* 40(17): 2553–2581. https://doi.org/10.1177%2F0192513X19860179

Meier, A., K. Musick, J. Fischer and S. Flood 2018. Mothers' and fathers' well-being in parenting across the arch of child development. *Journal of Marriage and Family* 80(4): 992–1004. https://doi.org/10.1111/jomf.12491

Musick, K., A. Meier and S. Flood 2016. How parents fare: Mothers' and fathers' subjective wellbeing in time with children. *American Sociological Review* 81(5): 1069–1095. https://doi.org/10.1177%2F0003122416663917

Negraia, D. V. and J. M. Augustine 2020. Unpacking the parenting well-being gap: The role of dynamic features of daily life across broader social contexts. *Social Psychology Quarterly* 83(3): 207–228. https://doi.org/10.1177%2F0190272520902453

Neilson, J. and M. Stanfors 2014. It's about time! Gender, parenthood, and household divisions of labor under different welfare regimes. *Journal of Family Issues* 35(8): 1066–1088. https://doi.org/10.1177%2F0192513X14522240

Nelson, S. K., K. Kushlev and S. Lyubomirsky 2014. The pains and pleasures of parenting: When, why, and how is parenthood associated with more or less well-being? *Psychological Bulletin* 140(3): 846–895. https://doi.org/10.1037/a0035444

Nomaguchi, K. and M. A. Milkie 2020. Parenthood and well-being: A decade in review. *Journal of Marriage and Family* 82(1): 198–223. https://doi.org/10.1111/jomf.12646

Offer, S. 2014. Time with children and employed parents' emotional well-being. *Social Science Research* 47: 92–203. https://doi.org/10.1016/j.ssresearch.2014.05.003

Offer, S. and B. Schneider 2011. Revisiting the gender gap in time-use patterns: Multitasking and wellbeing among mothers and fathers in dual-earner families. *American Sociological Review* 76(6): 809–833. https://doi.org/10.1177%2F0003122411425170

Sayer, L. 2007a. Gender differences in the relationship between long employee hours and multitasking. In *Workplace temporalities: Research in the sociology of work*, Vol. 17, ed B. A. Rubin, 405–435. Oxford, UK: Elsevier.

Sayer, L. 2007b. More work for mothers? Trends and gender differences in multitasking. In *Time competition: Disturbed balances and new options in work and care*, eds T. van der Lippe and P. Peters, 41–56. Cheltenham, UK: Edward Elgar.

Sullivan, O. and L. Gershuny 2013. Domestic outsourcing and multitasking: How much do they really contribute? *Social Science Research* 42(5): 1311–24. https://doi.org/10.1016/j.ssresearch.2013.05.004

Sullivan, O. and L. Gershuny 2018. Speed-up society? Evidence from the UK 2000 and 2015 time use diary surveys. *Sociology* 52(1): 20–38. https://doi.org/10.1177%2F0038038517712914

Vargha, L. and G. Donehower 2019. The quantity-quality tradeoff: A cross-country comparison of market and nonmarket investments per child in relation to fertility. *Population and Development Review* 45(2): 321–350. https://doi.org/10.1111/padr.12245

Wang, R. and S. M. Bianchi 2009. ATUS fathers' involvement in childcare. *Social Indicators Research* 93: 141–145. https://doi.org/10.1007/s11205-008-9387-4

Zannella, M., B. Hammer, A. Prskawetz and J. Sambt 2019. A quantitative assessment of the rush hour of life in Austria, Italy and Slovenia. *European Journal of Population* 35(4): 751–776. https://doi.org/10.1007/s10680-018-9502-4

Appendix

Table A.1:
Distribution of enjoyment episode scores by gender and type of childcare activity

	Men			Women			All		
Score	*N*	%	Cumul. %	*N*	%	Cumul. %	*N*	%	Cumul. %
				Overall childcare					
−3	25	0.3	0.3	174	0.8	0.8	199	0.6	0.6
−2	55	0.6	0.9	160	0.7	1.5	215	0.7	1.3
−1	139	1.5	2.4	484	2.2	3.7	623	2.0	3.3
0	1,120	12.5	14.9	3,655	16.3	20.0	4,775	15.2	18.5
1	1,501	16.7	31.6	4,214	18.9	38.9	5,715	18.3	36.8
2	2,315	25.8	57.4	5,823	26.1	65.0	8,138	26.0	62.8
3	3,819	42.6	100	7,825	35	100.0	11,644	37.2	100.0
Total	8,974	100.0		22,335	100		31,309	100.0	
				Routine care activities					
−3	14	0.6	0.6	66	0.8	0.8	80	0.7	0.7
−2	20	0.8	1.4	73	0.8	1.6	93	0.8	1.5
−1	56	2.3	3.7	214	2.5	4.1	270	2.5	4
0	352	14.8	18.5	1,532	17.8	21.9	1,884	17.1	21.1
1	479	20.1	38.6	1,709	19.8	41.7	2,188	19.9	41
2	610	25.6	64.2	2,231	25.8	67.5	2,841	25.8	66.8
3	855	35.8	100	2,806	32.5	100.0	3,661	33.2	100
Total	2,386	100.0		8,631	100		11,017	100.0	
				Educational care activities					
−3	0	0.0	0	9	1.2	1.2	9	0.9	0.9
−2	0	0.0	0	16	2.2	3.4	16	1.7	2.6
−1	5	2.3	2.3	33	4.4	7.8	38	4.0	6.6
0	37	17.5	19.8	125	16.8	24.6	162	16.9	23.5
1	52	24.5	44.3	219	29.5	54.1	271	28.4	51.9
2	61	28.8	73.1	189	25.4	79.5	250	26.2	78.1
3	57	26.9	100	152	20.5	100.0	209	21.9	100
Total	212	100.0		743	100		955	100.0	

Continued

Table A.1: Continued

	Men			Women			All		
Score	***N***	**%**	**Cumul. %**	***N***	**%**	**Cumul. %**	***N***	**%**	**Cumul. %**
				Recreational care activities					
−3	0	0.2	0.2	9	0.6	0.6	9	0.5	0.5
−2	0	0.3	0.5	16	0.6	1.2	16	0.5	1
−1	5	0.8	1.3	33	1.6	2.8	38	1.3	2.3
0	37	8.2	9.5	125	12.5	15.3	162	11.0	13.3
1	52	11.7	21.2	219	15.6	30.9	271	14.2	27.5
2	61	25.3	46.5	189	26.3	57.2	250	26.0	53.5
3	57	53.5	100	152	42.8	100.0	209	46.5	100
Total	212	100.0		743	100		955	100.0	
				Travel-related care activities					
−3	0	0.2	0.2	9	1	1.0	9	0.8	0.8
−2	0	1.0	1.2	16	0.6	1.6	16	0.7	1.5
−1	5	2.0	3.2	33	2.2	3.8	38	2.2	3.7
0	37	18.4	21.6	125	20.8	24.6	162	20.0	23.7
1	52	22.5	44.1	219	21.5	46.1	271	21.8	45.5
2	61	26.8	70.9	189	26.2	72.3	250	26.4	71.9
3	57	29.1	100	152	27.7	100.0	209	28.1	100
Total	212	100.0		743	100		955	100.0	

Source: Authors' calculations based on ITUS data (2013/4).

Table A.2:
Logit model – Overall childcare, men

	Parameter estimate	Standard error	Wald Chi square	Prob > Chi square
Intercept	−0.397	0.122	10.520	<0.001
Multitasking (Ref = not)				
Yes	0.060	0.058	1.043	0.307
Episode duration (Ref = > 60 min.)				
Less than 20 min	0.266	0.047	32.640	<0.001
20–39 min	0.304	0.050	36.792	<0.001
40–59 min	−0.258	0.074	12.137	<0.001
Presence of an adult (Ref = not)				
Yes	−0.198	0.029	47.316	<0.001
Daily childcare time (Ref = >240 min.)				
Up to 90 min.	0.098	0.046	4.554	0.033
91–150 min.	0.201	0.048	17.541	<0.001
151–240 min.	−0.161	0.054	8.746	0.003
Age group (Ref = 45–54 years)				
25–34	−0.207	0.062	11.233	<0.001
35–44	0.124	0.040	9.660	0.002
Household type (Ref = single parent)				
Couple	0.111	0.097	1.300	0.254
Number of children (Ref = two or more)				
One	−0.253	0.029	75.191	<0.001
Age of the youngest child (Ref = 11–14 years)				
0–2 years	−0.365	0.051	52.219	<0.001
3–5 years	−0.225	0.049	20.919	<0.001
6–10 years	−0.031	0.051	0.369	0.544
Educational level (Ref = low)				
High	0.066	0.046	2.045	0.153
Medium	−0.118	0.038	9.589	0.002
Employment status (Ref = unemployed)				
Employed	−0.314	0.046	46.335	<0.001
Life satisfaction (Ref = satisfied)				
Dissatisfied	0.244	0.045	29.424	<0.001
Neither satisfied nor dissatisfied	−0.015	0.043	0.117	0.732
Survey month (Ref = August–October)				
November–January	0.179	0.045	16.044	<0.001
February–April	−0.200	0.048	17.368	<0.001
May–July	−0.042	0.052	0.672	0.413
Week day (Ref = Saturday–Sunday)				
Monday–Friday	−0.093	0.031	8.778	0.003
Episode hour of start (Ref = 16–24)				
0–8	0.407	0.055	55.846	<0.001
9–15	−0.073	0.042	2.967	0.085

Source: Authors' calculations based on ITUS data (2013/4).

Table A.3:
Logit model – Overall childcare, women

	Parameter estimate	Standard error	Wald Chi square	Prob > Chi square
Intercept	−0.162	0.035	21.591	<0.001
Multitasking (Ref = not)				
Yes	0.081	0.020	15.856	<0.001
Episode duration (Ref = >60 min.)				
Less than 20 min.	0.089	0.024	13.215	<0.001
20–39 min.	0.085	0.027	10.191	0.001
40–59 min.	−0.121	0.040	9.337	0.002
Presence of an adult (Ref = not)				
Yes	−0.216	0.017	152.801	<0.001
Daily childcare time (Ref = >240 min.)				
Up to 90 min.	0.058	0.031	3.638	0.056
91–150 min.	−0.106	0.027	15.206	<0.001
151–240 min.	−0.092	0.025	14.125	<0.001
Age group (Ref = 45–54 years)				
25–34	−0.071	0.027	6.970	0.008
35–44	−0.120	0.022	31.197	<0.001
Household type (Ref = single parent)				
Couple	0.002	0.022	0.012	0.913
Number of children (Ref = two or more)				
One	−0.182	0.015	142.583	<0.001
Age of the youngest child (Ref = 11–14 years)				
0–2 years	−0.137	0.029	21.848	<0.001
3–5 years	−0.240	0.028	75.687	<0.001
6–10 years	0.139	0.026	28.659	<0.001
Educational level (Ref = low)				
High	0.098	0.024	17.287	<0.001
Medium	−0.053	0.020	7.494	0.006
Employment status (Ref = unemployed)				
Employed	−0.015	0.016	0.892	0.345
Life satisfaction (Ref = satisfied)				
Dissatisfied	0.141	0.023	38.331	<0.001
Neither satisfied nor dissatisfied	−0.005	0.023	0.042	0.838
Survey month (Ref = August–October)				
November–January	−0.077	0.024	9.849	0.002
February–April	−0.026	0.025	1.124	0.289
May–July	−0.033	0.026	1.700	0.192
Week day (Ref = Saturday–Sunday)				
Monday–Friday	0.024	0.018	1.914	0.166
Episode hour of start (Ref = 16–24)				
0–8	0.375	0.027	195.945	<0.001
9–15	−0.095	0.021	20.729	<0.001

Source: Authors' calculations based on ITUS data (2013/4).

Table A.4:
Logit model – Routine care, men

	Parameter estimate	Standard error	Wald Chi square	Prob > Chi square
Intercept	0.052	0.314	0.027	0.869
Multitasking (Ref = not)				
Yes	0.362	0.133	7.380	0.007
Episode duration (Ref = >60 min.)				
Less than 20 min.	0.070	0.099	0.495	0.482
20–39 min.	−0.035	0.105	0.115	0.734
40–59 min.	0.472	0.148	10.190	0.001
Presence of an adult (Ref = not)				
Yes	−0.098	0.055	3.185	0.074
Daily childcare time (Ref = >240 min.)				
Up to 90 min.	0.042	0.097	0.190	0.663
91–150 min.	0.240	0.091	6.892	0.009
151–240 min.	−0.377	0.101	13.987	<0.001
Age group (Ref = 45–54 years)				
25–34	−0.318	0.107	8.834	0.003
35–44	0.037	0.076	0.235	0.628
Household type (Ref = single parent)				
Couple	0.359	0.255	1.990	0.158
Number of children (Ref = two or more)				
One	−0.196	0.056	12.436	<0.001
Age of the youngest child (Ref = 11–14 years)				
0–2 years	−0.368	0.135	7.418	0.006
3–5 years	−0.262	0.132	3.898	0.048
6–10 years	−0.355	0.151	5.538	0.019
Educational level (Ref = low)				
High	−0.062	0.088	0.493	0.482
Medium	−0.134	0.075	3.178	0.075
Employment status (Ref = unemployed)				
Employed	−0.062	0.095	0.431	0.512
Life satisfaction (Ref = satisfied)				
Dissatisfied	0.210	0.089	5.517	0.019
Neither satisfied nor dissatisfied	0.139	0.085	2.647	0.104
Survey month (Ref = August–October)				
November–January	0.490	0.090	29.684	<0.001
February–April	−0.100	0.097	1.062	0.303
May–July	−0.068	0.102	0.448	0.503
Week day (Ref = Saturday–Sunday)				
Monday–Friday	−0.113	0.062	3.289	0.070
Episode hour of start (Ref = 16–24)				
0–8	0.622	0.094	43.609	<0.001
9–15	−0.144	0.088	2.702	0.100

Source: Authors' calculations based on ITUS data (2013/4).

Table A.5:
Logit model – Routine care, women

	Parameter estimate	Standard error	Wald Chi square	Prob > Chi square
Intercept	0.122	0.070	3.009	0.083
Multitasking (Ref = not)				
Yes	0.266	0.046	33.476	<0.001
Episode duration (Ref = >60 min.)				
Less than 20 min.	0.116	0.043	7.288	0.007
20–39 min.	0.186	0.044	17.673	<0.001
40–59 min.	–0.144	0.066	4.689	0.030
Presence of an adult (Ref = not)				
Yes	–0.190	0.027	47.927	<0.001
Daily childcare time (Ref = >240 min.)				
Up to 90 min.	–0.118	0.054	4.772	0.029
91–150 min.	–0.087	0.049	3.190	0.074
151–240 min.	–0.017	0.043	0.167	0.682
Age group (Ref = 45–54 years)				
25–34	0.026	0.046	0.319	0.572
35–44	–0.098	0.040	5.980	0.014
Household type (Ref = single parent)				
Couple	0.054	0.038	2.067	0.151
Number of children (Ref = two or more)				
One	–0.121	0.025	23.321	<0.001
Age of the youngest child (Ref = 11–14 years)				
0–2 years	–0.295	0.054	29.463	<0.001
3–5 years	–0.184	0.055	11.184	<0.001
6–10 years	0.171	0.057	9.119	0.003
Educational level (Ref = low)				
High	–0.026	0.038	0.449	0.503
Medium	0.024	0.033	0.523	0.470
Employment status (Ref = unemployed)				
Employed	–0.015	0.028	0.274	0.601
Life satisfaction (Ref = satisfied)				
Dissatisfied	0.171	0.038	20.253	<0.001
Neither satisfied nor dissatisfied	–0.001	0.039	0.001	0.971
Survey month (Ref = August–October)				
November–January	–0.149	0.041	12.917	0.000
February–April	–0.014	0.042	0.117	0.732
May–July	0.062	0.043	2.119	0.146
Week day (Ref = Saturday–Sunday)				
Monday–Friday	0.043	0.027	2.442	0.118
Episode hour of start (Ref = 16–24)				
0–8	0.401	0.040	101.724	<0.001
9–15	–0.143	0.037	15.044	<0.001

Source: Authors’ calculations based on ITUS data (2013/4).

Table A.6:
Logit model – Recreational care, men

	Parameter estimate	Standard error	Wald Chi square	Prob > Chi square
Intercept	−0.912	0.220	17.133	<0.001
Multitasking (Ref = not)				
Yes	0.295	0.075	15.519	<0.001
Episode duration (Ref = >60 min.)				
Less than 20 min.	0.306	0.088	12.210	<0.001
20–39 min.	0.190	0.089	4.588	0.032
40–59 min.	−0.353	0.125	7.965	0.005
Presence of an adult (Ref = not)				
Yes	−0.127	0.053	5.794	0.016
Daily childcare time (Ref = >240 min.)				
Up to 90 min.	0.204	0.094	4.706	0.030
91–150 min.	0.370	0.096	14.989	<0.001
151–240 min.	−0.267	0.113	5.565	0.018
Age group (Ref = 45–54 years)				
25–34	−0.155	0.103	2.269	0.132
35–44	0.011	0.073	0.024	0.877
Household type (Ref = single parent)				
Couple	0.180	0.186	0.932	0.334
Number of children (Ref = two or more)				
One	−0.321	0.054	34.895	<0.001
Age of the youngest child (Ref = 11–14 years)				
0–2 years	−0.179	0.100	3.208	0.073
3–5 years	−0.085	0.100	0.732	0.392
6–10 years	−0.082	0.105	0.611	0.435
Educational level (Ref = low)				
High	0.369	0.087	18.003	<0.001
Medium	−0.226	0.071	10.079	0.001
Employment status (Ref = unemployed)				
Employed	−0.460	0.084	29.935	<0.001
Life satisfaction (Ref = satisfied)				
Dissatisfied	0.306	0.082	13.805	<0.001
Neither satisfied nor dissatisfied	−0.226	0.084	7.319	0.007
Survey month (Ref = August–October)				
November–January	0.075	0.084	0.796	0.372
February–April	−0.128	0.093	1.905	0.168
May–July	0.016	0.093	0.031	0.860
Week day (Ref = Saturday–Sunday)				
Monday–Friday	−0.184	0.058	10.154	<0.001
Episode hour of start (Ref = 16–24)				
0–8	0.191	0.137	1.966	0.161
9–15	−0.125	0.099	1.594	0.207

Source: Authors' calculations based on ITUS data (2013/4).

Table A.7:
Logit model – Recreational care, women

	Parameter estimate	Standard error	Wald Chi square	Prob > Chi square
Intercept	−0.708	0.062	130.295	<0.001
Multitasking (Ref = not)				
Yes	0.295	0.029	101.412	<0.001
Episode duration (Ref = >60 min.)				
Less than 20 min.	0.234	0.046	25.345	<0.001
20–39 min.	0.059	0.051	1.311	0.252
40–59 min.	−0.158	0.073	4.702	0.030
Presence of an adult (Ref = not)				
Yes	−0.260	0.033	61.303	<0.001
Daily childcare time (Ref = >240 min.)				
Up to 90 min.	0.166	0.054	9.246	0.002
91–150 min.	−0.109	0.051	4.644	0.031
151–240 min.	−0.165	0.048	11.956	<0.001
Age group (Ref = 45–54 years)				
25–34	−0.083	0.052	2.587	0.108
35–44	−0.107	0.040	7.080	0.008
Household type (Ref = single parent)				
Couple	0.054	0.042	1.655	0.198
Number of children (Ref = two or more)				
One	−0.151	0.029	26.254	<0.001
Age of the youngest child (Ref = 11–14 years)				
0–2 years	−0.025	0.056	0.191	0.662
3–5 years	−0.149	0.053	7.945	0.005
6–10 years	0.161	0.050	10.320	0.001
Educational level (Ref = low)				
High	0.220	0.045	24.298	<0.001
Medium	−0.054	0.036	2.161	0.142
Employment status (Ref = unemployed)				
Employed	−0.048	0.031	2.396	0.122
Life satisfaction (Ref = satisfied)				
Dissatisfied	0.149	0.045	11.116	<0.001
Neither satisfied nor dissatisfied	−0.118	0.044	7.148	0.008
Survey month (Ref = August–October)				
November–January	−0.071	0.046	2.386	0.122
February–April	−0.112	0.048	5.315	0.021
May–July	−0.054	0.048	1.238	0.266
Week day (Ref = Saturday–Sunday)				
Monday–Friday	−0.041	0.032	1.606	0.205
Episode hour of start (Ref = 16–24)				
0–8	0.277	0.054	26.092	<0.001
9–15	−0.111	0.042	6.867	0.009

Source: Authors' calculations based on ITUS data (2013/4).

Table A.8:
Enjoyment scores OLS estimates

			Enjoyment score		H0: LSMean1=LSMean2		
			Mean	**Standard error**	**Effect**	***t* value**	**Pr > |*t*|**
			Overall childcare				
Model 1 Men	Multitasking	1 = Yes	1.73	0.086	−0.02	−0.27	0.785
		2 = Not	1.75	0.061			
Model 2 Women	Multitasking	1 = Yes	1.42	0.032	−0.16	−5.90	<0.001
		2 = Not	1.58	0.022			
			Routine care				
Model 3 Men	Multitasking	1 = Yes	1.30	0.214	−0.40	−2.65	<0.001
		2 = Not	1.70	0.156			
Model 4 Women	Multitasking	1 = Yes	1.15	0.066	−0.39	−6.54	<0.001
		2 = Not	1.54	0.038			
			Recreational care				
Model 5 Men	Multitasking	1 = Yes	1.83	0.098	−0.24	−3.64	<0.001
		2 = Not	2.07	0.083			
Model 6 Women	Multitasking	1 = Yes	1.61	0.042	−0.38	−11.01	<0.001
		2 = Not	1.99	0.036			

Source: Authors' calculations based on ITUS data (2013/4).

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 383–415

Well-being in Europe: Decompositions by country and gender for the population aged 50+

***Claudia Reiter*[1] *and Sonja Spitzer*[1,2,*]**

Abstract

The well-being of older Europeans is of increasing importance given the substantial ageing of the population. This paper comprehensively analyses well-being for the population aged 50+ in 26 European countries, using the newly proposed indicator "Years of Good Life" (YoGL), which measures the remaining years of life that an individual can expect to live in a "good" state. The indicator enables the decomposition of well-being into various dimensions, thereby revealing important heterogeneities between regions and genders. Results show that numbers of YoGL at age 50 vary considerably between European countries. They are highest in Northern and Western European countries and lowest in Central and Eastern European countries, where many "good" years are lost due to low life satisfaction. Interestingly, the high life expectancy levels in Southern Europe do not translate into higher numbers of YoGL, mainly due to the low levels of physical and cognitive health in this region. While women and men can expect to have similar numbers of YoGL, women are likely to spend a smaller proportion of their longer remaining lifetime in a good state. These results demonstrate the importance of using well-being indicators that consider population heterogeneity when measuring human well-being, especially for older populations.

Keywords: well-being; population 50+; Europe; SHARE data; Sullivan's method, cross-country analysis

[1]Department of Demography, University of Vienna, Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Vienna, Austria

[2]The authors are listed in alphabetical order with equal contributions.

[*]Correspondence to: Sonja Spitzer, sonja.spitzer@univie.ac.at

DOI: 10.1553/populationyearbook2021.res4.1

1 Introduction

Longevity is one of the biggest achievements of modern societies. In the past 20 years, the European population has, on average, gained more than five years of life expectancy. Consequently, old- and middle-aged people represent an ever-increasing proportion of the population in Europe, a trend that is further exacerbated by declining fertility rates and reinforcing migration patterns in some countries. In 2019, 39.6% of all Europeans were aged 50 or older (United Nations 2019). This share is expected to further increase: according to projections by the Wittgenstein Centre for Demography and Global Human Capital (2018), the majority of the European population will be aged 50+ by the end of the century.

In the light of these developments, gaining a deeper understanding of well-being in older age groups, as well as of the main factors that influence it, is not only interesting for researchers, but is highly relevant for policy-makers, and for society as a whole. Although the number of studies that seek to measure and quantify quality of life has been increasing, there is little existing research that has focused on the various aspects of well-being in older age groups. In particular, few studies have examined well-being from a more comprehensive perspective by going beyond a sole focus on the health-related aspects of quality of life. In this paper, we analyse comprehensive well-being based on a newly proposed indicator called "Years of Good Life" (YoGL), which considers mortality and physical and cognitive health, but also other aspects of quality of life, including poverty and subjective well-being. Furthermore, we contribute to the literature by decomposing well-being into various dimensions, thereby revealing old-age-specific country and gender differences in well-being for 26 European countries.

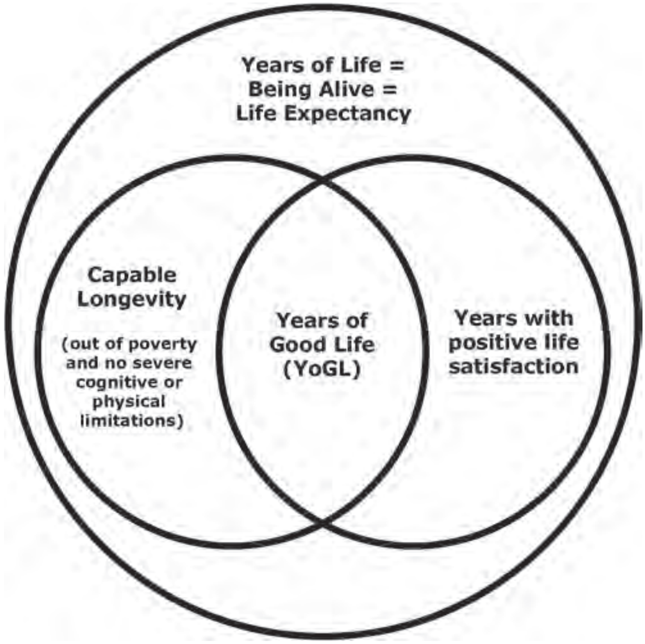
The remainder of the paper is structured as follows. Section 2 outlines the conceptual approach, and relates it to previous literature. Section 3 describes the data utilised, and Section 4 introduces the methods employed. Results and robustness analyses are presented and discussed in Section 5. Section 6 concludes and discusses potential limitations.

2 Conceptualisation and related work

In this paper, we make use of the novel comprehensive well-being indicator YoGL, a measure developed by Lutz et al. (2021), which aims at estimating the remaining years of life an individual can expect to live in a "good" state. By focusing on the changing composition of human populations with regard to the characteristics that reflect a society's aggregate level of human well-being, the indicator is tailor-made to serve as an outcome that can be used to judge the sustainability of long-term development trajectories (Matson et al. 2016).

The measure is based on the assumption that in order to be able to enjoy any quality of life, a person has to be alive. However, since mere survival does not sufficiently capture well-being, "years of good life" are made conditional on meeting minimum

Figure 1:
Dimensions of Years of Good Life



Note: Based on Lutz et al. (2021).

standards of objectively observable conditions (capable longevity), as well as of subjective life satisfaction. Building on earlier works of Sen (Desai et al. 1992), the objective conditions measuring "capable longevity" are further broken down into three separate dimensions: being out of poverty, being cognitively enabled, and having no serious physical disabilities. Only if people are above critical levels in all three objective dimensions and in their self-reported overall life satisfaction are the life years considered as "good" years in the calculation of YoGL. Figure 1 summarises this structure and the basic logic of the YoGL indicator, demonstrating that it represents the intersection of capable longevity and years with positive life satisfaction as a subset of overall years of life.[1]

Our analysis focuses on the expected years of good life at age 50 for men and women in 26 European countries. As people age, their physical and mental capabilities decline, even as their levels of financial dependence and social isolation

[1] More detailed information about the theoretical foundations of the indicator, as well as a comparison between YoGL and other existing well-being indicators, can be found in Lutz et al. (2021) and Reiter and Lutz (2020).

increase. Thus, the elderly are particularly vulnerable to experiencing limitations in their capable longevity and deterioration in their life satisfaction (Netuveli and Blane 2008). While we acknowledge that most 50-year-olds are still in their prime of life, are active on the labour market, and are in good health, the methodology of YoGL allows us to not only capture age-specific well-being at age 50, but to take into account expected well-being in later years of life (for a more detailed description of the methodology and calculation of YoGL, see Section 4.1).

The previous studies that analysed the well-being of older adults often had a strong focus on health dimensions; i.e., they used summary measures for population health as measures of well-being. Prominent examples of these indicators include Quality Adjusted Life Expectancy (Zeckhauser and Shepard 1976), Disability Adjusted Life Expectancy (Murray and Lopez 1994), and Healthy Life Years (Robine 2006) – all of which were calculated by combining life table information with information on the prevalence of health states or diseases. Analyses based on these indicators have found considerable heterogeneities in health-related well-being between European countries, as well as between sub-populations by age and gender (Heijink et al. 2011; Jagger et al. 2008).

Overall, we expect the prevalence of the objective dimensions of YoGL – i.e., capable longevity – to decrease with rising age. Individual ageing leads to a gradual decrease in physical and mental capacity and a growing risk of disease (World Health Organization 2018). Thus, as a population ages, a smaller proportion of the total population is cognitively enabled and free of serious physical disabilities. In addition, for older people, particularly for older women, the risk of poverty increases with age (Eurostat 2020c). However, in many highly developed European countries, extensive welfare systems and efficient redistribution measures ensure that the actual proportion of the population living below the poverty line is kept to a minimum, including among the elderly population.

The relationship between ageing and subjective life satisfaction is more ambiguous, with previous research suggesting that average levels of life satisfaction may decrease, increase, or remain relatively constant across the lifespan (Horley and Lavery 1995; Lelkes 2008; Steptoe et al. 2015). For example, in a study on self-assessed life satisfaction among the elderly population in Sweden, Austria and Germany, Kutubaeva (2019) found no evidence that ageing itself necessarily worsens an individual's perception of life. By contrast, Angelini et al. (2012) came to the conclusion that age influences life satisfaction among Europeans through two counterbalancing channels: i.e., on the one hand, increasing age leads to an increase in the perceived level of life satisfaction; and, on the other, it results in a shift in the individual thresholds that determine whether an individual is satisfied with her life. Nevertheless, although there are inconsistencies in the previous research findings, a certain impact of old-age-specific burdens (e.g., dependency or reduced social contact through isolation or the deaths of friends and family) on life satisfaction cannot be ruled out.

The indicator YoGL is designed in such a way that it does not apply a standard weighting structure to its constituents or rely on implicitly assumed trade-offs

(Lutz et al. 2021; Reiter and Lutz 2020). Therefore, a high level of life satisfaction does not, for example, counterbalance severe physical limitations. On the contrary, when an individual's life satisfaction is low solely because of her strong functional limitations, "double-counting" is also avoided by design. While we appreciate that YoGL is a substantively justified combined measure, we also want to highlight the usefulness of decomposing the indicator into its four dimensions (see Section 4.2). This is especially relevant when analysing older age groups, as the challenges posed by population ageing at various levels of society are crucial concerns for policy-makers. By gaining a better understanding of why a population is losing good years of life, sustainable prevention strategies can be formulated to help individuals preserve their well-being in old age.

3 Data

The analysis is based on two different data sources. Individual-level information on the objective and subjective dimensions of YoGL is based on the Survey of Health, Ageing and Retirement in Europe (SHARE). Additional aggregated data for the model-based out-of-sample predictions of physical health, as well as country-specific life tables, are taken from Eurostat.

3.1 Survey of Health, Ageing and Retirement in Europe

We analyse YoGL for Europe based on SHARE, a multidisciplinary cross-country survey providing micro-level information on the health, well-being and socio-economic characteristics of around 140,000 individuals, based on data collected in around 380,000 interviews (Börsch-Supan et al. 2013). These survey data are especially well-suited for analysing YoGL in European countries: first, because the data are ex-ante harmonised; and, second, because the data include information on all four dimensions needed to compute the indicator. The survey's target population consists of all non-institutionalised individuals aged 50 and older. This analysis is based on the most recent Wave 7 from 2017, which includes European countries as well as Israel (Bergmann et al. 2019; Börsch-Supan 2019b). We keep all European individuals, which results in 70,191 observations from 26 countries: namely, Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Luxembourg, Malta, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden and Switzerland. In addition, information from Wave 5 (2013) is utilised for out-of-sample predictions of physical health (Börsch-Supan 2019a; Malter and Börsch-Supan 2015). Wave 5 includes only 14 European countries: namely, Austria, Belgium, Czechia, Denmark, Estonia, France, Germany, Italy, Luxembourg, Netherlands, Slovenia, Spain, Sweden and Switzerland. In total, they provide 59,713 observations.

Before describing in detail the operationalisation of the different YoGL dimensions, we should stress that the concept of the indicator is based on one specific interpretation of human well-being that focuses on longevity combined with being above minimum thresholds in further dimensions of well-being. This focus on basic material needs, basic physical and mental functioning, and having at least a medium level of life satisfaction is highly sensitive to improvements at the tail end of the population distribution, but is insensitive to improvements in the upper parts. Our decision to focus on the tail end rather than the mean of the distribution is in line with the "leave no-one behind" principle, which is also promoted in the Sustainable Development Goals (UN System Chief Executives Board for Coordination 2017).

3.1.1 Objective YoGL dimensions

Being out of poverty. Following Lutz et al. (2021), individuals are considered poor if their income falls below the World Bank poverty line for upper-middle income countries of US$5.50 PPP per day (Ferreira and Sánchez-Páramo 2017). More specifically, poverty is assessed based on an individual's total monthly household income after taxes and contributions. Income is equivalised by employing the square root scale that is also used in OECD (2011) and OECD (2008), in which household income is divided by the square root of the household size. It is then adjusted for differences in purchasing power and converted to international dollars per day. Table 1 displays the shares of men and women who are out of poverty in each country. It shows that in Europe, most individuals aged 50 and older are above the World Bank poverty line, except in poor countries such as Bulgaria, Latvia and Romania. Table 2 provides summary statistics for income, as well as for all of the other variables utilised.

Without cognitive limitations. Cognitive limitations are measured via a memory performance test. At the beginning of the test, the interviewer reads aloud a list of 10 words. The participants are then asked to repeat as many of these words as possible within one minute. On average, individuals recall 5.2 words. Memory scores are highest in Switzerland (6.0 words) and lowest in Portugal (3.9 words), which is also the only country in which no participant recalled more than eight words. As suggested in the previous literature, participants are considered cognitively impaired if they recall only three words or less (Grodstein et al. 2001; Purser et al. 2005; Spitzer and Weber 2019). For robustness analyses, the threshold is set at two words or less (Section 5.3). Again, descriptive results and summary statistics are provided in Tables 1 and 2.

Without physical limitations. Finally, physical limitations are assessed via a chair stand performance test. The test is introduced by the interviewer saying: "The next test measures the strength and endurance in your legs. I would like you to fold your arms across your chest and sit so that your feet are on the floor; then stand up keeping your arms folded across your chest. Like this..." The survey participants are asked if they think it would be safe to try to stand up from a chair, and the

Table 1:
Overview of well-being dimensions by country for the population aged 50+ (2017, Wave 7)

Country	Out of poverty %		Without cognitive limitations %		Without physical limitations %		With positive life satisfaction %		Above threshold in all YoGL dimensions %	
	M	W	M	W	M	W	M	W	M	W
Austria	100	100	94.1	92.2	87.6	81.8	91.9	91.4	78.4	73.3
Belgium	100	100	90.1	89.8	88.3	82.4	93.5	92.0	77.3	72.5
Bulgaria	98.8	98.0	77.8	76.1	77.0	70.6	68.0	57.2	49.8	39.3
Croatia	99.2	99.3	82.0	80.6	76.2	67.5	75.1	70.4	53.0	46.6
Cyprus	100	100	70.5	71.7	81.0	73.6	90.9	85.6	57.8	53.4
Czechia	100	100	89.4	92.1	82.4	77.1	86.4	80.1	68.2	60.7
Denmark	100	100	89.7	93.4	91.6	89.7	94.3	94.1	80.2	81.5
Estonia	100	99.9	83.7	87.3	78.1	71.1	69.4	74.4	52.3	52.8
Finland	100	100	85.5	90.0	91.2	87.5	95.9	94.6	77.2	77.6
France	100	100	89.2	88.7	88.2	83.5	88.3	86.2	73.1	68.4
Germany	100	100	90.5	93.6	87.4	83.7	87.4	87.5	72.8	72.1
Greece	100	100	88.5	84.5	81.8	72.5	86.7	82.3	67.3	55.9
Hungary	100	100	83.9	82.4	79.8	68.4	81.2	67.1	61.1	43.9
Italy	100	100	80.3	76.8	82.6	75.2	91.3	87.6	65.5	56.7
Latvia	97.6	98.6	82.6	82.7	78.6	71.7	70.9	65.4	52.1	45.7
Lithuania	99.7	99.4	75.9	78.8	77.1	69.5	65.2	62.1	43.1	41.8
Luxembourg	100	100	87.9	86.0	88.5	82.9	90.1	87.4	75.1	66.4
Malta	100	100	79.1	80.8	82.7	76.5	94.2	90.3	65.4	61.1
Poland	100	99.9	77.0	79.0	78.3	73.4	78.7	75.5	53.9	51.6
Portugal	100	100	62.1	54.2	82.3	77.9	90.3	86.0	49.6	42.7
Romania	91.4	89.5	73.1	65.3	71.7	61.4	84.8	78.2	50.3	39.5
Slovakia	100	100	83.8	78.2	83.6	75.6	86.3	82.0	65.4	56.0
Slovenia	100	100	85.9	85.1	81.0	75.1	84.0	81.8	63.3	59.5
Spain	100	100	75.4	74.5	80.0	73.2	88.8	85.9	57.4	54.4
Sweden	100	100	88.9	90.8	92.2	89.1	93.5	93.7	78.7	78.2
Switzerland	100	99.9	93.2	93.3	91.7	89.1	96.9	94.9	84.2	80.5

Note: Cross-sectional individual SHARE survey weights are applied.

Table 2:
Summary statistics (2017, Wave 7)

	N	Mean	SD	Min.	Max.
Age (in number of years)	69,897	66.61	10.156	51	105
Gender (1 = woman)	69,897	0.54	0.498	0	1
Yearly equivalised household income (in Intl$)	69,400	18,408	14,565	428	417,137
Number of words recalled	69,181	5.19	1.812	0	10
Predicted ability to stand up from a chair	65,450	0.81	0.218	0	1
Life satisfaction	69,897	7.56	1.818	0	10

Note: Cross-sectional individual SHARE survey weights are applied.

test is conducted only if they agree. Following Lutz et al. (2021), individuals are considered to be free from physical limitations if they are able to stand up from the chair without using their arms. By contrast, participants are considered to be physically impaired if they are unable to stand up from the chair, if they have to use their arms to stand up from the chair, or if they think that it is unsafe to participate in the test in the first place. Previous literature has shown that the share of participants who are categorised as physically impaired hardly changes when individuals who use their arms to stand up are considered to be unimpaired (Spitzer and Weber 2019).

The chair stand test was conducted in Wave 2 (2006-2007) and in Wave 5 (2013), but not in Wave 7 (2017). Hence, we use information from Wave 5, the most recent wave in which the chair stand test was administered, to extrapolate data to Wave 7. Details on the methods and robustness analyses are provided in Sections 4.3 and 5.3. Based on these out-of-sample predictions, 81% of the participants are considered to be without physical limitations. Descriptive results and summary statistics are again provided in Tables 1 and 2, along with information for the robustness analyses in the Appendix.

3.1.2 Subjective dimension

Subjective well-being is assessed via a standard 10-step Likert-scale based on the question: “On a scale from 0 to 10 where 0 means completely dissatisfied and 10 means completely satisfied, how satisfied are you with your life?” On average over all of the countries in the sample, individuals reported a value of 7.6 (Table 2). Individuals are considered to be satisfied if they reported a value larger than five. For robustness analyses, a cut-off of four is also considered (Section 5.3).

After assessing the objective and subjective dimensions, we add a binary code to each observation that indicates whether the individual is above the critical threshold in all four YoGL dimensions. Finally, country-specific proportions by 10-year age groups[2] and gender are aggregated using calibrated cross-sectional survey weights, as provided by SHARE. We find that, on average across all of the countries in the sample, 63.6% of individuals are simultaneously out of poverty, report positive life satisfaction, have basic cognitive abilities, and have no severe activity limitations. Simple correlation matrices for all dimensions per country and gender are provided in Table A.1 in the Appendix.

3.2 Eurostat

3.2.1 Eurostat data on health expenditure

For the model-based out-of-sample predictions (see Section 4.3), country-level information on health expenditures is utilised. These data are provided by Eurostat (2020a). In particular, we utilise total general government expenditure on health as a percentage of GDP in the respective year; i.e., 2013 for Wave 5 and 2017 for Wave 7.

3.2.2 Eurostat life tables

Gender-specific period life tables for the year 2017 for all 26 countries in our sample also come from Eurostat (2020b). As these life tables are reported by single-year age groups, standard life table techniques were applied to transform them into abridged life tables with 10-year age intervals.

4 Method

4.1 Sullivan method

The calculation of YoGL is based on demographic life table methods (Sullivan method) in which age-specific person-years lived at each age are multiplied by age-specific proportions considered to be above the critical threshold in all four YoGL dimensions (Sullivan 1971). Consequently, a year is only counted as a good year for those individuals who are above the critical thresholds in all four dimensions; and

---

[2] Using five-year age groups would reduce the number of countries for which we can estimate the remaining years of good life at age 50, since SHARE data do not include observations of men and women for the age group 50–54 for Hungary, Portugal and Sweden. Instead, 10-year age groups are used to calculate prevalence rates and life tables. This has the additional benefit of increasing the number of observations for each age group. Sensitivity analyses revealed that the results hardly differ depending on whether five-year age groups or 10-year age groups are used.

no trade-offs or compensatory mechanisms are assumed between the dimensions. Summing up the age-specific person-years of good life for all of the remaining age groups above the age at which remaining life expectancy will be assessed results in the expected total years of good life. Formally, the calculation can be summarised in the following mathematical notation:

$$\frac{1}{l_x}\sum_{i=x}^{A}\pi_i L_i \tag{1}$$

where l_x denotes the number of survivors at age x_i (beginning of the interval i); L_i indicates the number of person-years lived in the age group i; and π_i denotes the prevalence of the state of interest, i.e., country-specific proportions of the population above a critical level in all four dimensions by 10-year age groups and gender.

As with total life expectancy, YoGL can be assessed at birth, as well as at any other age considered appropriate. As the focus of this paper is, however, on the well-being of older adults, we calculate YoGL at age 50.

4.2 Decomposition

By decomposing YoGL into four different dimensions, we are able to better understand why a population is losing good years of life. The methodology we apply here resembles the calculation of YoGL using the Sullivan method; however, in this case, the prevalence of the state of interest is decomposed into four different proportions: proportion of the population who are out of poverty (${\pi_i}^p$), proportion of the population with basic cognitive ability (${\pi_i}^c$), proportion of the population without severe activity limitations (${\pi_i}^h$) and proportion of the population with positive life satisfaction (${\pi_i}^s$).

This approach enables us to calculate four sub-indicators, which we call “years of life out of poverty”, “years of life without severe activity limitations”, “years of life with basic cognitive ability” and “years of life with positive life satisfaction”. Given that the YoGL indicator is based on no implicit trade-off assumptions or weighting structures, it must hold that the number of years of good life is less than or equal to each of the four sub-indicators for every (sub-)population.

4.3 Model-based out-of-sample predictions

The YoGL calculations are based on data from Wave 7, because it is the most recent wave, and includes more countries than any previous SHARE survey wave. Wave 7 does not, however, include information on the performance-tested ability to stand up from a chair. Thus, the data are extrapolated from Wave 5, which is the most recent SHARE wave in which the chair stand test was conducted. For the out-of-sample predictions, the information on the ability to stand up from a chair from Wave 5 (2013) is regressed on crucial determinants on both the individual and the country level. These determinants include demographic and socio-economic characteristics, information on physical and cognitive health (most importantly, the self-reported ability to stand up from a chair), country-level health expenditure, region dummies, as well as the other three well-being dimensions (being out of poverty, having positive life satisfaction and basic cognitive ability). Using the same determinants from Wave 7 (2017) enables us to predict levels of physical health for 2017. A detailed list of variables along with summary statistics is provided in Table A.2 in the Appendix.

Since the outcome variable is binary, a logistic regression model is utilised for the regression. Standard errors are clustered at the household level. The output table is provided in the Appendix (Table A.3). Overall, the non-linear extrapolation model yields a relatively high Pseudo R^2 of 0.27. The model fits the underlying data well, as shown in Figure A.1 in the Appendix. In all of the countries that participated in Wave 5, the share of individuals with physical limitations based on the predicted data is virtually identical to the share based on the observed data. Exceptions include Czechia, Luxembourg, Austria and Spain; but even in those countries, the difference never exceeds four percentage points. Nevertheless, we conduct robustness analyses in which we compare YoGL based on predicted physical limitations with YoGL based on observed physical limitations for Wave 5 (Section 5.3).

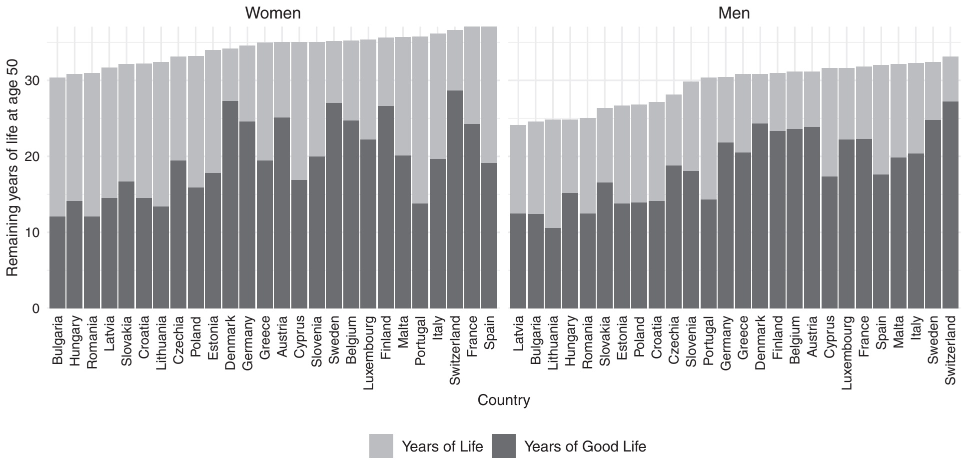
5 Results

5.1 Country and gender differences

Figure 2 displays YoGL at age 50 by country and gender. It also shows life expectancy at age 50, and ranks countries accordingly, which allows us to analyse the relationship between remaining years of life and remaining years of good life. On average, the YoGL at age 50 are 19.7 years for women and 18.6 years for men. The results also show, however, that the numbers of YoGL vary considerably between countries. In 2017, 50-year-old Swiss women can expect to live another 36.6 years, of which 28.7 are considered good years. Hence, these women can expect to spend 78.4% of their remaining lifetime as good years. By contrast, 50-year-old Bulgarian women have a life expectancy of 30.4 additional years in 2017, of which just 12.1 years are considered good years, resulting in a proportion of only 39.8%. The findings for men are very similar. The projected numbers of YoGL for men are highest in Switzerland (27.2 years) and are lowest in Lithuania (10.6 years) and Bulgaria (12.5 years).

Overall, life expectancy at age 50 is highest in Southern Europe and Switzerland, followed by Northern and other Western European countries. For example, 50-year-old women in Spain can expect to live another 37.1 years, and 50-year-old men in Switzerland can expect to live another 33.1 years. The high life expectancy in

Figure 2:
Remaining Years of Good Life at age 50 by country and gender (2017, Wave 7)

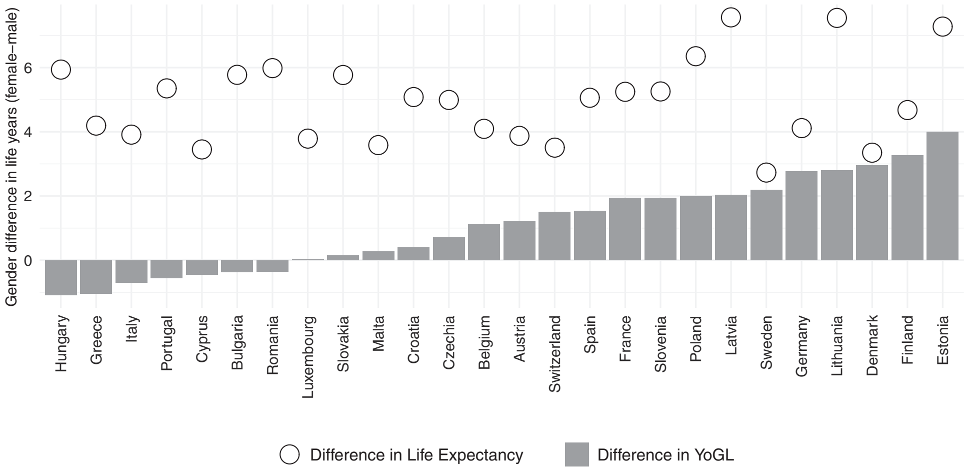


Note: Cross-sectional individual SHARE survey weights are applied.

Southern Europe does not, however, translate into more YoGL. Indeed, despite its relatively high life expectancy, especially for women, Portugal is among the countries with the lowest numbers of YoGL. Instead, the Northern and Western European countries are the leaders in YoGL for both men and women. The highest numbers of YoGL are observed in Switzerland and in Northern European countries such as Sweden and Denmark, followed by in other Western European countries. Consequently, these are also the countries with the highest proportions of remaining years of good life compared to the total remaining years of life. However, Central and Eastern European countries have both low life expectancy and lower numbers of YoGL. These differences between overall life expectancy and YoGL confirm the importance of differentiating between mere survival and well-being.

Figure 3 shows the gender differences in YoGL and life expectancy at age 50 by country. In general, women have higher life expectancy than men. While this pattern holds for all countries, the gender gap in life expectancy is largest in Latvia (7.6 years) and Lithuania (7.5 years), and is smallest in Sweden (2.7 years). On average over all of the countries in our sample, women's life expectancy exceeds men's life expectancy by 4.9 years. Interestingly, however, the differences in YoGL are very small. On average, the gap in YoGL between women and men is only 1.1 years. This suggests that women's advantage in life expectancy does not translate into an equal advantage in good years. The only exceptions to this general pattern are Denmark and Sweden, where the gender gap in YoGL is only slightly smaller than the gender gap in life expectancy. In seven countries (Hungary, Greece, Italy,

Figure 3:
Gender differences in Years of Good Life and life expectancy at age 50 by country (2017, Wave 7)



Note: Cross-sectional individual SHARE survey weights are applied.

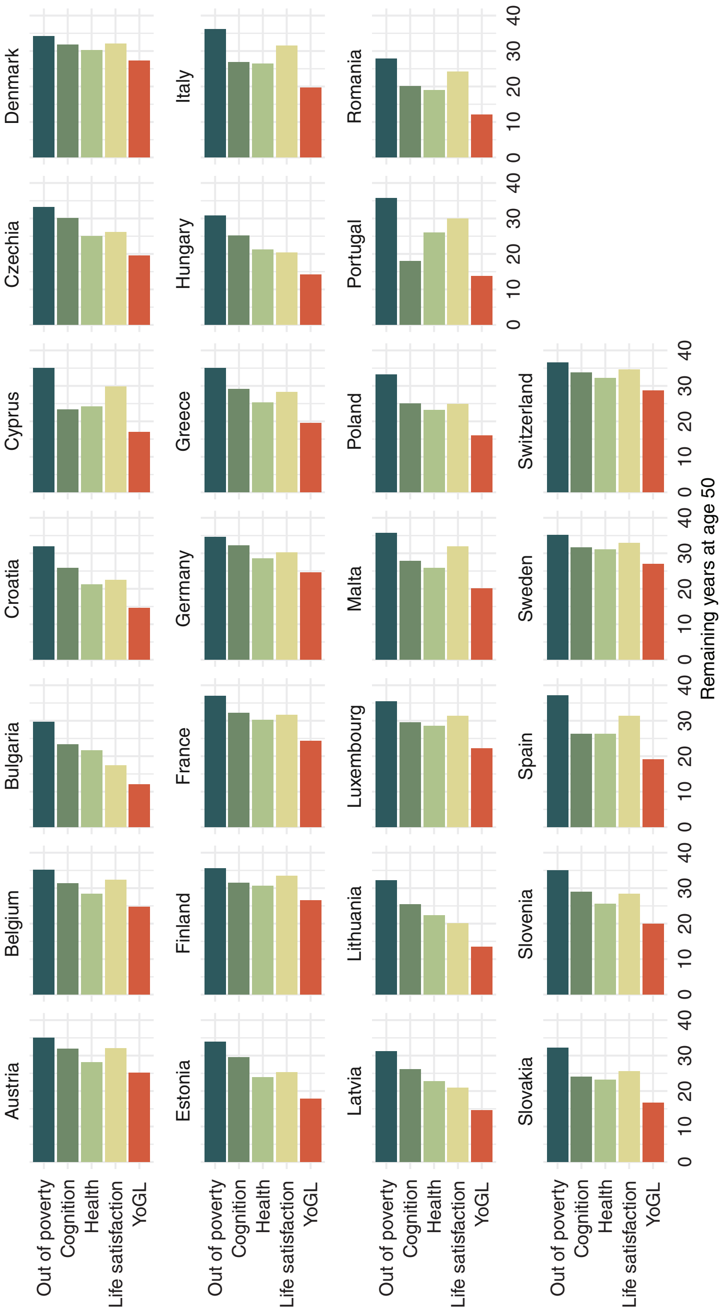
Portugal, Cyprus, Bulgaria and Romania), the numbers of YoGL at age 50 are even lower for women than for men.

As a result of the gender disparities noted above, women have, on average, a much lower proportion of remaining years of good life than of remaining years of life. While men at age 50 can expect, on average, to spend 62.7% of their remaining time as good years, the corresponding share for women is just 57.1%. The proportions of remaining years of good life are particularly low for women in Portugal (38.6%), Romania (39.1%) and Bulgaria (39.8%). These results are related to the "male-female health-mortality paradox" (di Lego et al. 2020), a term that is used to refer to the phenomenon that, on average, women live longer than men, but spend a larger proportion of their life in poorer health (Luy and Minagawa 2014; Pongiglione et al. 2015).

5.2 Decomposition

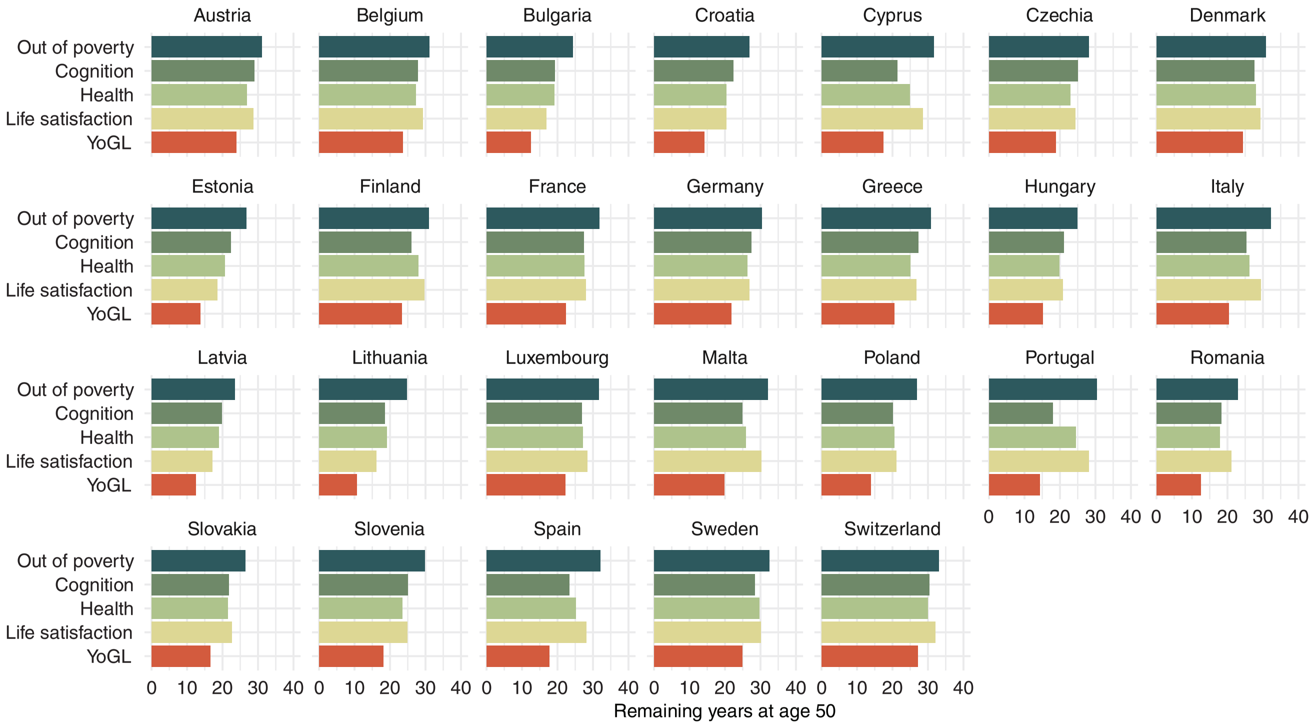
Decomposing YoGL into four dimensions allows us to better understand why countries are losing good years of life. The results presented in Figures 4 and 5 clearly show that, overall, most good years are lost due to limitations in physical health, followed by limitations in cognitive health. By contrast, poverty has only small effects on YoGL, since most Europeans are considered to be out of poverty according to the definitions described above. The countries with the lowest numbers

Figure 4:
Decomposing remaining Years of Good Life at age 50 by country, women (2017, Wave 7)



Note: Cross-sectional individual SHARE survey weights are applied.

Figure 5:
Decomposing remaining Years of Good Life at age 50 by country, men (2017, Wave 7)



Note: Cross-sectional individual SHARE survey weights are applied.

of years of life out of poverty are all Central and Eastern European countries; but even in those countries, more YoGL are lost due to deductions in the other dimensions. Related literature has, however, suggested that poverty plays an important role in the lower levels of well-being observed in low-income countries outside of Europe, and in trends in YoGL over time (Reiter and Lutz 2020; United Nations Development Programme 2019).

Analysing the decomposition separately by country and gender reveals that there are big differences in the composition of YoGL. While the reduction in good years due to low life satisfaction plays a small role in most countries, it is highly relevant for Central and Eastern European countries. In Bulgaria, Hungary, Lithuania and Latvia, it is even the main reason for the reduction in good years; thus, in these countries, life satisfaction has a bigger impact on YoGL than physical and cognitive health. These results are very much in line with previous findings showing that in Central and Eastern European countries, life satisfaction is persistently low (Dingemans and Henkens 2018; Guriev and Zhuravskaya 2009). This pattern has often been attributed to the impact of the post-communist transition and the related increases in inequality and uncertainty (Djankov et al. 2016; Sanfey and Teksoz 2007).

Women lose most of their YoGL due to their physical limitations. Among men, however, the negative effects of physical and cognitive limitations are almost identical. These results have to be interpreted against the backdrop of differences in overall life expectancy. As we mentioned above, women spend a larger proportion of their overall lifetime in poor health. Similar results are found for life satisfaction. While women spend more years with positive life satisfaction than men, they spend a smaller proportion of their lifetime with positive life satisfaction than men.

5.3 Robustness analyses

Sensitivity analyses displayed in Table A.4 and Figures A.2, A.3 and A.4 in the Appendix show that the results discussed above are robust. As we discussed in Section 3.1, we consider different thresholds for physical limitations, cognitive limitations and subjective well-being. Furthermore, we compare YoGL based on predicted and observed physical limitations for countries that participated in Wave 5.

The results of the main analysis show that 63.6% of the population are simultaneously out of poverty, report positive life satisfaction, have basic cognitive ability and suffer from no severe activity limitations. This share hardly changes when the robustness analyses are applied. Decreasing the threshold for cognitive limitations to two words instead of three words increases the share of individuals who are above the threshold in all YoGL dimensions to 68.6%; while decreasing the threshold for positive life satisfaction from five to four increases the share to 68.7%.

As can be seen in Figure A.2, changing the threshold for cognitive limitations leaves the results virtually unchanged for most countries. Although YoGL increase slightly when the more lenient threshold is applied, the country ranking and related results are hardly affected. One exception is Portugal, where YoGL increases substantially based on the new threshold, especially for men. The reason for this increase is that cognitive ability in Portugal has a different distribution than it does in the other countries. First, as mentioned above, Portugal is the only country in which no participant remembered more than eight words – in all other countries, the maximum is 10 words. Second, while the number of words recalled is normally distributed in most countries, the Portuguese distribution has an additional concentration around three words. Hence, whether the threshold is set at two or three words makes a greater difference for Portugal than it does for the other countries. When we look at the distribution of the number of words in previous waves in Portugal, this curious distribution does not appear. Thus, we conclude that the results for Portugal need to be treated with caution.

Decreasing the threshold for positive life satisfaction also increases YoGL slightly for all countries simultaneously; however, since the country ranking hardly changes (Figure A.3), the conclusions drawn from the results are virtually unaffected.

Figure A.4 shows the difference in the YoGL when using predicted values of physical limitations rather than observed values of physical limitations for Wave 5, which covered a much smaller group of countries than Wave 7. The results show that when the predicted rather than the observed values are used, the YoGL hardly change, and the country ranking remains exactly the same for both men and women. We thus conclude that our predictions are robust.

6 Conclusion

Given that the proportion of the population aged 50+ is increasing in European countries, the well-being of older people is an issue that is gaining in relevance. Our aim in the current paper was to contribute to our understanding of quality of life in older age groups by analysing well-being in the population aged 50+ in 26 European countries. Based on the novel indicator "Years of Good Life" (YoGL), we estimated the remaining years of life that an individual can expect to live in a "good" state. Our results uncovered considerable differences in YoGL at age 50 between regions and genders, with the numbers of YoGL being highest in Northern and Western European countries, and lowest in Central and Eastern European countries. While the absolute numbers of YoGL were found to be quite similar for women and men, women were shown to spend a smaller proportion of their higher remaining lifetime in a good state.

In addition to looking at the overall well-being of older people, we further decomposed the indicator into its four dimensions in order to gain a better understanding of why a given population is losing good years of life. Our results show that in Central and Eastern Europe, the numbers of years spent with positive life satisfaction are particularly low; while in Southern Europe, the loss of YoGL is mainly attributable to low levels of physical and cognitive health. Taking this population heterogeneity into account when measuring human well-being may also help policy-makers in formulating sustainable prevention strategies aimed at helping individuals maintain their levels of well-being in old age. Depending on the specific dimensions from which country-specific deficiencies in well-being emerge, policies may seek to increase opportunities for lifelong learning, reduce barriers to social and societal participation in later life, or improve health literacy among older adults. By shedding light on the existing deficiencies in the well-being of older Europeans, and on how these deficiencies differ between countries and between men and women, we hope to inform the development of policies that ultimately aim to improve the quality of life of the elderly in a more holistic way.

Although a range of sensitivity analyses suggest that our results are robust, this study has potential limitations. As our estimates are largely based on survey data, our results are heavily dependent on the selection of participants. For example, highly educated individuals are more likely to participate in surveys, which could, in turn, distort the results through the non-representativeness of the population (Spitzer 2020). Furthermore, our results are built on an indicator that is tailor-made for global analyses, with the thresholds for each dimension designed to reflect the diverse development stages of countries worldwide. While this approach is helpful when making global comparisons, it may lose some of its informative value when the thresholds are applied to a specific set of countries only. For example, given that the World Bank poverty line for upper-middle income countries is US$5.50, it would appear that virtually no European lives in poverty. Thus, our estimates hardly reflect the relatively high poverty risk that older women in Europe face. Future research may expand our estimates of well-being in the population aged 50+ to a global scale.

Acknowledgements

The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no 741105). This paper uses data from SHARE Waves 5 and 7 (DOIs: 10.6103/SHARE.w5.700, and 10.6103/SHARE.w7.700). The SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982) and Horizon 2020 (SHARE-DEV3: GA N°676536, SERISS: GA N°654221) and by DG Employment, Social Affairs & Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

References

Angelini, V., D. Cavapozzi, L. Corazzini and O. Paccagnella 2012. Age, health and life satisfaction among older Europeans. *Social Indicators Research* 105(2): 293–308. https://doi.org/10.1007/s11205-011-9882-x

Bergmann, M., A. Scherpenzeel and A. Börsch-Supan (eds) 2019. *SHARE Wave 7 Methodology: Panel Innovations and Life Histories*. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Börsch-Supan, A. 2019a. *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 7.0.0. SHARE-ERIC. Data set*. Retrieved from 10.6103/SHARE.w5.700

Börsch-Supan, A. 2019b. *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7. Release version: 7.0.0. SHARE-ERIC. Data set*. Retrieved from 10.6103/SHARE.w7.700

Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, ... S. Zuber 2013. Data resource profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology* 42(4): 992–1001.

Desai, M., A. Sen and J. Boltvinik 1992. *Social Progress Index: A Proposal*. Bogotá: United Nations Development Programme.

di Lego, V., P. Di Giulio and M. Luy 2020. The male-female health-mortality paradox. *Encyclopedia of Geronotology and Population Aging*, eds D. Gu and M. Dupre. Springer.

Dingemans, E. and K. Henkens 2018. Life satisfaction of working retirees in Europe. *Netspar Academic Series*, (DP 06/2018-032).

Djankov, S., E. Nikolova and J. Zilinsky 2016. The happiness gap in Eastern Europe. *Journal of Comparative Economics* 44(1): 108–124. https://doi.org/10.1016/j.jce.2015.10.006

Eurostat 2020a. General government expenditure by function (COFOG). Retrieved April 8, 2020, from https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=gov_10a_exp&lang=en

Eurostat 2020b. Life table. Retrieved April 7, 2020, from https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_mlifetable&lang=en

Eurostat 2020c. People at risk of poverty or social exclusion. Retrieved April 27, 2020, from https://ec.europa.eu/eurostat/databrowser/view/t2020_50/default/table?lang=en

Ferreira, F. and C. Sánchez-Páramo 2017. A richer array of international poverty lines. Retrieved April 9, 2020, from https://blogs.worldbank.org/developmenttalk/richer-array-international-poverty-lines

Grodstein, F., J. Chen, R. S. Wilson and J. E. Manson 2001. Type 2 diabetes and cognitive function in community-dwelling elderly women. *Diabetes Care* 24(6): 1060–1065.

Guriev, S. and E. Zhuravskaya 2009. (Un)Happiness in Transition. *Journal of Economic Perspectives Figure* 23(2): 143–168. https://doi.org/10.1257/jep.23.2.143

Heijink, R., P. Van Baal, M. Oppe, X. Koolman and G. Westert 2011. Decomposing cross-country differences in quality adjusted life expectancy: The impact of value sets. *Population Health Metrics* 9. https://doi.org/10.1186/1478-7954-9-17

Horley, J. and J. J. Lavery 1995. Subjective well-being and age. *Social Indicators Research* 34: 275–282. https://doi.org/10.1007/BF01079200

Jagger, C., C. Gillies, F. Moscone, E. Cambois, H. Van Oyen, W. Nusselder and J. M. Robine 2008. Inequalities in healthy life years in the 25 countries of the European Union in 2005: A cross-national meta-regression analysis. *The Lancet* 372(9656): 2124–2131. https://doi.org/10.1016/S0140-6736(08)61594-9

Kutubaeva, R. Z. 2019. Analysis of life satisfaction of the elderly population on the example of Sweden, Austria and Germany. *Population and Economics* 3(3): 102–116. https://doi.org/10.3897/popecon.3.e47192

Lelkes, O. 2008. Happiness across the life cycle: Exploring age-specific preferences. *European Centre Policiy Brief* (2).

Lutz, W., E. Striessnig, A. Dimitrova, S. Ghislandi, A. Lijadi, C. Reiter, S. Spitzer and D. Yildiz 2021. Years of Good Life (YoGL) is a wellbeing indicator designed to serve research on sustainability. *Proceedings of the National Academy of Sciences of the United States of America*, 118(12): e1907351118. https://doi.org/10.1073/pnas.1907351118

Luy, M. and Y. Minagawa 2014. Gender gaps - life expectancy and proportion of life in poor health. *Health Reports* 25(12): 12–19.

Malter, F. and A. Börsch-Supan (eds) 2015. *SHARE Wave 5: Innovations & Methodology*. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Matson, P., W. C. Clark and K. Andersson 2016. *Pursuing sustainability: A guide to the science and practice*. Princeton University Press.

Murray, C. J. and A. D. Lopez 1994. Quantifying disability: Data, methods and results. *Bulletin of the World Health Organization* 72(3): 481–494.

Netuveli, G. and D. Blane 2008. Quality of life in older ages. *British Medical Bulletin* 85(1): 113–126. https://doi.org/10.1093/bmb/ldn003

OECD 2008. *Growing unequal? Income distribution and poverty in OECD countries*. Retrieved from https://read.oecd-ilibrary.org/social-issues-migration-health/growing-unequal_9789264044197-en#page1

OECD 2011. *Divided we stand: Why inequality keeps rising*. Paris: OECD Publishing. https://doi.org/10.1787/9789264119536-en

Pongiglione, B., B. L. De Stavola and G. B. Ploubidis 2015. A systematic literature review of studies analyzing inequalities in health expectancy among the older population. *PLoS ONE* 10(6): 1–21. https://doi.org/10.1371/journal.pone.0130747

Purser, J. L., G. G. Fillenbaum, C. F. Pieper and R. B. Wallace 2005. Mild cognitive impairment and 10-year trajectories of disability in the Iowa established populations for epidemiologic studies of the elderly cohort. *Journal of the American Geriatrics Society* 53(11): 1966–1972.

Reiter, C. and W. Lutz 2020. Survival and years of good life in Finland in the very long run. *Finnish Yearbook of Population Research* 54: 1–27. https://doi.org/10.23979/fypr.87148

Robine, J.-M. 2006. Summarizing health status. *Oxford Handbook of Public Health Practice* (2nd ed.), eds D. Pencheon, C. Guest, D. Melzer and J. Gray. Oxford: Oxford University Press.

Sanfey, P. and Teksoz, U. 2007. Does transition make you happy? *Economics of Transition* 15(4): 707–731. https://doi.org/10.1111/j.1468-0351.2007.00309.x

Spitzer, S. 2020. Biases in health expectancies due to educational differences in survey participation of older Europeans: It's worth weighting for. *European Journal of Health Economics* 21: 573–605. https://doi.org/10.1007/s10198-019-01152-0

Spitzer, S. and D. Weber 2019. Reporting biases in self-assessed physical and cognitive health status of older Europeans. *PLoS ONE* 14(10): e0223526. https://doi.org/10.1371/journal.pone.0223526

Steptoe, A., A. Deaton and A. A. Stone 2015. Subjective wellbeing, health, and ageing. *The Lancet* 385: 640–648. https://doi.org/10.1016/S0140-6736(13)61489-0

Sullivan, D. F. 1971. A single index of mortality and morbidity. *HSMHA Health Reports* 86(4): 347–354.

UN System Chief Executives Board for Coordination 2017. *Leaving no one behind: Equality and Non-discrimination at the Heart of Sustainable Development*. New York. Retrieved August 10, 2020, from https://www.unsceb.org/CEBPublicFiles/CEB%20equality%20framework-A4-web-rev3.pdf

United Nations 2019. World population prospects 2019. Retrieved May 13, 2020, from https://population.un.org/wpp/

United Nations Development Programme 2019. *Human Development Report 2019: Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century*.

Wittgenstein Centre for Demography and Global Human Capital 2018. Wittgenstein Centre Data Explorer Version 2.0 (Beta). Retrieved February 22, 2020, from http://www.wittgensteincentre.org/dataexplorer

World Health Organization 2018. Ageing and health. Retrieved April 28, 2020, from https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

Zeckhauser, R. and D. Shepard 1976. Where now for saving lives? *Law and Contemporary Problems* 40(4): 5–45.

Appendix

Table A.1:
Simple correlation matrices of well-being dimensions by country and gender (2017, Wave 7)

		Men				Women			
Country	Variable	Income per day	Outcome chair stand test	Outcome memory test	Subjective well-being	Income per day	Outcome chair stand test	Outcome memory test	Subjective well-being
Austria	Income per day	1.000				1.000			
	Outcome chair stand test	0.150	1.000			0.107	1.000		
	Outcome memory test	0.224	0.360	1.000		0.132	0.400	1.000	
	Subjective well-being	0.104	0.317	0.049	1.000	0.026	0.290	0.102	1.000
Belgium	Income per day	1.000				1.000			
	Outcome chair stand test	0.188	1.000			0.203	1.000		
	Outcome memory test	0.220	0.269	1.000		0.270	0.381	1.000	
	Subjective well-being	0.151	0.227	0.034	1.000	0.187	0.260	0.093	1.000
Bulgaria	Income per day	1.000				1.000			
	Outcome chair stand test	0.344	1.000			0.384	1.000		
	Outcome memory test	0.264	0.306	1.000		0.270	0.367	1.000	
	Subjective well-being	0.305	0.451	0.187	1.000	0.323	0.422	0.202	1.000
Croatia	Income per day	1.000				1.000			
	Outcome chair stand test	0.175	1.000			0.254	1.000		
	Outcome memory test	0.176	0.387	1.000		0.237	0.417	1.000	
	Subjective well-being	0.212	0.240	0.142	1.000	0.197	0.320	0.207	1.000
Cyprus	Income per day	1.000				1.000			
	Outcome chair stand test	0.110	1.000			0.204	1.000		
	Outcome memory test	0.088	0.460	1.000		0.149	0.531	1.000	
	Subjective well-being	0.057	0.287	0.094	1.000	0.102	0.285	0.186	1.000

Continued

Table A.1:
Continued

		Men				Women			
Country	**Variable**	**Income per day**	**Outcome chair stand test**	**Outcome memory test**	**Subjective well-being**	**Income per day**	**Outcome chair stand test**	**Outcome memory test**	**Subjective well-being**
Czechia	Income per day	1.000				1.000			
	Outcome chair stand test	0.207	1.000			0.258	1.000		
	Outcome memory test	0.211	0.360	1.000		0.251	0.374	1.000	
	Subjective well-being	0.158	0.225	0.135	1.000	0.182	0.273	0.138	1.000
Denmark	Income per day	1.000				1.000			
	Outcome chair stand test	0.126	1.000			0.079	1.000		
	Outcome memory test	0.135	0.254	1.000		0.058	0.291	1.000	
	Subjective well-being	0.009	0.250	0.049	1.000	0.003	0.299	0.091	1.000
Estonia	Income per day	1.000				1.000			
	Outcome chair stand test	0.263	1.000			0.331	1.000		
	Outcome memory test	0.304	0.360	1.000		0.307	0.416	1.000	
	Subjective well-being	0.224	0.351	0.165	1.000	0.233	0.346	0.130	1.000
Finland	Income per day	1.000				1.000			
	Outcome chair stand test	0.236	1.000			0.175	1.000		
	Outcome memory test	0.290	0.368	1.000		0.254	0.318	1.000	
	Subjective well-being	0.193	0.314	0.122	1.000	0.083	0.239	0.032	1.000
France	Income per day	1.000				1.000			
	Outcome chair stand test	0.105	1.000			0.092	1.000		
	Outcome memory test	0.181	0.408	1.000		0.105	0.407	1.000	
	Subjective well-being	0.125	0.296	0.131	1.000	0.098	0.273	0.120	1.000

Continued

Table A.1: Continued

		Men				Women			
Country	**Variable**	**Income per day**	**Outcome chair stand test**	**Outcome memory test**	**Subjective well-being**	**Income per day**	**Outcome chair stand test**	**Outcome memory test**	**Subjective well-being**
Germany	Income per day	1.000				1.000			
	Outcome chair stand test	0.186	1.000			0.212	1.000		
	Outcome memory test	0.224	0.266	1.000		0.250	0.365	1.000	
	Subjective well-being	0.281	0.264	0.111	1.000	0.239	0.243	0.139	1.000
Greece	Income per day	1.000				1.000			
	Outcome chair stand test	-0.001	1.000			0.044	1.000		
	Outcome memory test	0.026	0.443	1.000		0.043	0.445	1.000	
	Subjective well-being	0.044	0.278	0.169	1.000	0.050	0.287	0.172	1.000
Hungary	Income per day	1.000				1.000			
	Outcome chair stand test	0.174	1.000			0.280	1.000		
	Outcome memory test	0.162	0.373	1.000		0.269	0.465	1.000	
	Subjective well-being	0.275	0.416	0.207	1.000	0.310	0.390	0.261	1.000
Italy	Income per day	1.000				1.000			
	Outcome chair stand test	0.064	1.000			0.091	1.000		
	Outcome memory test	0.137	0.429	1.000		0.104	0.438	1.000	
	Subjective well-being	0.062	0.310	0.199	1.000	0.079	0.305	0.199	1.000
Latvia	Income per day	1.000				1.000			
	Outcome chair stand test	0.321	1.000			0.358	1.000		
	Outcome memory test	0.258	0.324	1.000		0.323	0.397	1.000	
	Subjective well-being	0.282	0.337	0.140	1.000	0.292	0.318	0.225	1.000

Continued

Table A.1:
Continued

		Men				Women			
Country	**Variable**	**Income per day**	**Outcome chair stand test**	**Outcome memory test**	**Subjective well-being**	**Income per day**	**Outcome chair stand test**	**Outcome memory test**	**Subjective well-being**
Lithuania	Income per day	1.000				1.000			
	Outcome chair stand test	0.283	1.000			0.316	1.000		
	Outcome memory test	0.207	0.379	1.000		0.293	0.444	1.000	
	Subjective well-being	0.264	0.305	0.119	1.000	0.291	0.283	0.146	1.000
Luxembourg	Income per day	1.000				1.000			
	Outcome chair stand test	0.152	1.000			0.219	1.000		
	Outcome memory test	0.213	0.316	1.000		0.255	0.333	1.000	
	Subjective well-being	0.187	0.296	0.205	1.000	0.113	0.257	0.055	1.000
Malta	Income per day	1.000				1.000			
	Outcome chair stand test	0.203	1.000			0.054	1.000		
	Outcome memory test	0.202	0.338	1.000		0.168	0.376	1.000	
	Subjective well-being	0.041	0.157	0.011	1.000	0.044	0.206	0.011	1.000
Poland	Income per day	1.000				1.000			
	Outcome chair stand test	0.217	1.000			0.242	1.000		
	Outcome memory test	0.227	0.377	1.000		0.251	0.469	1.000	
	Subjective well-being	0.229	0.259	0.177	1.000	0.180	0.257	0.139	1.000
Portugal	Income per day	1.000				1.000			
	Outcome chair stand test	0.197	1.000			0.113	1.000		
	Outcome memory test	0.221	0.395	1.000		0.236	0.380		
	Subjective well-being	0.268	0.292	0.052	1.000	0.225	0.261		

Continued

Table A.1: Continued

Country	Variable	Men				Women			
		Income per day	Outcome chair stand test	Outcome memory test	Subjective well-being	Income per day	Outcome chair stand test	Outcome memory test	Subjective well-being
Romania	Income per day	1.000				1.000			
	Outcome chair stand test	0.335	1.000			0.401	1.000		
	Outcome memory test	0.193	0.404	1.000		0.281	0.461	1.000	
	Subjective well-being	0.185	0.335	0.120	1.000	0.234	0.307	0.176	1.000
Slovakia	Income per day	1.000				1.000			
	Outcome chair stand test	0.134	1.000			0.146	1.000		
	Outcome memory test	0.105	0.321	1.000		0.060	0.422	1.000	
	Subjective well-being	0.095	0.459	0.195	1.000	0.104	0.410	0.110	1.000
Slovenia	Income per day	1.000				1.000			
	Outcome chair stand test	0.229	1.000			0.295	1.000		
	Outcome memory test	0.255	0.414	1.000		0.338	0.477	1.000	
	Subjective well-being	0.284	0.309	0.176	1.000	0.294	0.324	0.231	1.000
Spain	Income per day	1.000				1.000			
	Outcome chair stand test	0.217	1.000			0.273	1.000		
	Outcome memory test	0.305	0.456	1.000		0.359	0.524	1.000	
	Subjective well-being	0.104	0.252	0.180	1.000	0.137	0.323	0.209	1.000
Sweden	Income per day	1.000				1.000			
	Outcome chair stand test	0.256	1.000			0.290	1.000		
	Outcome memory test	0.317	0.315	1.000		0.328	0.292	1.000	
	Subjective well-being	0.134	0.213	0.016	1.000	0.134	0.223	0.010	1.000
Switzerland	Income per day	1.000				1.000			
	Outcome chair stand test	0.024	1.000			0.045	1.000		
	Outcome memory test	0.117	0.313	1.000		0.128	0.309	1.000	
	Subjective well-being	0.053	0.244	0.088	1.000	0.056	0.215	0.099	1.000

Table A.2:
Summary statistics for the predictor variables (2013, Wave 5)

	N	Mean	SD	Min.	Max.
Out of poverty (1 = yes)	58,947	1.00	0.014	0	1
Cognitively able (1 = yes)	58,823	0.84	0.362	0	1
Satisfied with life (1 = yes)	59,377	0.85	0.353	0	1
Gender (1 = woman)	59,377	0.54	0.499	0	1
Age (in number of years)	59,377	66.11	10.645	51	104
Low education (1 = yes)	58,504	0.42	0.493	0	1
Medium education (1 = yes)	58,504	0.37	0.484	0	1
High education (1 = yes)	58,504	0.21	0.406	0	1
Retired (1 = yes)	58,874	0.51	0.500	0	1
Married (1 = yes)	57,447	0.66	0.474	0	1
Partner in same household (1 = yes)	59,377	0.66	0.473	0	1
Household size	59,377	2.10	1.001	1	11
Wealth quintile	59,377	2.94	1.431	1	5
Subjective ability chair stand (1 = able)	59,362	0.82	0.383	0	1
Number of chronic diseases	59,280	1.21	1.272	0	10
Number of instrumental activity limitations	59,359	0.50	1.528	0	13
Heart attack (1 = yes)	59,280	0.11	0.309	0	1
Diabetes (1 = yes)	59,280	0.12	0.329	0	1
Lung disease (1 = yes)	59,280	0.06	0.246	0	1
Cancer (1 = yes)	59,280	0.06	0.235	0	1
Stroke (1 = yes)	59,280	0.03	0.179	0	1
Parkinson (1 = yes)	59,280	0.01	0.082	0	1
Alzheimer (1 = yes)	59,280	0.01	0.077	0	1
Walking 100 metres (1 = able)	59,362	0.10	0.305	0	1
Sitting for two hours (1 = able)	59,362	0.10	0.303	0	1
Climbing stairs (1 = able)	59,362	0.12	0.320	0	1
Stooping, kneeling or crouching (1 = able)	59,362	0.30	0.458	0	1
Pulling or pushing large objects (1 = able)	59,362	0.13	0.333	0	1
Lifting weights over 5 kilos (1 = able)	59,362	0.20	0.399	0	1
Walking across a room (1 = able)	59,359	0.02	0.134	0	1
Getting in or out of bed (1 = able)	59,359	0.03	0.171	0	1
Doing work around house or garden (1 = able)	59,359	0.10	0.300	0	1
Region Northern Europe (1 = yes)	59,377	0.05	0.209	0	1
Region Southern Europe (1 = yes)	59,377	0.32	0.465	0	1
Region Central and Eastern Europe (1 = yes)	59,377	0.04	0.200	0	1
Region Western Europe (1 = yes)	59,377	0.60	0.491	0	1

Note: Cross-sectional individual SHARE survey weights are applied; region categories: Northern Europe (Sweden, Denmark, Finland), Southern Europe (Spain, Italy, Greece, Portugal, Cyprus, Malta), Central and Eastern Europe (Czechia, Poland, Hungary, Slovenia, Estonia, Croatia, Lithuania, Bulgaria, Latvia, Romania, Slovakia), Western Europe (Austria, Germany, Netherlands, France, Switzerland, Belgium, Ireland, Luxembourg)

Table A.3:
Output table prediction model for chair stand (2013, Wave 5)

	Coef.	SE
Self-reported ability chair stand (ref = unable)	0.539***	(0.037)
Women	0.001	(0.029)
Age 55-59 (ref = age 50-54)	–0.098	(0.065)
Age 60-64 (ref = age 50-54)	–0.053	(0.067)
Age 65-69 (ref = age 50-54)	–0.200**	(0.072)
Age 70-74 (ref = age 50-54)	–0.228**	(0.075)
Age 75-79 (ref = age 50-54)	–0.405***	(0.076)
Age 80-84 (ref = age 50-54)	–0.704***	(0.080)
Age 85-89 (ref = age 50-54)	–0.992***	(0.092)
Age 90+ (ref = age 50-54)	–1.253***	(0.136)
Medium education (ref = low education)	0.074*	(0.034)
High education (ref = low education)	0.190***	(0.044)
Retired (ref = no)	0.006	(0.040)
Married (ref = no)	0.117*	(0.059)
2nd wealth quintile (ref = 1st wealth quintile)	0.174***	(0.043)
3rd wealth quintile (ref = 1st wealth quintile)	0.192***	(0.045)
4th wealth quintile (ref = 1st wealth quintile)	0.284***	(0.047)
5th wealth quintile (ref = 1st wealth quintile)	0.358***	(0.049)
Number of chronic diseases	0.017	(0.016)
Number of instrumental activity limitations	–0.060**	(0.019)
Heart attack (ref = no)	–0.039	(0.046)
Diabetes (ref = no)	–0.268***	(0.044)
Lung disease (ref = no)	–0.081	(0.056)
Cancer (ref = no)	–0.179**	(0.057)
Alzheimer (ref = no)	0.271	(0.148)
Health expenditure in mio. Euro (country-level)	–0.006	(0.011)
Southern Europe (ref = Northern)	–0.825***	(0.062)
Central and Eastern Europe (ref = Northern)	–0.767***	(0.058)
Western Europe (ref = Northern)	–0.234***	(0.052)
Lives with partner (ref = no)	–0.040	(0.064)
Household size	0.031	(0.021)
Stroke (ref = no)	–0.284***	(0.069)
Parkinson (ref = no)	–0.437**	(0.137)
Walking 100 metres (ref = no)	–0.638***	(0.045)
Sitting for two hours (ref = no)	0.001	(0.044)

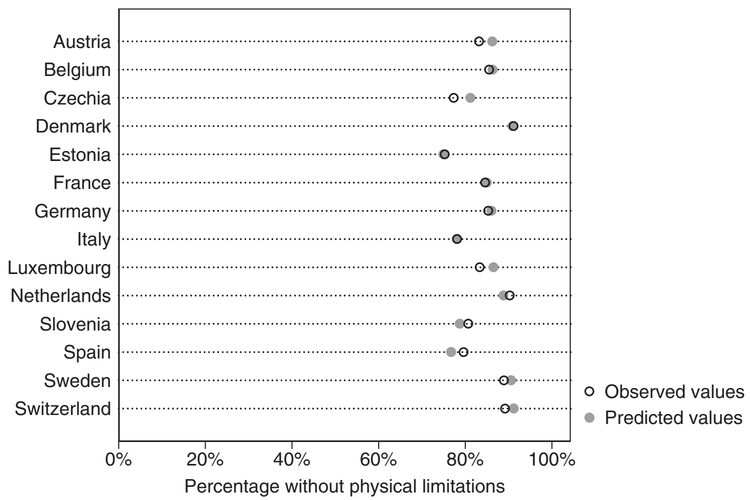
Continued

Table A.3:
Continued

	Coef.	SE
Climbing stairs (ref = no)	−0.652***	(0.042)
Stooping, kneeling or crouching (ref = no)	−0.589***	(0.034)
Pulling or pushing large objects (ref = no)	−0.376***	(0.042)
Lifting weights over 5 kilos (ref = no)	−0.282***	(0.039)
Walking across a room (ref = no)	−0.569***	(0.132)
Getting in or out of bed (ref = no)	−0.040	(0.083)
Doing work around house or garden (ref = no)	−0.438***	(0.056)
Cognitively able (ref = no)	0.571***	(0.038)
Out of poverty (ref = no)	2.921*	(1.279)
Satisfied with life (ref = no)	0.273***	(0.037)
Constant	−1.160	(1.287)
Pseudo R^2	0.271	
N	56,082	
SE	clustered	

Note: Dependent variable = tested ability chair stand. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure A.1:
Comparison of observed and predicted shares without physical limitations in the population aged 50+ (2013, Wave 5)



Note: Cross-sectional individual SHARE survey weights are applied.

Table A.4:
Overview of well-being dimensions robustness analyses (2017, Wave 7)

	Without cognitive limitations %				With positive life satisfaction %						Above threshold in all YoGL dimensions %			
	Original		Robustness		Original		Robustness		Original		Robustness cognition		Robustness satisfaction	
Country	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
Austria	94.1	92.2	97.8	96.2	91.9	91.4	96.7	97.4	78.5	73.5	80.8	75.3	81.7	76.9
Belgium	90.1	89.8	96.0	95.9	93.5	92.0	98.5	97.3	77.5	72.6	81.5	76.1	80.6	75.4
Bulgaria	77.8	76.1	93.5	90.8	68.0	57.2	83.9	78.1	49.8	39.3	55.8	43.3	57.8	50.7
Croatia	82.0	80.6	94.2	91.5	75.1	70.4	91.6	89.8	53.1	46.6	57.7	49.6	62.2	56.4
Cyprus	70.5	71.7	86.8	88.5	90.9	85.6	95.9	95.7	57.8	53.4	67.8	60.9	59.2	57.9
Czechia	89.4	92.1	97.5	97.3	86.4	80.1	94.9	96.4	68.3	60.8	71.6	62.6	73.2	72.1
Denmark	89.7	93.4	96.0	97.2	94.3	94.1	97.3	98.2	80.2	81.4	85.0	83.8	82.2	84.4
Estonia	83.7	87.3	91.6	94.0	69.4	74.4	89.0	91.2	52.4	52.9	55.2	55.4	64.3	61.8
Finland	85.5	90.0	93.4	95.3	95.9	94.6	98.6	98.8	77.3	77.6	83.6	81.0	79.1	80.9
France	89.2	88.7	95.0	95.0	88.3	86.2	97.1	96.5	73.2	68.5	76.6	71.9	79.1	74.5
Germany	90.5	93.6	97.1	97.2	87.4	87.5	95.3	96.5	72.9	72.2	76.6	74.0	77.9	78.1
Greece	88.5	84.5	95.7	93.5	86.7	82.3	95.5	92.7	67.3	56.0	70.9	59.7	72.8	61.3
Hungary	83.9	82.4	95.1	91.4	81.2	67.1	88.7	81.5	61.3	44.0	66.5	47.2	65.3	51.5
Italy	80.3	76.8	91.9	89.0	91.3	87.6	96.2	95.0	65.6	56.7	72.9	63.4	67.8	60.3
Latvia	82.6	82.7	91.7	91.7	70.9	65.4	87.6	85.3	52.2	45.8	56.7	48.8	62.0	56.7
Lithuania	75.9	78.8	89.5	90.5	65.2	62.1	86.8	86.5	43.2	41.8	49.0	44.9	55.4	54.2
Luxembourg	87.9	86.0	94.3	93.1	90.1	87.4	97.4	96.6	75.2	66.5	79.2	71.1	78.8	72.5
Malta	79.1	80.8	92.2	94.8	94.2	90.3	98.9	97.3	65.5	61.2	74.5	69.2	68.0	64.7

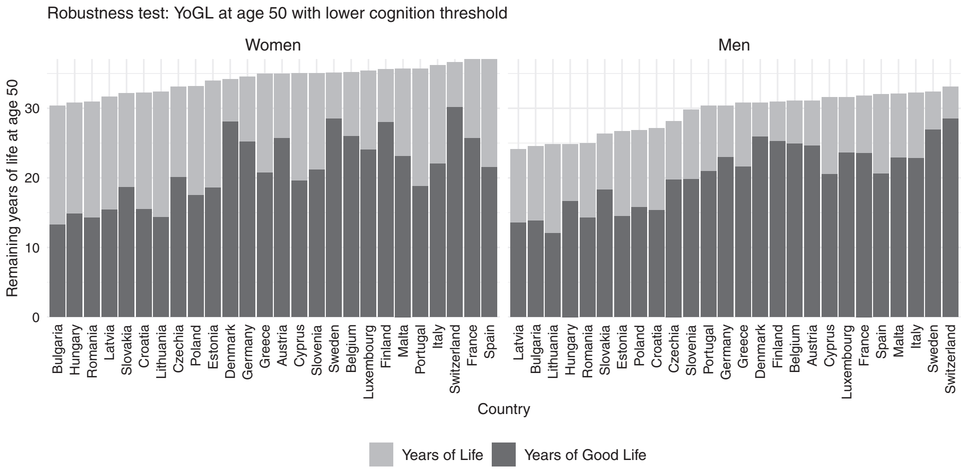
Continued

Table A.4: Continued

Country	Without cognitive limitations %				With positive life satisfaction %						Above threshold in all YoGL dimensions %			
	Original		Robustness		Original		Robustness		Original		Robustness cognition		Robustness satisfaction	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F
Poland	77.0	79.0	91.2	90.7	78.7	75.5	93.7	92.0	53.9	51.6	61.0	56.4	61.6	60.0
Portugal	62.1	54.2	90.6	76.3	90.3	86.0	94.9	93.7	49.7	42.8	70.7	57.8	51.2	45.8
Romania	73.1	65.3	88.6	83.8	84.8	78.2	92.2	89.6	50.4	39.6	58.1	46.8	53.1	43.8
Slovakia	83.8	78.2	94.4	91.0	86.3	82.0	94.9	93.0	65.3	56.0	71.9	62.2	70.2	62.1
Slovenia	85.9	85.1	95.8	93.2	84.0	81.8	97.0	95.7	63.4	59.6	69.0	62.9	71.4	66.2
Spain	75.4	74.5	89.2	87.2	88.8	85.9	97.1	94.6	57.5	54.6	66.4	61.1	62.4	58.2
Sweden	88.9	90.8	96.9	96.0	93.5	93.7	98.2	98.4	78.8	78.3	84.8	82.2	81.8	81.5
Switzerland	93.2	93.3	97.7	98.0	96.9	94.9	99.5	98.5	84.4	80.7	87.6	84.2	86.2	83.2

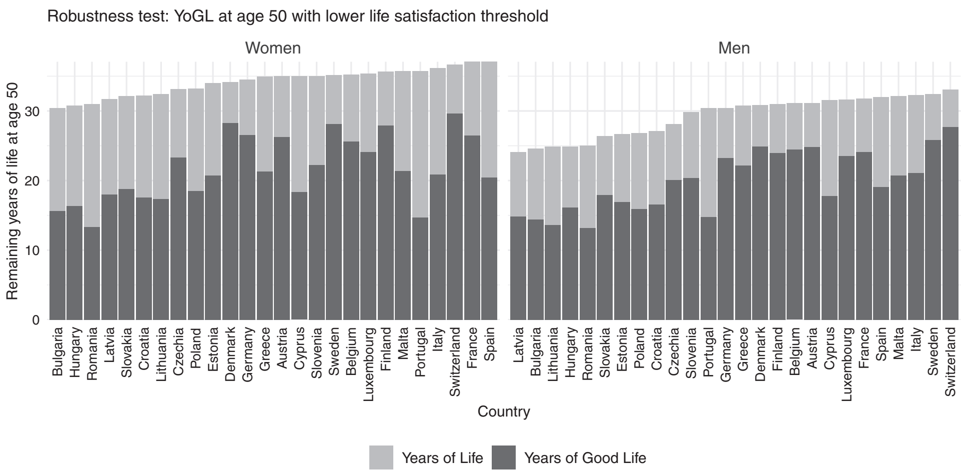
Note: Cross-sectional individual SHARE survey weights are applied

Figure A.2:
Remaining Years of Good Life at age 50 by country and gender – robustness analysis applying a different threshold for cognitive limitations (2017, Wave 7)



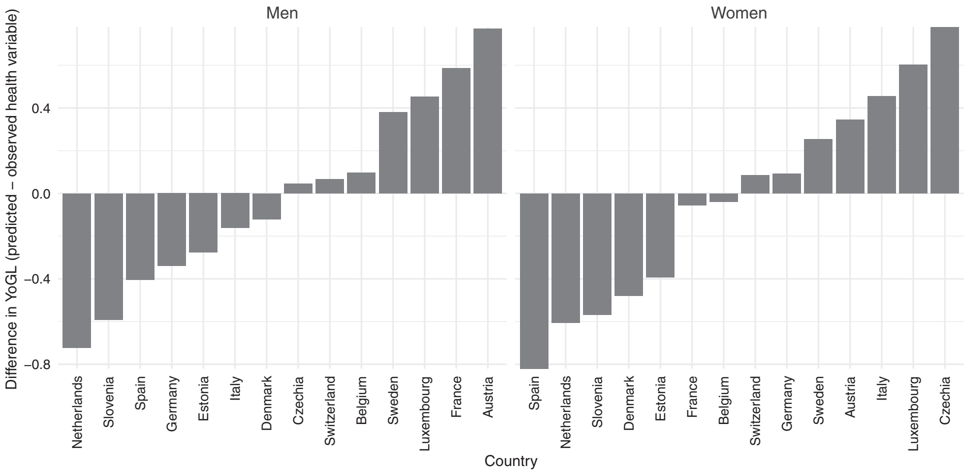
Note: Cross-sectional individual SHARE survey weights are applied.

Figure A.3:
Remaining Years of Good Life at age 50 by country and gender – robustness analysis applying a different threshold for positive life satisfaction (2017, Wave 7)



Note: Cross-sectional individual SHARE survey weights are applied.

Figure A.4:
Remaining Years of Good Life at age 50 by country and gender – robustness analysis comparing results based on observed and predicted physical limitations (2013, Wave 5)



Note: Cross-sectional individual SHARE survey weights are applied.

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 417–466

The life course and subjective well-being across generations – an analysis based on cross-national surveys (2002–2016)

Hideko Matsuo[1,*] ***and Koen Matthijs***[1]

Abstract

This paper identifies subjective well-being trajectories through happiness measures as influenced by time, socio-economic, demographic and behavioural determinants. Hierarchical age-period-cohort models are applied to European Social Survey (2002–2016) data on the population aged 30 and older in 10 countries. A U-shaped relationship between age and happiness is found for some countries, but a rather flat pattern and considerable diversity beyond age 80 are detected for other countries. Lower happiness levels are found for baby boomers (1945–1964) than for pre-boomers and post-boomers, and also for late boomers (1955–1964) than for early boomers (1945–1954). Women, highly educated and native people are shown to have higher happiness levels than men, less educated and non-native people, respectively. Moreover, a positive assessment of income, having a partner, and being a parent, in good health, employed and socially active are all found to have a positive impact on happiness levels. We find evidence of gaps in happiness levels due to differences in socio-economic characteristics over the life course in some, but not in all of the countries analysed.

Keywords: subjective well-being; social inequalities; generations; healthy ageing

1 Introduction

Our life experience and overall well-being are central to our quality of life. Generally speaking, individuals with higher levels of life satisfaction have more positive emotional feelings and a higher quality of life (Skevington and Böhnke 2018). The recent OECD report 'How is life' (2020) – which takes a broader

[1]Center for Sociological Research, Research Unit Family and Population Studies (https://soc.kuleuven.be/ceso/fapos), Leuven, Belgium

[*]Correspondence to: Hideko Matsuo, Hideko.Matsuo@kuleuven.be

DOI: 10.1553/populationyearbook2021.res4.2

approach to assessing well-being, and takes into account multiple items in addition to subjective well-being, including income, health, the environment and social connections – shows that well-being has improved in many countries since 2010, when the financial crisis affected many OECD countries. However, the report also notes that less progress has been made in reducing income inequality, and that large gaps in well-being remain depending on gender, educational level and age. Quality of life is affected by demographic changes (e.g., increasing life expectancy, below-replacement fertility and international migration), growing inequality and social welfare programme constraints – and by the impact of these developments on intrinsic capacity[1] and the environment[2] (WHO 2015). In response to these trends, a number of studies – many of them policy-oriented have been carried out, including research on the development of preventive strategies for the disadvantaged and early assistance for those with urgent needs (Zaidi et al. 2017; UNECE/EC 2019). Healthy ageing[3] has therefore emerged as the leading global and EU policy framework (EC 2014; WHO 2015; OECD 2018).

Well-established approaches to the measurement of subjective well-being (SWB) already exist. These approaches usually involve examining levels of well-being for the overall population or sub-groups based on objective and/or subjective measures (Thompson and Marks 2008; Abdallah et al. 2011). Thus, levels of SWB or happiness emerge as core components of the multi-dimensional well-being concept. Levels of SWB (or happiness) are often described as following a U-shaped pattern over the life course (Helliwell et al. 2019): i.e., SWB starts at a higher level before age 20, declines slowly as people reach their mid-thirties, and then increases again after age 40. In reality, such patterns follow diverse age trajectories in different countries (Steptoe et al. 2015). This implies that more research is needed to generate more robust insights into this pattern. It has, for instance, been suggested that SWB should be measured using large-scale representative samples across countries and over time. Moreover, SWB is a multi-dimensional concept that is often measured through self-reported survey items. Much of the research on SWB has focused on its individual socio-economic and demographic determinants. Such studies have focused on individual countries or involved cross-national and/or cross-sectional analyses of one, several or all aspects of SWB (e.g., Diener 2009; Deeming 2013; OECD 2013; Helliwell et al. 2019).

Although these individual-level determinants of SWB have been studied extensively, relatively little is known about how SWB levels evolve over the life course;

1 This term refers to 'all the mental and physical capacities that a person can draw on and includes their ability to walk, think, see, hear and remember. The level of intrinsic capacity is influenced by a number of factors such as the presence of diseases, injuries and age-related changes' (WHO 2015, 28).

2 Environment is defined as 'the home, community and broader society, and all the factors within them such as the built environment, people and their relationships, attitudes and values, health and social policies, the systems that support them and the services that they implement' (ibid., 29).

3 Healthy ageing can be described as 'the process of developing and maintaining the functional ability that enables well-being in older age' (ibid., 28).

about how SWB trends differ across birth cohorts; or about how SWB patterns differ across countries and regions. In other words, we currently lack a full comparative understanding of heterogeneous life course patterns of SWB with time trends.

Therefore, our aim in this paper is to contribute to the further development of SWB research by focusing on time trends and socio-economic and demographic determinants from a multiple-country, life course and cohort perspective. SWB is assessed through the 'people's level of happiness'. Individual-based, repeated cross-sectional data spanning more than a decade from the European Social Survey 2002–2016, which collects information on levels of happiness, are analysed by applying an extension of hierarchal age period cohort models (i.e., a modification of Bell and Jones 2015) to 10 individual countries. Two research questions are formulated: 1. What are the effects of different dimensions of time (age, period, cohort) on people's levels of happiness? 2. And, do we observe differences in people's happiness levels depending on their socio-economic, demographic and behavioural characteristics, controlled for time effects?

2 General research background

2.1 Subjective well-being

The Stiglitz report (2009), an influential document shaping European policy, defined the key dimensions of well-being as follows: material living standards, health, education, personal activities, social connections and relationships, the environment and economics, and the physical nature of insecurity. Well-being can be understood as people's subjective evaluations of their overall state of physical and mental health (OECD 2018). The concept of SWB constitutes the core of the multi-dimensional concept of well-being (Thompson and Marks 2008; Abdallah et al. 2011), which contains three sub-components: life evaluation, affect and eudemonia (Abdallah et al. 2011; OECD 2013). These sub-components overlap with empirical measures of SWB: hedonic and eudemonic. The former emphasises emotions and the evaluation of SWB through the pursuit of goals leading to happiness and life satisfaction; while the latter focuses on human flourishing, the fulfilment of one's potential in the form of individual autonomy and self-actualisation (Vanhoutte 2014).

The well-being of the ageing European population has been linked to 'three main components: low probability of disease and disease-related disability, high cognitive and physical functional capacity, and active engagement with life' (Rowe and Kahn 1997, 433). It has been shown that individuals with higher levels of well-being tend to be more economically productive, and to have higher incomes, better health and higher life expectancy (Huppert 2009). The relationship between physical health and SWB is bidirectional: older people who are in poor health tend to experience lower levels of SWB, whereas having higher levels of SWB can contribute to better health (Steptoe et al. 2015).

2.2 Time trends and the individual- and country-level characteristics of subjective well-being

2.2.1 Time trends

Taking time trajectories into account is fundamental for understanding levels of SWB. There is extensive evidence indicating that age affects SWB. In an analysis of British panel data from 14 data collection waves starting in 1991 on respondents aged 16–64 years, Clark (2007) found a U-shaped relationship between age and SWB (i.e., overall mental well-being, overall life satisfaction). A U-shaped age-SWB relationship has also been confirmed for other countries: Blanchflower and Oswald (2008) found evidence of the existence of such a pattern for the US and Western Europe between 1976 and 2006 based on data from the US General Social Survey (1972–2006), the Eurobarometer survey (1976–2002) and the World Values Survey (1981–2004) covering respondents aged 20 and older in 80 countries. A U-shaped pattern was also observed for Eastern European, Latin American and Asian nations, but not for some developing countries. More recently, Blanchflower (2021) re-examined this U-shaped happiness-age curve globally based on recent data for 145 nations, including 109 developing countries. After controlling for socio-demographic variables and analysing samples covering respondents under age 70, a U-shaped pattern was confirmed globally.

The existence of such a relationship may seem counterintuitive, as the elderly are more likely than younger people to experience negative life events, such as reduced income, deteriorating health or the death of a partner. This pattern has been explained through socioemotional selectivity theory (Carstensen et al. 2003), which argues that compared to younger people, the elderly pay more attention to emotional goals and experiences, and are more likely to remember emotional and positive information.

It appears, however, that this U-shaped curve is not universal. Frijters and Beatton (2012) rejected the claim that there is a U-shaped relationship between age and life satisfaction (i.e., individual happiness) based on an analysis of panel survey data for respondents aged 18 and older in Germany, the United Kingdom and Australia. When using conventional regressions of fixed-effects models, the U-shaped relation disappears due to reverse causality. They argued that ‘happiness-increasing variables, like getting a job, a high income, and getting married, appear to happen mostly to middle-aged individuals who were already happy’ (ibid., 540). There is other evidence that the U-shaped curve does not exist across all countries. Based on the results of a Gallup poll among people aged 15 in 160 countries, Steptoe et al. (2015) examined the relationship between age and three aspects of SWB (i.e., evaluative, hedonic and eudemonic). They found a U-shaped relationship in high-income and English-speaking countries, but a negative association in other regions, including former Soviet states and Eastern Europe.

Levels of SWB also appear to be determined in part by people's birth cohort. Empirical evidence from the US generally shows that levels of SWB are lower among baby boomers than among members of earlier birth cohorts, keeping other factors constant; even though selection effects exist (Yang 2008). Based on data from the English Longitudinal Study of Ageing (ELSA), Jivraj et al. (2014) found that, after controlling for age, older cohorts (aged 70+ in 2002–2011; i.e., pre-baby boomers) had the same or higher levels of SWB (i.e., quality of life, depressive symptomatology, life satisfaction) than younger cohorts (aged 50–59 in 2002–2011; i.e., late baby boomers).

There are many reasons why baby boomers might be distinct from members of other cohorts. The share of elderly in the population is currently growing primarily because the baby boom cohorts (i.e., cohorts born in 1945–1964) are so large. Baby boomers have a socio-economic and cultural identity related to their educational and occupational careers that is distinct from the identities of the previous and the following generations. Compared to the preceding generation, baby boomers have tended to report lower levels of well-being, which can likely be attributed to their early and formative life conditions and experiences, including being faced with rising economic inequality in the post-war era resulting from a larger cohort size that created greater competition for education and jobs (Easterlin et al. 1993; Fukuda 2013). Moreover, at least in Western countries, baby boomers can be seen as the forerunners in the dramatic changes in family life that characterised the Second Demographic Transition (SDT) regime (Lesthaeghe 2014). Based on data from three US censuses conducted between 1980 and 2000, Lin and Brown (2012) found that one out of every three baby boomers was unmarried, largely due to divorce or having never married, with a very small proportion having been widowed. Baby boomers continue to experience unstable marriages as rates of divorce increase in later life, as part of the so-called 'grey divorce' trend (Brown and Wright 2017; Stepler 2017).

While baby boomers as a group differ from other cohorts in terms of their socio-economic, cultural and demographic characteristics, considerable heterogeneity can also be found within the group of the baby boomers (i.e., early boomers were born in 1945–1954 and late boomers were born in 1955–1964). Based on the 'Midlife in the U.S.' surveys of 1995–1996 and 2004–2006, Barrett and Toothman (2014) examined the subjective experiences of ageing and their health consequences (i.e., individual views of the life course, such as the structure, timing and advancement of careers) during the midlife period. The early boomers reported having significantly more youthful identities and later boundaries for middle age than the late boomers. The late baby boomers also had lower socio-economic positions than the early baby boomers, keeping age constant. These findings imply that the early and the late boomers have experienced different economic and social conditions that have produced different life course trajectories (ibid.).

While age and cohort effects are rather well-established, evidence on the impact of historical events or trends on levels of SWB has been rather mixed. While economic prosperity has increased over time, happiness has not risen in a linear manner (Easterlin 2010). This non-linear relationship between income and happiness over time has also been documented by Yang (2008). An analysis of General Social Survey (GSS) data showed that happiness levels decreased in the 1970s–1980s, and gradually rebounded thereafter. However, Ortiz-Ospina and Roser (2017) provided another perspective. Based on an analysis of data from the Gallup World Poll and the World Values Survey, they found evidence that happiness has increased globally over the past three decades (1984–2014). In the same study, and based on data from the Eurobarometer survey on life satisfaction, the authors also reported observing fluctuations across historical periods (1973–2016). However, they generally found decade-long positive trends for most European countries except Greece, where there was a substantial decrease in life satisfaction during the financial crisis (2007–2012).

2.2.2 Individual socio-economic, demographic and behavioural differences in SWB levels, and time variations

Levels of SWB are highly dependent on people's socio-economic and demographic characteristics. Multi-dimensional concepts and theoretical frameworks can provide valuable insights into this relationship by highlighting a number of determinants of SWB, such as income, health, social contacts, employment status, personality and culture (religion) (OECD 2013, 33, Figure 1.1); as well as marital status, education and living arrangements (Thompson and Marks 2008). These social differentials of SWB levels are also strongly associated with the individual's place of origin (OECD 2017).

Clear evidence exists on the effects of socio-economic position on levels of SWB. According to a systematic review of literature from the 1995–2013 period, among the population aged 60+, having a poor socio-economic position (as reflected in, for example, levels of education, income, wealth, financial assets, social class and neighbourhood characteristics) is associated strongly with poor SWB (i.e., subjective health, life satisfaction, quality of life, well-being), whereas this association weakens when mediated by social support and a positive self-perception of health (Read et al. 2016). A significant positive relationship has been found between SWB (i.e., happiness, life satisfaction, morale) and levels of education in particular (Witter et al. 1984). However, while there is strong evidence that higher education leads to greater happiness, either directly or indirectly (e.g., via income), this relationship is complex due to the difficulties that can arise in the operationalisation of education (Michalos 2008).

Empirical research has identified gaps in levels of SWB between native and non-native people, even after controlling for country and individual characteristics (Hadjar and Backes 2013). Migrant populations account for around 10% of Europe's total population, although the proportions and socio-economic backgrounds of these migrant groups differ substantially across countries. Based on an analysis of data from the Survey of Health, Aging and Retirement in Europe (SHARE), Sand and

Gruber (2018) presented empirical evidence indicating that non-European migrants and migrants from Southern and Eastern Europe have substantially lower levels of SWB than natives. They also found, however, that migrants from Northern and Central European countries have levels of SWB comparable to those of non-migrants. In general, migrants have been shown to have lower socio-economic status and SWB levels than the native population. However more empirical evidence of this general finding is required because of the significant methodological issues in this research related to non-response and measurement errors (OECD 2017).

An earlier study has provided evidence of gaps in levels of SWB by gender, albeit with large differences by country (Inglehart 2002). The analysis of data from the World Values Survey in 65 countries between 1980 and 1999 using a life satisfaction measure found that women had higher levels of SWB than men in the more developed countries. However, it appears that gender can have mixed effects. On the one hand, based on a meta-analysis of developmental and gerontological literature obtained through electronic databases (i.e., PSYCINFO, MEDLINE, PSYINDEX), Pinquart and Sörensen (2000) found that associations between socio-economic status (e.g., education, income and social class) and SWB levels (i.e., life satisfaction, happiness, self-esteem) were generally stronger for men than for women. The authors attributed this finding to men having a traditional occupational path focused on personal career development. Another view has been provided by Witter and others (1984). Based on a meta-analysis of US literature prior to 1980, they found a significant positive relationship between education and SWB levels that was stronger for women than for men. The authors noted, however, that the strength of the relationship was reduced when occupational status was controlled for in the model.

What is less clear is how these differences in levels of SWB by socio-economic characteristics described above vary over the life course. On the one hand, cumulative advantage and disadvantage theory argues that socio-economic gradients in SWB levels can vary over the life course (Dannefer 2003). This implies that relatively small differences in education or income during adulthood can widen during the life course. It has, for example, been shown that there is a strong negative relationship between financial distress and well-being (i.e., control autonomy, self-realisation, pleasure) that widens in old age (Niedzwiedz et al. 2015). Interestingly, this study found that this association is weaker for Northern countries.

On the other hand, a number of studies have shown that SWB levels tend to converge over the life course. For example, a study for the US found that having advantageous socio-economic characteristics (e.g., being male, white and highly educated) does not continue to confer advantages at older ages (Yang 2008). Sand and Gruber (2018) also found evidence that the gap in levels of SWB between natives and migrants diminishes at older ages. Another study by Inglehart (2002) found that in rich Western societies – but not in many non-Western (Asian, Latin American, African) societies – the relationship between gender and well-being reverses, from women having an advantage at younger ages to women having a disadvantage at older ages. Gender differences in SWB (i.e., life satisfaction, happiness) have been attributed to women generally reporting higher frequencies of depression because of their disadvantaged socio-economic position (Van de Velde et al. 2010), and experiencing and expressing emotion more than men (Plant et al. 2004).

Research has shown that demographic factors play an important role in SWB levels (and the reverse). Lower levels of SWB, measured in a multi-dimensional manner (i.e., life satisfaction, emotional well-being, vitality), were found for divorced than for married people in Northern Belgium (Jenkinson et al. 2020). The relationship between fertility and SWB has been studied extensively, mostly in North America and Europe. An analysis of data from the British Household Panel Survey found that having children has large positive effects on life satisfaction (Angeles 2010), and that these effects become stronger as the number of children increases. Recent cross-national research has shown that high levels of life satisfaction instigate fertility, even in low-fertility countries (Mencarini et al. 2018), although this relationship can vary depending on individual factors and across institutional contexts. Based on data from a large number of OECD countries, Glass and others (2016) found large variations in the happiness of parents, depending on the generosity of family policies. There is evidence that the association between parenthood and SWB can be weakened or strengthened depending on cultural practices and institutional arrangements, in particular when gender inequality is present (Aassve et al. 2015). Thus, the generally positive relationship between the number of children and happiness can be influenced by contextual factors that lead to happiness levels being higher at younger ages (i.e., around the birth of first child), but decreasing thereafter (Myrskylä and Margolis 2014). Moreover, it has been shown that life-cycle conditions can affect the relationship between happiness and fertility due to the financial and emotional costs of childrearing, but that the high costs for parents when children are young can be compensated when the children are older and provide their parents with social protection (Margolis and Myrskylä 2011).

2.2.3 Country institutional characteristics and individual subjective well-being

In an analysis of the World Database of Happiness that controlled for individual factors, including income, unemployment, inequality, social capital and life satisfaction, Spruk and Kešeljević (2016) found that countries with better economic institutions and higher levels of economic freedom (e.g., property rights, open markets, limited governance) have higher individual happiness levels. Based on data from the World Values Surveys in OECD countries, Welsch and Kühling (2016) found evidence of a positive relationship between macro-economic performance and life satisfaction levels, with trade openness and institutional quality playing a crucial role. Rather than mediating effects, Pacek and Radcliff (2008) found that welfare programmes have direct effects on individual happiness and life satisfaction.

These findings imply that a country's social welfare characteristics also have an effect on individual well-being levels.

2.3 Study framework, research questions and general hypotheses

2.3.1 Study framework

In order to examine SWB empirically, this paper focuses from a life course perspective (Giele and Elder 1998; Shanahan et al. 2016) on one aspect of SWB, happiness. Specifically, we examine the time, socio-economic, demographic and behavioural determinants of happiness. The framework comprises four life course components 'individual agency', 'linked lives', 'historical period and time' and 'timing' – that correspond broadly to age, period and cohort. Our study focuses in particular on age and cohort. Cohort is an important concept for understanding social change because each cohort has distinct experiences and characteristics, driven by the changing contexts of formal education, socialisation and historical events (Ryder 1965). Compared to other generations, the baby boomers experienced higher levels of competition in schools and jobs, which resulted in socio-economic and demographic differentials in education, income, family and health status.

Over the life course, gaps in levels of SWB by socio-economic characteristics can either increase or decrease. On the one hand, cumulative advantage and disadvantage theory argues that relatively small differences that emerge during adulthood can widen during the remainder of the life course. These relative advantages and/or disadvantages can result in widening differences in later life (Dannefer 2003). For instance, even though in the post-war period overall educational levels increased and the welfare state expanded, having low educational attainment early in life continued to affect people's subsequent occupational trajectories, resulting in economic hardship even after retirement (Oris et al. 2017) that could influence their SWB levels. On the other hand, there may be age effects of happiness, which suggests that socially advantaged groups are not necessarily better off than the advantaged groups later in life. This implies that crossovers or smaller gaps in SWB levels by socio-economic gradients can occur during the life course, resulting in converging SWB levels.

2.3.2 Research questions and hypotheses

Two research questions, and corresponding hypotheses, are formulated as follows.

1. What are the effects of different dimensions of time (age, period, cohort) on people's levels of happiness? Following previous empirical research on the impact of age, we expect (H1a) to observe a general U-shaped pattern. (H1b) We also assume that the cohort effects (i.e., effects resulting from changes that occur and influence people in their earlier years) are more substantial than the period effects (i.e., effects resulting from changes that influence all age groups equally).

This means that we anticipate finding some negative period effects for the years of the economic-financial crisis (2008 and later), and that we expect to observe more substantial negative effects for baby boomers, and particularly for late baby boomers (born in 1955–1964) relative to pre-baby boomers (born before 1945). (H1c) These time-specific trends are, however, assumed to differ cross-nationally: we expect to find U-shaped age effects in Northern and Western European countries, but smaller effects in Eastern European countries; and lower levels of happiness in Eastern European countries than in Northern and Western European countries.

2. Do we observe differences in happiness levels depending on people's socio-economic, demographic and behavioural characteristics, controlled for time effects? And, if so, do we observe widening or decreasing differences in people's happiness levels over time? (H2a) We assume that people with weaker socio-economic profiles have lower levels of happiness in general, but that these patterns are diverse during the life course, and to some extent across cohorts: i.e., we expect to find that some socio-economic profiles (e.g., low education) exhibit growing disparities; and that others (e.g., female, non-native) show crossover trends whereby socio-economic advantages turn into disadvantages, and vice versa. (H2b) We also expect to observe lower happiness levels among people with lower socio-economic and non-traditional family status; albeit with substantial cross-national differences, as discussed above. We furthermore expect socio-economic gaps due to gender and educational level to decrease moderately in Northern and Western European countries because of the effects of general social welfare programmes, but to remain large in Eastern and Southern European countries, where social welfare programmes are rather limited. While non-native status is largely related to the socio-economic characteristics and country origin of the migrants in each country, decreasing native-non-native gaps are generally anticipated.

3 Data, measures, methods

3.1 Data

The data files of ESS Rounds 1 to 8 (2002–2016),[4] and in particular the subsample data file on non-institutionalised individuals (i.e., those not residing in institutions) aged 30–100 in 10 countries (DE, ES, GB, HU, IE, NL, NO, PT, SE, SI), were analysed. ESS is a repeated cross-sectional and multi-country survey, with a comparable European sample. It aims to ensure high methodological quality with respect to survey design and measurement for cross-national analysis. The cumulative datasets include 15 countries for all rounds, which means that five

[4] This analysis makes use of the most recent publicly released cumulative ESS data file on version 6.6 (edition of 12 December 2018, European Social Survey 2018). This version consists of 15 countries (BE, CH, DE, ES, FI, GB, FR, HU, IE, NL, NO, PL, PT, SE, SI).

countries were removed from the analysis for the purposes of this paper.[5] The total number of sample units analysed for 10 countries was 120,810 units, as only sample units with complete cases were included in our analysis. Missing data accounted for less than 3% of the sample in each country.

3.2 Measures

Measures were selected and conceptualised based on Yang (2008), Bell (2014) and Bell and Jones (2015), taking into consideration other empirical research, as discussed in Section 2. All measures selected were collected in each round of ESS.

3.2.1 Individual-level dependent variable

We operationalised the hedonic aspect of SWB through measures of happiness, and created an ordinal response variable (1 = very happy, 2 = happy, 3 = unhappy) based on an originally 11-point scale variable (i.e., how happy are you?; 0: extremely unhappy to 10: extremely happy). The ordinal response variable ranges from 1 = very happy (original points scale of 8, 9, 10), over 2 = happy (original points scale of 5, 6, 7) to 3 = not too happy (original points scale of less than 4). We chose to use an ordinal response variable because we observed that there were few responses for some of the response choices for the continuous variable. This was considered a major disadvantage in the application of cross-national analysis.

3.2.2 Individual-level independent variables

Key socio-economic individual items – age, birth year, gender, level of educational attainment and native/non-native status – were included in the fixed part of the initial model. These variables were mostly transformed into dummies, except for age and birth year, which were grand mean centred and squared. In addition, the interaction effects between these socio-economic variables and time (i.e., age and birth year) were included in the extended models (models 2, 3). The interaction effects between age and birth year that test for life course trends were incorporated in order to examine varying cohort effects (Bell 2014; Bell and Jones 2015). This item made it possible to assess whether and how life course trajectories change across cohorts. In addition to the above, the respondent's assessment of income, employment status,

5 Five countries are removed from the analysis due to the use of a non-standardised questionnaire on partnership status across rounds in Finland and France; and because no values for covariate parameters were obtained in the random models in Belgium, Poland and Switzerland due to insufficient sample sizes in the sub-groups. However, all of the geographical regions of Europe are represented by at least one of the 10 countries remaining in the analysis.

partnership status, child status, health condition and social interaction level were all included in the full model.

- Age at the time of the survey and polynomial age
- Birth year and polynomial birth year
- Gender (1 = female; 0 = male)
- Years of education: low [1 = below 12 years; 0 = otherwise); high [1 = more than 16 years; 0 = otherwise]
- Non-native status (1 = self and/or parents not born in country; 0 = otherwise)[6]
- Feeling about current household income (1 = difficult to manage; 0 = otherwise)
- Current partnership status (1 = divorced or separated; 0 = otherwise); (1 = widowed; 0 = otherwise) (1 = never married; 0 = otherwise)
- Current child status [1 = have child/ren (i.e., either non-residential and/or residential child; 0 = otherwise)]
- Current employment status (1 = part-time work; 0 = otherwise), (1 = unemployed; 0 = otherwise), (1 = retired; 0 = otherwise)
- Self-perception of health condition (1 = excellent; 0 = otherwise), (1 = fair; 0 = otherwise), (1 = poor; 0 = otherwise)
- Social interaction [take part in social activities compared to others of same age: 1 (much less than most) to 5 (much more than most)]
- Interaction of, respectively, gender, education and native/non-native status with age
- Interaction of, respectively, gender, education and native/non-native status with birth year
- Age interacting with birth year

3.2.3 Random level variables

Period (survey rounds) and cohort groups (10-year groups) were included in the random part of the hierarchical model. In each survey round, data collection took place in the same year in all countries (European Social Survey 2016).

[6] The country of birth of the respondent and of both parents are taken into account in determining the respondent's non-native/native status. This means, for instance, that second-generation (children of first-generation migrants who were born in another country) respondents are considered non-native in our analysis.

3.3 Methods

3.3.1 Descriptive analysis

Profiles of high happiness levels are shown first, distinguished by periods and birth cohorts in each country. Each survey round (eight rounds corresponding to 2002–2016) and each cohort group (comprising 10 years, corresponding to birth years between 1905 and 1985) are shown.

Analytical models

Hierarchical age-period-cohort (HAPC) models applying Bell and Jones's modifications are used (Snijders and Bosker 1999; Yang and Land 2013; Bell 2014; Bell and Jones 2013, 2015) to examine the dynamic changes in SWB levels through happiness measures by time trends. Cohort effects, which result from events and experiences that occur and influence people in the earlier years of their life course, are considered more substantial than period effects, which result from general historical events that influence all age groups equally. This assumption is justified because it enables us to take into account the upbringing and experiences of particular cohorts by examining pre- and post-baby boomers' levels of happiness.

In order to address the linearly dependent issue when modelling age, period and cohort, our model constrains the period linear trends to zero by including the birth year as a linear fixed effect (Bell 2014; Bell and Jones 2015). Basic level-1 consists of (Yang and Land 2013; Bell and Jones 2015, 203):

$$\Pr(Happy\ Y_{i(j_1 j_2)} = 1) = \beta_{0 j_1 j_2} + \beta_1 Age_{i(j_1 j_2)} + \beta_2 Age^2_{i(j_1 j_2)} + \mathrm{e}_{i(j_1 j_2)} \tag{1}$$

The random intercept model is:

$$\beta_{0 j_1 j_2} = \beta_0 + \beta_3 Cohort_{j_1} + u_{j_1} + u_{j_2} \tag{2}$$

The dependent variable is the ordinal response outcome of happiness of the ith respondent in the j_1th period (survey rounds) (j_1 = 1–8) and the j_2th cohort (j_2 = 1–9) (ibid. 203). The addition of a single birth year (cohort) in the fixed level is a modification by Bell (2014) and Bell and Jones (2015) of the model proposed by Yang and Land (2013). Period and cohort residuals remain in the model to illustrate contextual effects. Based on the combined 2-level models, the likelihood of 'being very happy' is predicted. The model includes gender (female/not), migrant status (native origin/not) and levels (low/high) of educational attainment; the interaction effects between those variables and age/birth year; and socio-economic, demographic and behavioural items (e.g., partnership status, child status, employment status, health status, income assessment, social interaction levels).

Happiness levels are estimated with three models for each country using the SAS programme and implementing proc glimmix. All models consist of individual fixed and random parts. The basic model consists of time and individual fixed core socio-economic determinants (model 1), to which interaction effects are added (model 2), followed by additional socio-economic, demographic and behavioural determinants (model 3). Time trends are examined by studying the statistical significance of individual estimated coefficients for birth cohorts and periods that identify patterns in graphs, as proposed by Frenk and others (2013). Models are estimated with the hierarchical ordinal logit models of happiness using residual pseudo-likelihood estimation. Fit statistics (i.e., −2 Res Long Pseudo-likelihood) are documented and examined across models. All model results incorporate design weights, which are publicly available in datasets.

4 Results

The overall picture of the levels of happiness is presented first, in the form of descriptive trends. The core socio-economic determinants controlled by time are then shown, followed by varying happiness levels by core socio-economic characteristics over the life course and across birth cohorts, as well as additional covariates. All results are distinguished by countries, and are presented in the respective tables and figures.

4.1 Descriptive statistics

Table 1 presents the characteristics of the sample average for each country individually and for all countries together. Across all countries, the mean age is 54 years. Slightly more women (53%) than men are included in the sample. Most individuals are native (10% are non-native). The respondents' educational backgrounds are diverse: on average, the share of those with fewer years of schooling (<12 years) is 46%, and the share of those with more years of schooling (+16 years) is about 23%. On average, 10% of respondents are divorced and widowed, and 17% are never married. A majority of individuals (roughly 81%) have children. More than one-fifth of respondents assess their household income negatively. More than one-quarter of respondents are retired (28%), whereas around 10% work part-time and 5% are unemployed. On average, 28% of respondents rate their health status as fair, while 10% assess it as poor. These general sample characteristics of the 10 countries combined are, however, subject to substantial cross-national differences, as shown by several measures for each country in Table 1. There are particularly large differences with respect to educational levels, assessments of income and self-reported health levels.

Classified by period and cohort, the proportion of people with high levels of happiness differs across countries (Figure 1). For Western and Northern European countries, relatively similar high shares are observed (50–80% of each birth cohort), whereas the shares are generally lower for Southern (PT) and Eastern European

Table 1:
Sample socio-economic, demographic and behavioural characteristics (% or mean) in ten countries

	DE	ES	GB	HU	IE	NL	NO	PT	SE	SI	Average
High happiness, %	56.34	55.8	59.67	31.9	57.68	67.06	71.19	32.7	68.65	48.5	54.95
Age, mean (SD)	54.31	52.99	55.24	54.39	53.34	54.32	52.38	56.94	54.63	54.34	54.29
	(14.27)	(15.39)	(15.73)	(14.95)	(15.06)	(14.98)	(14.55)	(15.75)	(15.21)	(14.64)	
Female, %	49.85	51.14	55.06	56.44	53.94	54.96	47.23	60.51	50.28	55.48	53.49
Non-native, %	11.33	8.78	13.68	3.27	13.16	10.35	9.81	6.33	14.07	13.34	10.41
Low educated, %	25.92	54.72	44.03	48.26	32	37.39	30.84	76.14	38.23	49.49	45.68
High educated, %	26.92	22.81	26.34	17.56	27.16	27.31	33.56	10.01	23.91	15.58	23.12
Divorced, %	10.81	6.34	14.94	13.65	6.68	11.97	12.78	8.11	13.74	6.16	10.52
Widowed, %	8.5	8.81	13.3	16.27	10.23	11.29	6.33	15.33	6.68	11.63	10.84
Never-married, %	14.15	16.54	16.72	11.06	21.74	16.93	19.68	11.48	23.42	15.14	16.69
Parent, %	77.77	78.2	78.06	85.21	74.56	74.78	85.51	79.13	85.27	87.38	80.59
Neg. income assessment, %	14.19	23.75	15.94	46.12	21.58	12.24	7.07	44.27	8	17.01	21.02
Excellent health, %	11.78	14.31	27.53	7.53	35.88	14.51	30.4	5.97	29.02	11.59	18.85
Fair health, %	32.82	29.16	22.91	37.12	17.21	25.56	19.41	40.21	21.1	37.56	28.31
Poor health, %	11.16	12.49	9.01	18.67	3.48	5.45	6.4	16.7	4.91	12.49	10.08
Unemployed, %	5.42	8.14	3.49	4.75	7.52	3.38	2.09	7.05	2.98	5.12	4.99
Part-time, %	10.93	10.11	13.07	5.97	12.28	16.47	10.64	8.12	5.84	4.32	9.78
Retired, %	29.34	19.25	32.68	35.93	21.93	23.73	20.09	35.9	25.94	35.01	27.98
Social interaction, mean (SD)	2.71	2.61	2.72	2.38	2.68	2.81	2.91	2.53	2.88	2.58	2.68
	(0.92)	(0.92)	(0.98)	(0.94)	(0.89)	(0.96)	(0.78)	(0.87)	(0.90)	(0.96)	
N	18,292	11,400	14,355	9,900	13,985	12,561	10,359	11,114	11,157	7,687	120,810

(HU, SI) countries (below 60% for each birth cohort). For the latter group of countries, the more recent the cohort, the higher the proportion, which leads to a positive linear trend. High percentages of 'very happy' people are also found to some extent across different periods for several countries (DE, IE, PT). The low share for Portugal in the 2010–2012 period stands out.

4.2 Time trends

Starting with the analysis of age effects, a key question is whether life really begins after people have made it through their forties, as formative years of struggle

Figure 1:
Proportion high happiness (very happy) by birth cohort (1905–'85) and survey rounds (2002–16) in ten countries

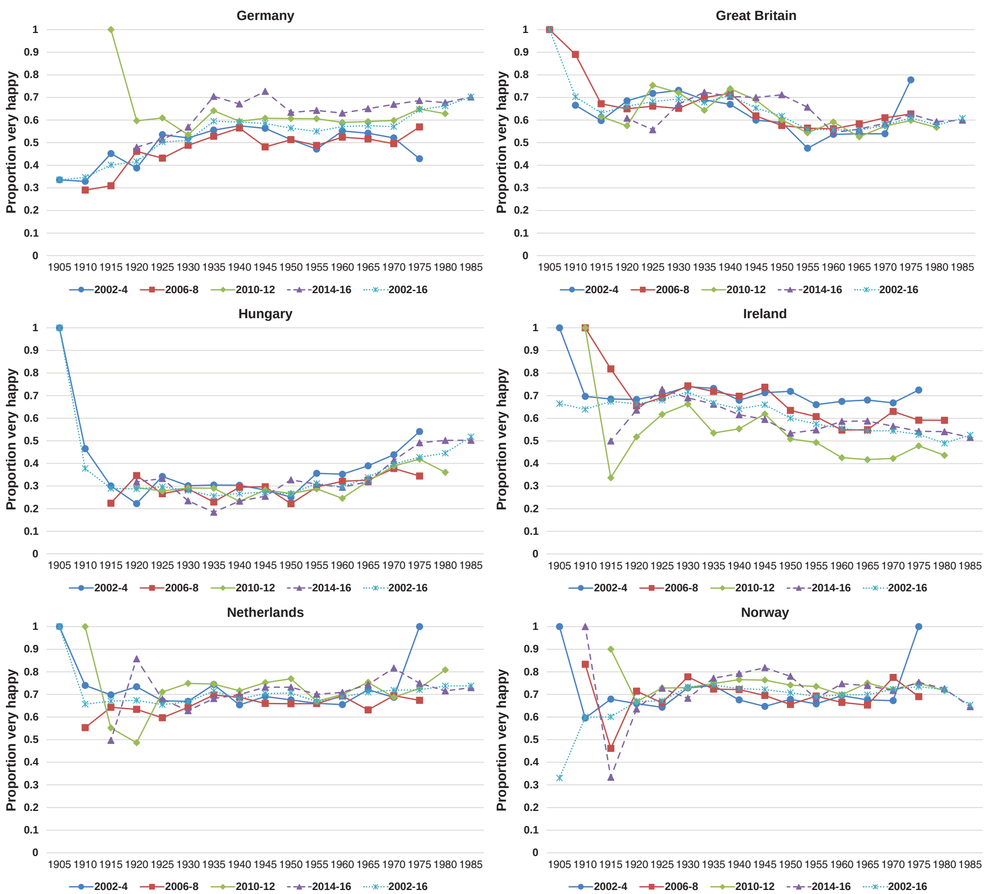
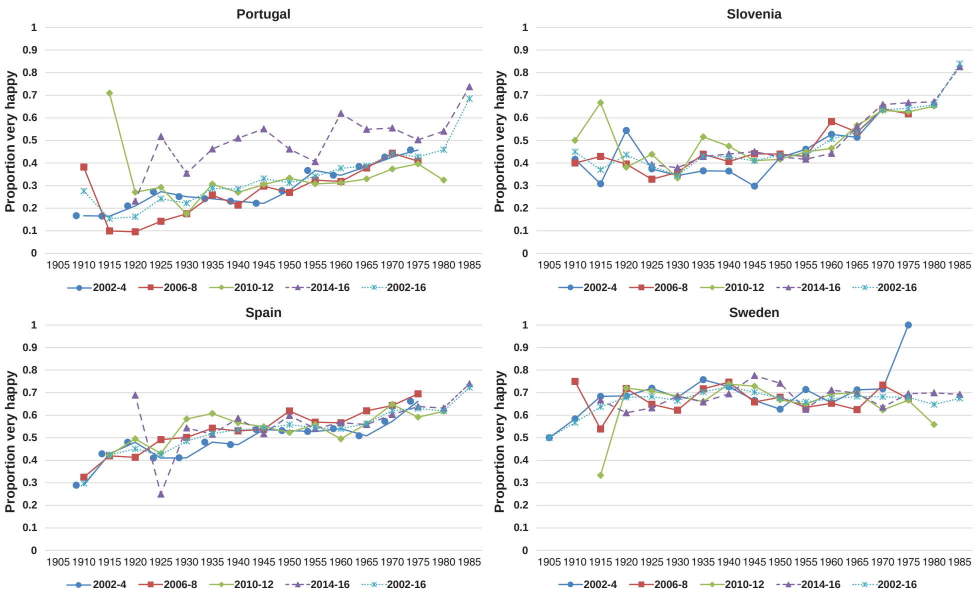


Figure 1:
Continued



Note: This category of 'very happy' corresponds to the dichotomization of response items in the model.
Source: European Social Survey cumulative data file in 10 countries.

and challenges are compensated for with lower stress and higher income levels, especially as children start leaving the parental home. Thus, a period of happiness could begin from the mid-fifties onwards. Figure 2 shows the overall age trends for the predicted probabilities of being very happy estimated from model (1) presented in Table 2.[7] The fixed effects of age and polynomial age in Table 2 do not show significant effects in all countries, although Figure 2 clearly displays age trends by country. This figure shows that happiness evolves, but differently across countries. A U-shaped relationship between age and happiness (i.e., reflecting lower levels of happiness in mid-life, which increase thereafter) is found in several countries (GB, HU, SI) (H1a), while other countries (DE, ES, IE, NL) show rather flat age patterns until ages 60–70 (i.e., younger seniors). A general decline with age, and negative age effects for older seniors (i.e., ages 80+), are observed for Southern European countries (PT). Contrary to expectations (H1c), a non-U-shaped pattern is not observed among Eastern European countries (HU, SI), whereas two Southern

[7] The predicted logits (denoted β) are converted to predicted probabilities by calculating probability = $1/(1 + \exp\beta)$.

European countries (ES and PT) exhibit this pattern. These different age trends across countries may be due to different reasons, one of which may relate to the sample selection, as the proportion of very old people varies substantially across countries.

Figure 2:
Overall age trends for predicted probabilities of being very happy based on model 1 in ten countries

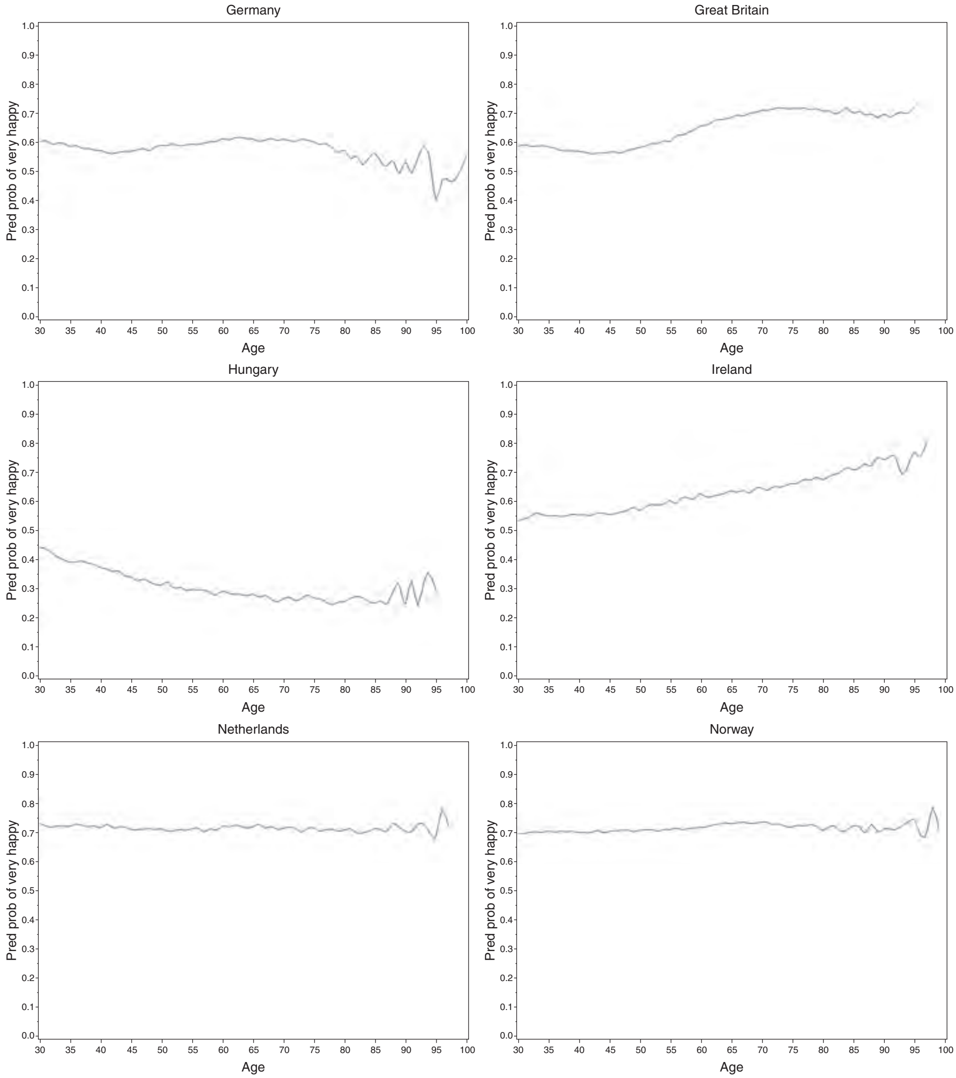
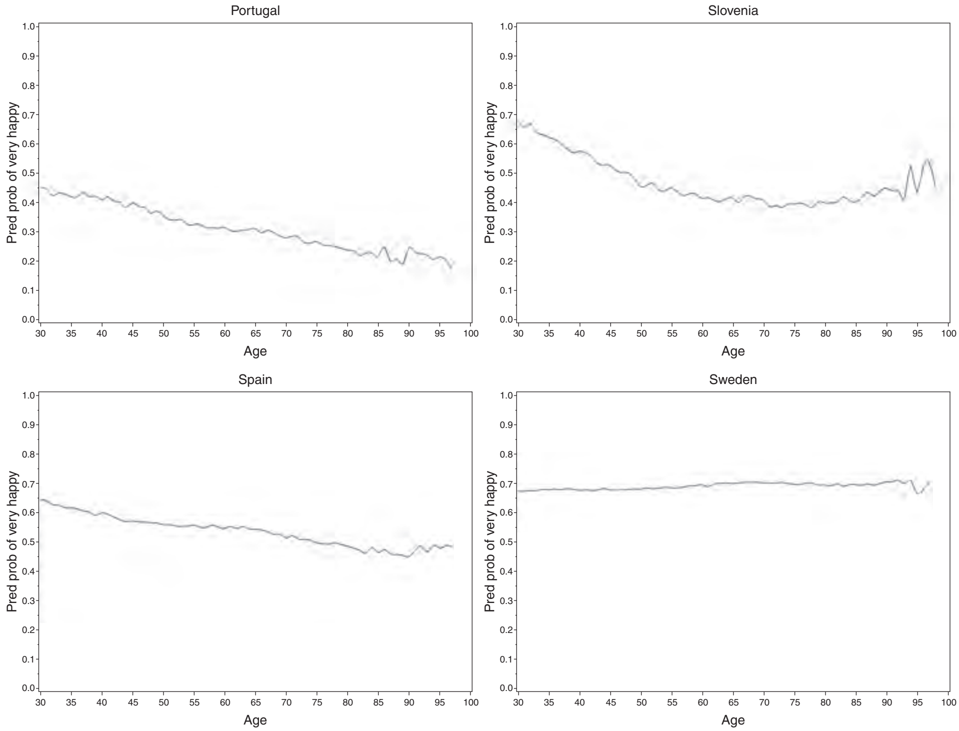


Figure 2:
Continued



Both period and cohort effects on happiness levels are shown in Figure 3[8] based on model 1. We examine period and cohort residuals through the coefficients for each period and cohort level in order to examine individual estimated period and cohort effects (Frenk et al. 2013). The corresponding figures in each country graphically plot the estimated cohorts and time period effects within the 95% confidence bounds. The coefficients of the period and cohort residuals are, in general, not large, although period and cohort variations are observed. In line with our hypothesis (H1b) and previous empirical research on cohort characteristics (Barrett and Toothman 2014), we find relatively low levels of happiness among baby boomers. In contrast to pre-baby boomers (i.e., those born before 1945), baby boomers in most countries have lower levels of happiness compared to the average values across all cohorts in each country. Moreover, in most countries, we find that late baby boomers (i.e., those born in 1955–1964) have lower levels of happiness

[8] Following Frenk and others (2013), time trends are examined by studying the statistical significance of individual estimated coefficients for birth cohorts and periods, and identifying patterns in graphs.

Table 2:
Effects on high levels of happiness: Age-period-cohort models 1–3 in ten countries

	DE		
	Model 1	Model 2	Model 3
Intercept	0.295 (0.099)*	0.301 (0.097)*	−0.015 (0.106)
Intercept	2.614 (0.102)***	2.625 (0.100)***	2.691 (0.110)***
Age	0.439 (0.106)***	0.524 (0.117)***	0.608 (0.124)***
Age^2	0.459 (0.256)^	0.464 (0.256)^	0.424 (0.270)
Birth year	0.041 (0.011)**	0.042 (0.012)**	0.049 (0.012)***
Birth $year^2$	0.004 (0.003)^	0.005 (0.003)^	0.005 (0.003)^
Gender	0.085 (0.029)**	0.073 (0.029)*	0.141 (0.032)***
Non-native	−0.186 (0.42)***	−0.159 (0.044)**	−0.040 (0.047)
Low educated	−0.347 (0.037)***	−0.333 (0.039)***	−0.111 (0.041)**
High educated	0.280 (0.036)***	0.235 (0.038)***	0.094 (0.040)*
Age ∗ birth year	0.092 (0.051)^	0.096 (0.051)^	0.087 (0.054)^
Age ∗ gender		−0.118 (0.067)^	0.019 (0.071)
Age ∗ non-native		0.054 (0.098)	0.090 (0.104)
Age ∗ low educated		0.142 (0.086)^	0.066 (0.091)
Age ∗ high educated		−0.262 (0.082)**	−0.298 (0.086)**
Birth year ∗ gender		−0.00 (0.007)	0.002 (0.007)
Birth year ∗ non-native		−0.0005 (0.009)	0.000 (0.010)
Birth year ∗ low educated		0.015 (0.009)^	0.010 (0.009)
Birth year ∗ high educated		−0.015 (0.008)^	−0.019 (0.008)*
Income assessment			−1.083 (0.047)***
Divorced			−0.750 (0.049)***
Widowed			−1.021 (0.062)***
Never married			−0.789 (0.051)***
Have child			0.060 (0.040)
Excellent health status			0.615 (0.054)***
Fair health status			−0.598 (0.035)***
Poor health status			−1.271 (0.051)***
Part-time work			0.015 (0.049)
Unemployed			−0.548 (0.075)***
Retired			−0.050 (0.053)
Social interaction			0.304 (0.017)***
Random-period	0.017 (0.012)^	0.017 (0.012)^	0.019 (0.013)^
Random-cohort	0.018 (0.013)^	0.016 (0.011)^	0.004 (0.004)
−2 Res log pseudo-likelihood	155,710.671	155,816.065	164,609.403

Continued

Table 2:
Continued

	ES		
	Model 1	Model 2	Model 3
Intercept	0.415 (0.109)**	0.432 (0.110)**	−0.155 (0.134)
Intercept	3.188 (0.116)***	3.211 (0.117)***	2.840 (0.139)***
Age	0.063 (0.116)	0.179 (0.146)	0.289 (0.147)*
Age^2	−0.074 (0.269)	−0.088 (0.268)	−0.133 (0.263)
Birth year	0.019 (0.012)^	0.021 (0.015)	0.035 (0.014)*
Birth $year^2$	−0.001 (0.003)	−0.001 (0.003)	−0.001 (0.003)
Gender	−0.149 (0.036)***	−0.158 (0.037)***	−0.029 (0.041)
Non-native	−0.260 (0.065)***	−0.180 (0.079)*	−0.038 (0.083)
Low educated	−0.140 (0.049)**	−0.155 (0.051)**	0.044 (0.054)
High educated	0.049 (0.055)	−0.018 (0.066)	−0.137 (0.069)*
Age ∗ birth year	−0.019 (0.054)	−0.025 (0.054)	−0.033 (0.052)
Age ∗ gender		−0.193 (0.080)*	−0.072 (0.084)
Age ∗ non-native		0.377 (0.148)*	0.382 (0.155)*
Age ∗ low educated		−0.105 (0.106)	−0.049 (0.111)
Age ∗ high educated		0.082 (0.122)	0.090 (0.127)
Birth year ∗ gender		−0.007 (0.008)	−0.005 (0.008)
Birth year ∗ non-native		0.025 (0.014)^	0.024 (0.015)^
Birth year ∗ low educated		−0.006 (0.010)	−0.002 (0.011)
Birth year ∗ high educated		0.015 (0.012)	0.013 (0.012)
Income assessment			−0.583 (0.047)***
Divorced			−0.821 (0.079)***
Widowed			−0.781 (0.080)***
Never married			−0.576 (0.070)***
Have child			0.222 (0.222)**
Excellent health status			0.759 (0.064)***
Fair health status			−0.344 (0.045)***
Poor health status			−0.884 (0.065)***
Part-time work			0.044 (0.065)
Unemployed			−0.420 (0.073)***
Retired			0.059 (0.072)
Social interaction			0.240 (0.022)***
Random-period	0.023 (0.016)^	0.023 (0.016)^	0.022 (0.015)^
Random-cohort	0.009 (0.009)	0.009 (0.009)	0.003 (0.004)
−2 Res log pseudo-likelihood	98,292.313	98,450.663	100,356.5

Continued

Table 2:
Continued

	GB		
	Model 1	Model 2	Model 3
Intercept	0.497 (0.127)**	0.506 (0.122)**	−0.159 (0.118)
Intercept	2.661 (0.130)***	2.675 (0.125)***	2.238 (0.121)***
Age	0.152 (0.067)*	0.070 (0.098)	0.043 (0.112)
Age^2	0.095 (0.133)	0.096 (0.138)	−0.110 (0.166)
Birth year	0.006 (0.007)	−0.002 (0.010)	−0.003 (0.011)
Birth year^2	0.0001(0.001)	0.000 (0.001)	−0.002 (0.002)
Gender	0.144 (0.032)***	0.135 (0.035)***	0.171 (0.037)***
Non-native	−0.162 (0.044)**	−0.192 (0.050)**	−0.072 (0.053)
Low educated	−0.123 (0.039)**	−0.118 (0.042)**	0.106 (0.44)*
High educated	0.214 (0.043)***	0.198 (0.048)***	0.049 (0.051)
Age ∗ birth year	0.010 (0.026)	0.009 (0.027)	−0.028 (0.033)
Age ∗ gender		0.108 (0.075)	0.093 (0.078)
Age ∗ non-native		0.082 (0.104)	0.161 (0.108)
Age ∗ low educated		0.004 (0.089)	0.096 (0.093)
Age ∗ high educated		0.053 (0.100)	−0.027 (0.104)
Birth year ∗ gender		0.019 (0.007)**	0.015 (0.007)*
Birth year ∗ non-native		0.013 (0.010)	0.016 (0.010)
Birth year ∗ low educated		−0.011 (0.009)	−0.004 (0.009)
Birth year ∗ high educated		0.006 (0.010)	−0.002 (0.010)
Income assessment			−0.830 (0.046)***
Divorced			−0.684 (0.055)***
Widowed			−0.904 (0.082)***
Never married			−0.444 (0.059)***
Have child			0.040 (0.051)
Excellent health status			0.599 (0.044)***
Fair health status			−0.504 (0.044)***
Poor health status			−0.841 (0.064)***
Part-time work			0.045 (0.051)
Unemployed			−0.299 (0.084)**
Retired			0.393 (0.069)***
Social interaction			0.290 (0.018)***
Random-period	0.003 (0.003)	0.003 (0.003)	0.005 (0.005)
Random-cohort	0.068 (0.055)	0.058 (0.046)^	0.020 (0.017)
−2 Res log pseudo-likelihood	126,688.981	126,884.011	129,745.022

Continued

Table 2:
Continued

	HU		
	Model 1	Model 2	Model 3
Intercept	−0.953 (0.133)**	−0.943 (0.132)***	−1.226 (0.143)***
Intercept	1.260 (0.133)***	1.273 (0.132)***	1.349 (0.144)***
Age	−0.073 (0.180)	0.048 (0.193)	0.043 (0.163)
Age^2	0.640 (0.018)	0.644 (0.394)$^\wedge$	0.446 (0.319)
Birth year	0.005 (0.018)	0.018 (0.019)	0.005 (0.016)
Birth year^2	0.007 (0.0045)$^\wedge$	0.007 (0.004)$^\wedge$	0.005 (0.003)
Gender	0.151 (0.036)***	0.132 (0.036)**	0.375 (0.039)***
Non-native	0.080 (0.099)	0.123 (0.104)	0.228 (0.109)*
Low educated	−0.472 (0.040)***	−0.470 (0.41)***	−0.074 (0.044)$^\wedge$
High educated	0.440 (0.053)***	0.428 (0.055)***	0.155 (0.057)**
Age ∗ birth year	0.133 (0.079)$^\wedge$	0.133 (0.079)$^\wedge$	0.088 (0.064)
Age ∗ gender		−0.129 (0.080)$^\wedge$	−0.100 (0.083)
Age ∗ non-native		0.477 (0.222)*	0.537 (0.231)*
Age ∗ low educated		−0.088 (0.88)	−0.056 (0.092)
Age ∗ high educated		−0.135 (0.121)	−0.099 (0.126)
Birth year ∗ gender		−0.005 (0.008)	−0.007 (0.008)
Birth year ∗ non-native		0.041 (0.021)$^\wedge$	0.045 (0.022)*
Birth year ∗ low educated		−0.018 (0.009)*	−0.011 (0.009)
Birth year ∗ high educated		−0.013 (0.012)	−0.003 (0.012)
Income assessment			−0.947 (0.042)***
Divorced			−0.585 (0.057)
Widowed			−0.555 (0.065)***
Never married			−0.425 (0.079)***
Have child			0.188 (0.068)**
Excellent health status			0.861 (0.079)***
Fair health status			−0.507 (0.046)***
Poor health status			−1.279 (0.062)***
Part-time work			0.017 (0.077)
Unemployed			−0.172 (0.091)$^\wedge$
Retired			0.160 (0.060)**
Social interaction			0.278 (0.021)***
Random-period	0.057 (0.038)$^\wedge$	0.056 (0.037)$^\wedge$	0.035 (0.027)$^\wedge$
Random-cohort	0.002 (0.004)	0.002 (0.004)	0.002 (0.004)
−2 Res log pseudo-likelihood	80,577.945	80,689.226	85,067.865

Continued

Table 2:
Continued

	IE		
	Model 1	Model 2	Model 3
Intercept	0.155 (0.133)	0.201 (0.130)	−0.635 (0.144)**
Intercept	2.449 (0.136)***	2.499 (0.133)***	1.894 (0.146)***
Age	−0.215 (0.192)	−0.271 (0.200)	−0.331 (0.178)^
Age^2	1.189 (0.411)**	1.166 (0.405)**	0.761 (0.362)*
Birth year	−0.038 (0.019)*	−0.054 (0.02)**	−0.061 (0.018)**
Birth year^2	0.011 (0.004)**	0.011 (0.004)**	0.007 (0.004)*
Gender	0.076 (0.033)*	0.053 (0.034)	0.078 (0.037)*
Non-native	0.021 (0.050)	−0.027 (0.059)	0.155 (0.062)*
Low educated	−0.298 (0.041)***	−0.326 (0.042)***	−0.088 (0.044)*
High educated	0.227 (0.041)***	0.116 (0.047)*	−0.024 (0.049)
Age ∗ birth year	0.229 (0.082)**	0.223 (0.081)**	0.141 (0.072)*
Age ∗ gender		−0.024 (0.08)	−0.006 (0.083)
Age ∗ non-native		0.012 (0.127)	0.013 (0.132)
Age ∗ low educated		−0.014 (0.097)	0.037 (0.101)
Age ∗ high educated		0.270 (0.0102)**	0.268 (0.106)**
Birth year ∗ gender		0.006 (0.008)	0.003 (0.008)
Birth year ∗ non-native		0.009 (0.012)	0.008 (0.012)
Birth year ∗ low educated		0.005 (0.010)	0.010 (0.010)
Birth year ∗ high educated		0.038 (0.010)**	0.034 (0.010)**
Income assessment			−0.656 (0.043)***
Divorced			−0.531 (0.082)***
Widowed			−0.625 (0.090)***
Never married			−0.265 (0.062)***
Have child			0.176 (0.057)**
Excellent health status			0.768 (0.040)***
Fair health status			−0.466 (0.050)***
Poor health status			−0.923 (0.102)***
Part-time work			−0.052 (0.051)
Unemployed			−0.658 (0.066)***
Retired			−0.070 (0.069)
Social interaction			0.273 (0.021)***
Random-period	0.067 (0.042)^	0.063 (0.040)^	0.044 (0.029)^
Random-cohort	0.004 (0.005)	0.003 (0.004)	0.005 (0.006)
−2 Res log pseudo-likelihood	123,900.008	123,894.129	126,807.087

Continued

Table 2:
Continued

	NL		
	Model 1	Model 2	Model 3
Intercept	0.646 (0.569)	0.645 (0.571)	−0.321 (0.537)
Intercept	3.679 (0.572)**	3.681 (0.574)***	2.980 (0.541)**
Age	−2.920 (0.446)***	−3.037 (0.452)***	−2.705 (0.468)***
Age^2	2.124 (0.603)**	2.099 (0.604)**	2.386 (0.637)**
Birth year	−0.293 (0.045)***	−0.306 (0.045)***	−0.271 (0.047)***
Birth year^2	0.022 (0.006)**	0.021 (0.006)**	0.024 (0.006)**
Gender	0.092 (0.039)*	0.072 (0.040)^	0.179 (0.044)***
Non-native	−0.426 (0.061)***	−0.376 (0.068)***	−0.018 (0.074)
Low educated	−0.303 (0.046)***	−0.290 (0.047)***	−0.113 (0.050)*
High educated	0.149 (0.05)**	0.144 (0.055)**	0.034 (0.058)
Age ∗ birth year	0.428 (0.120)**	0.421 (0.121)***	0.479 (0.127)*
Age ∗ gender		0.109 (0.089)	0.138 (0.094)
Age ∗ non-native		0.067 (0.139)	0.137 (0.150)
Age ∗ low educated		−0.013 (0.104)	0.101 (0.110)
Age ∗ high educated		0.197 (0.115)^	0.242 (0.120)*
Birth year ∗ gender		0.020 (0.009)*	0.015 (0.009)^
Birth year ∗ non-native		−0.001 (0.013)	0.002 (0.015)
Birth year ∗ low educated		−0.007 (0.010)	0.002 (0.011)
Birth year ∗ high educated		0.019 (0.011)^	0.022 (0.012)^
Income assessment			−0.881 (0.063)***
Divorced			−0.909 (0.077)***
Widowed			−1.175 (0.104)***
Never married			−0.488 (0.075)***
Have child			0.171 (0.060)**
Excellent health status			0.774 (0.073)***
Fair health status			−0.607 (0.048)***
Poor health status			−1.110 (0.090)***
Part-time			0.008 (0.055)
Unemployed			−0.551 (0.114)***
Retired			0.172 (0.086)*
Social interaction			0.346 (0.022)***
Random-period	2.42 (1.497)*	2.442 (1.505)*	2.084 (1.330)^
Random-cohort	0.006 (0.006)	0.005 (0.005)	0.003 (0.005)
−2 Res log pseudo-likelihood	126,774.107	126,774.406	127,196.860

Continued

Table 2:
Continued

	NO		
	Model 1	Model 2	Model 3
Intercept	0.931 (0.078)***	0.928 (0.078)***	−0.454 (0.132)*
Intercept	3.558 (0.096)***	3.556 (0.095)***	2.410 (0.141)***
Age	0.313 (0.082)**	0.323 (0.114)**	0.313 (0.110)**
Age^2	0.181 (0.201)	0.176 (0.200)	0.212 (0.174)
Birth year	0.029 (0.008)**	0.025 (0.011)*	0.033 (0.011)**
Birth year^2	0.001 (0.002)	0.001 (0.002)	0.002 (0.002)
Gender	0.081 (0.043)^	0.082 (0.044)^	0.149 (0.047)**
Non-native	−0.179 (0.072)*	−0.170 (0.081)*	0.080 (0.087)
Low educated	−0.096 (0.056)^	−0.082 (0.059)	0.178 (0.062)**
High educated	−0.070 (0.053)	−0.078 (0.055)	−0.221 (0.059)**
Age ∗ birth year	0.033(0.040)	0.030 (0.040)	0.040 (0.034)
Age ∗ gender		−0.059 (0.097)	−0.032 (0.102)
Age ∗ non-native		0.352 (0.166)*	0.325 (0.174)^
Age ∗ low educated		−0.044 (0.127)	−0.053 (0.133)
Age ∗ high educated		0.003 (0.117)	−0.016 (0.123)
Birth year ∗ gender		0.000 (0.009)	−0.002 (0.010)
Birth year ∗ non-native		0.030 (0.06)*	0.022 (0.016)
Birth year ∗ low educated		0.000 (0.012)	−0.002 (0.013)
Birth year ∗ high educated		0.002 (0.011)	−0.001 (0.012)
Income assessment			−0.889 (0.082)***
Divorced			−0.724 (0.067)***
Widowed			−0.793 (0.105)***
Never married			−0.573 (0.067)***
Have child			0.400 (0.068)***
Excellent health status			0.636 (0.059)***
Fair health status			−0.460 (0.060)***
Poor health status			−0.931 (0.088)***
Part-time			0.097 (0.078)
Unemployed			−0.754 (0.142)***
Retired			0.290 (0.092)**
Social interaction			0.411 (0.030)***
Random-period	0.007 (0.007)	0.007 (0.007)	0.004 (0.005)
Random-cohort	0.005 (0.006)	0.004 (0.006)	0.001 (0.003)
−2 Res log pseudo-likelihood	93,022.950	93,086.183	95,724.636

Continued

Table 2:
Continued

	PT		
	Model 1	Model 2	Model 3
Intercept	−0.086 (0.319)	−0.046 (0.312)	−0.685 (0.264)*
Intercept	2.730 (0.320)***	2.783 (0.313)***	2.337 (0.265)***
Age	0.364 (0.349)	0.453 (0.365)	0.462 (0.322)
Age^2	−1.279 (0.592)*	−1.181 (0.591)*	−0.419 (0.545)
Birth year	0.054 (0.035)	0.047 (0.036)	0.046 (0.032)
Birth year^2	−0.014 (0.006)*	−0.013 (0.006)*	−0.005 (0.005)
Gender	−0.214 (0.033)***	−0.246 (0.034)***	0.007 (0.035)
Non-native	0.153 (0.072)*	0.210 (0.096)*	0.373 (0.099)**
Low educated	−0.238 (0.051)***	−0.307 (0.067)***	−0.005 (0.069)
High educated	0.282 (0.074)**	0.299 (0.104)**	0.199 (0.106)^
Age ∗ birth year	−0.271 (0.118)*	−0.255 (0.118)*	−0.099 (0.109)
Age ∗ gender		0.184 (0.075)*	0.265 (0.077)**
Age ∗ non-native		0.498 (0.168)**	0.612 (0.173)**
Age ∗ low educated		−0.255 (0.118)*	−0.194 (0.121)
Age ∗ high educated		−0.026 (0.178)	−0.031 (0.181)
Birth year ∗ gender		0.034 (0.007)***	0.035 (0.007)***
Birth year ∗ non-native		0.045 (0.017)**	0.051 (0.017)**
Birth year ∗ low educated		−0.017 (0.011)	−0.018 (0.012)
Birth year ∗ high educated		−0.005 (0.017)	−0.008 (0.017)
Income assessment			−0.574 (0.037)***
Divorced			−0.619 (0.076)***
Widowed			−0.687 (0.072)***
Never married			−0.324 (0.070)***
Have child			0.042 (0.059)
Excellent health status			0.430 (0.077)***
Fair health status			−0.331 (0.040)***
Poor health status			−1.027 (0.057)***
Part-time			−0.109 (0.058)^
Unemployed			−0.146 (0.065)*
Retired			0.078 (0.055)
Social interaction			0.222 (0.020)***
Random-period	0.541 (0.383)^	0.525 (0.374)^	0.293 (0.240)
Random-cohort	0.055 (0.045)	0.033 (0.029)	0.016 (0.015)
−2 Res log pseudo-likelihood	103,419.412	103,808.587	105,352.371

Continued

Table 2:
Continued

	SE		
	Model 1	Model 2	Model 3
Intercept	0.758 (0.061)***	0.760 (0.062)***	−0.131 (0.118)
Intercept	3.357 (0.078)***	3.363 (0.079)***	2.798 (0.127)***
Age	−0.012 (0.050)	0.106 (0.093)	0.062 (0.104)
Age^2	0.273 (0.126)*	0.271 (0.128)*	0.325 (0.155)*
Birth year	−0.004 (0.005)	−0.001 (0.009)	−0.001 (0.010)
Birth $year^2$	0.003 (0.001)*	0.002 (0.001)^	0.003 (0.002)^
Gender	0.057 (0.041)	0.057 (0.041)	0.201 (0.045)***
Non-native	−0.260 (0.057)***	−0.257 (0.059)***	−0.033 (0.064)
Low educated	−0.070 (0.050)	−0.065 (0.052)	0.158 (0.056)**
High educated	0.063 (0.054)	0.034 (0.059)	−0.141 (0.063)*
Age ∗ birth year	0.053 (0.025)*	0.048 (0.025)^	0.053 (0.031)^
Age ∗ gender		−0.188 (0.091)*	−0.227 (0.097)*
Age ∗ non-native		0.212 (0.128)^	0.231 (0.137)^
Age ∗ low educated		−0.209 (0.110)^	−0.167 (0.118)
Age ∗ high educated		0.096 (0.121)	0.117 (0.128)
Birth year ∗ gender		−0.011 (0.009)	−0.020 (0.009)*
Birth year ∗ non-native		0.018 (0.012)	0.015 (0.013)
Birth year ∗ low educated		−0.011 (0.011)	−0.006 (0.012)
Birth year ∗ high educated		0.013 (0.012)	0.014 (0.013)
Income assessment			−0.949 (0.075)***
Divorced			−0.744 (0.064)***
Widowed			−0.992 (0.096)***
Never married			−0.690 (0.060)***
Have child			0.376 (0.063)***
Excellent health status			0.725 (0.058)***
Fair health status			−0.712 (0.055)***
Poor health status			−1.573 (0.093)***
Part-time			0.235 (0.098)*
Unemployed			−0.452 (0.118)**
Retired			0.244 (0.087)**
Social interaction			0.277 (0.025)***
Random-period	0.001 (0.003)	0.001 (0.003)	0.002 (0.004)
Random-cohort	0.003 (0.004)	0.003 (0.004)	0.0001 (0.003)
−2 Res log pseudo-likelihood	97,714.566	97,895.511	102,930.973

Continued

Table 2:
Continued

	SI		
	Model 1	Model 2	Model 3
Intercept	−0.037 (0.101)	−0.041 (0.101)	−0.815 (0.101)**
Intercept	2.487 (0.107)***	2.485 (0.107)***	1.988 (0.107)***
Age	0.025 (0.115)	−0.021 (0.147)	0.060 (0.147)
Age^2	−0.107 (0.286)	−0.109 (0.287)	−0.129 (0.287)
Birth year	0.020 (0.012)^	0.013 (0.015)	0.016 (0.015)
Birth year^2	−0.001 (0.003)	−0.001 (0.003)	−0.001 (0.003)
Gender	0.081 (0.045)^	0.089 (0.046)*	0.267 (0.046)***
Non-native	0.149 (0.067)*	0.144 (0.066)*	0.180 (0.066)**
Low educated	−0.550 (0.051)***	−0.545 (0.051)***	−0.216 (0.051)***
High educated	0.404 (0.073)***	0.402 (0.078)***	0.223 (0.078)**
Age ∗ birth year	−0.027 (0.057)	−0.026 (0.057)	−0.035 (0.057)
Age ∗ gender		0.018 (0.102)	0.059 (0.102)
Age ∗ non-native		−0.039 (0.148)	0.0001(0.148)
Age ∗ low educated		0.098 (0.113)	0.058 (0.113)
Age ∗ high educated		−0.057 (0.164)	−0.088 (0.164)
Birth year ∗ gender		0.004 (0.010)	0.000 (0.010)
Birth year ∗ non-native		−0.011 (0.014)	0.000 (0.014)
Birth year ∗ low educated		0.013 (0.011)	0.010 (0.011)
Birth year ∗ high educated		−0.004 (0.016)	−0.005 (0.016)
Income assessment			−0.714 (0.150)***
Divorced			−0.636 (0.153)***
Widowed			−0.696 (0.141)***
Never married			−0.452 (0.257)***
Have child			0.189 (0.014)*
Excellent health status			0.760 (0.003)***
Fair health status			−0.637 (0.049)***
Poor health status			−1.142 (0.069)***
Part-time			0.097 (0.054)
Unemployed			−0.237 (0.082)*
Retired			0.182 (0.106)*
Social interaction			0.316 (0.154)***
Random-period	0.018 (0.014)^	0.018 (0.014)^	0.013 (0.011)
Random-cohort	0.006 (0.007)	0.006 (0.007)	0.004 (0.237)
−2 Res log pseudo-likelihood	60,276.65	60,313.82	62,773.84

Note: ^$p < .1$; *$p < .05$; **$p < .01$; ***$p < .0001$.

Figure 3:
Graphical plot of the estimated cohort and period effects on high happiness [overall 95% confidence interval (low/high)] in ten countries based on model 1

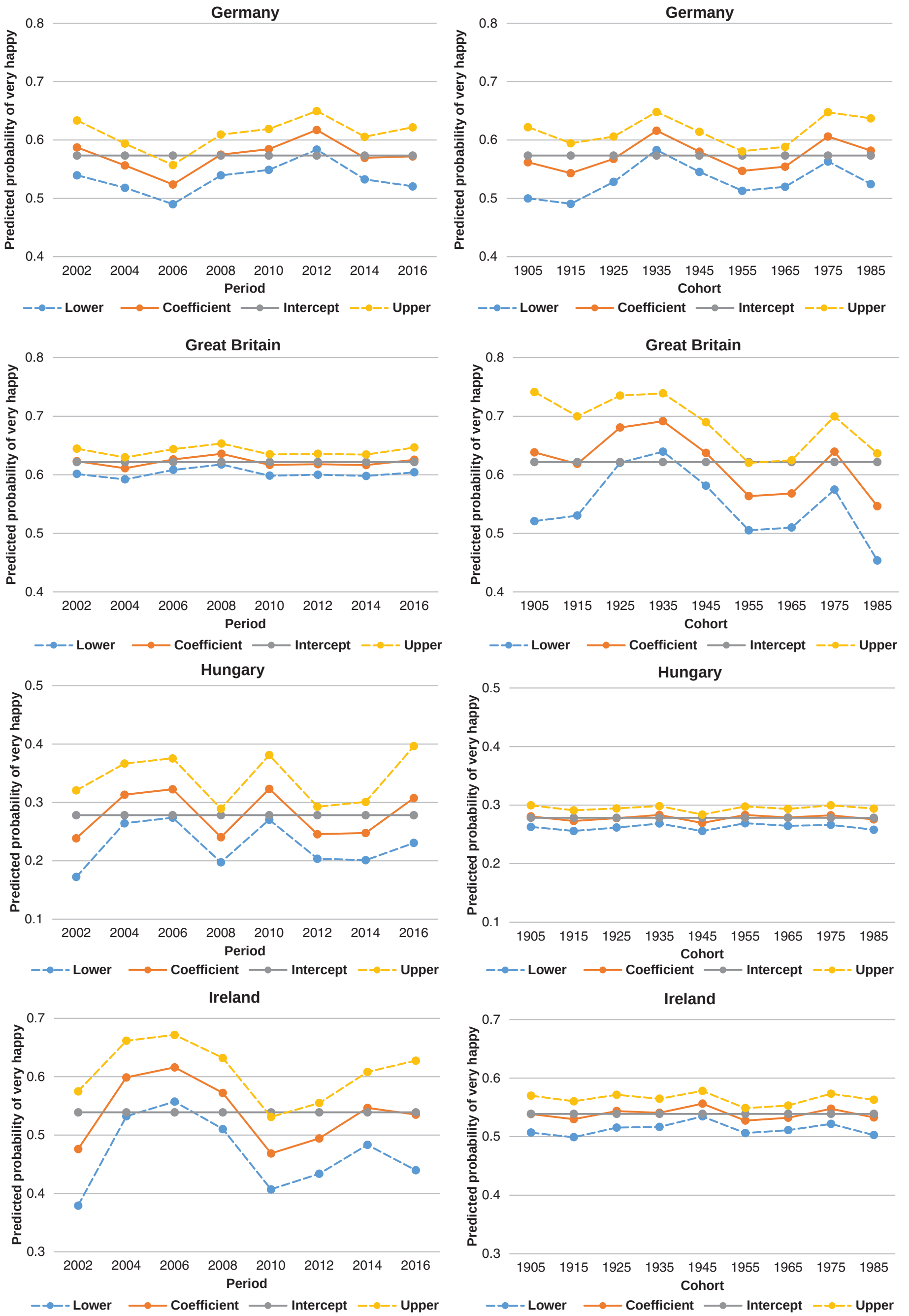


Figure 3: Continued

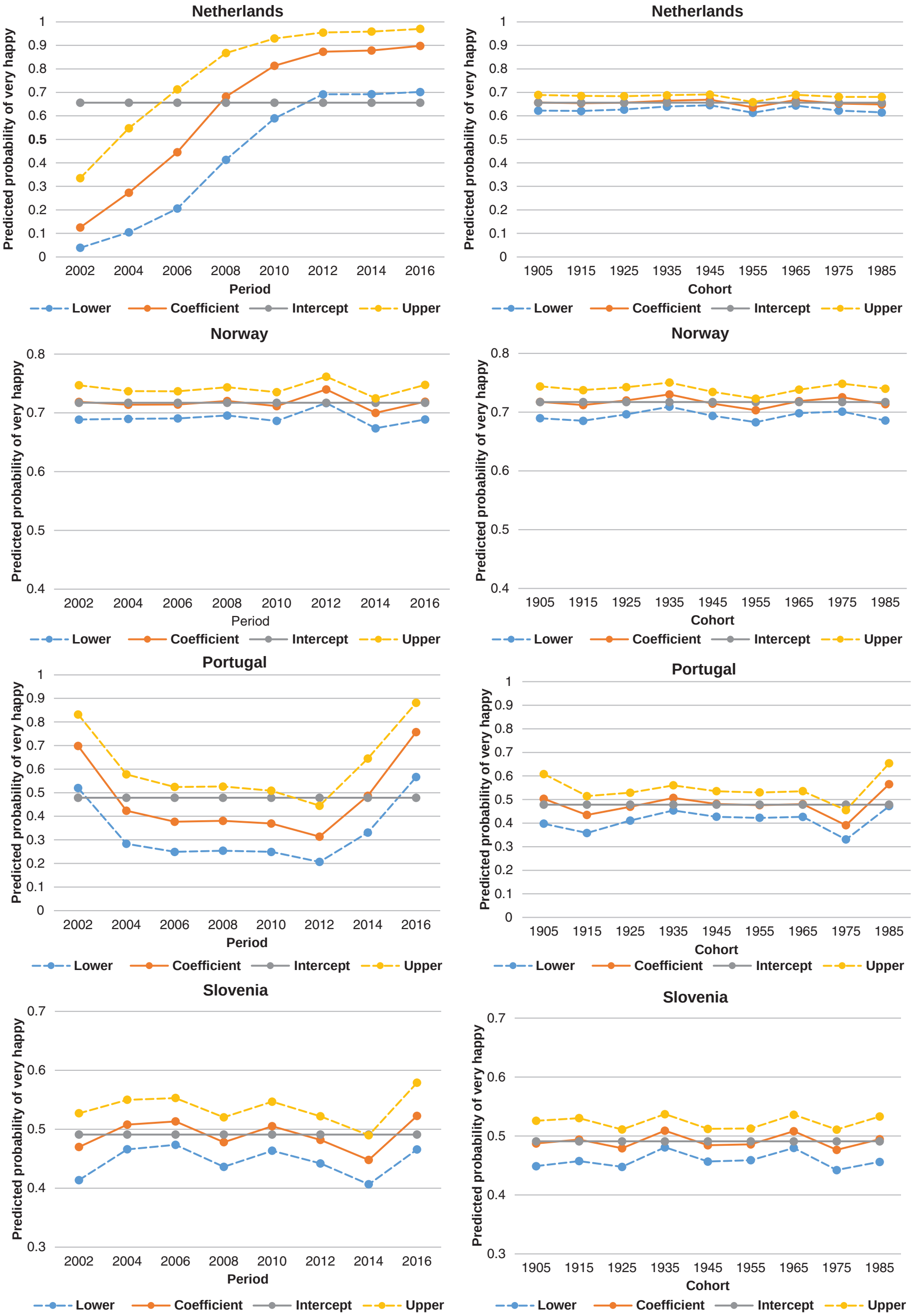
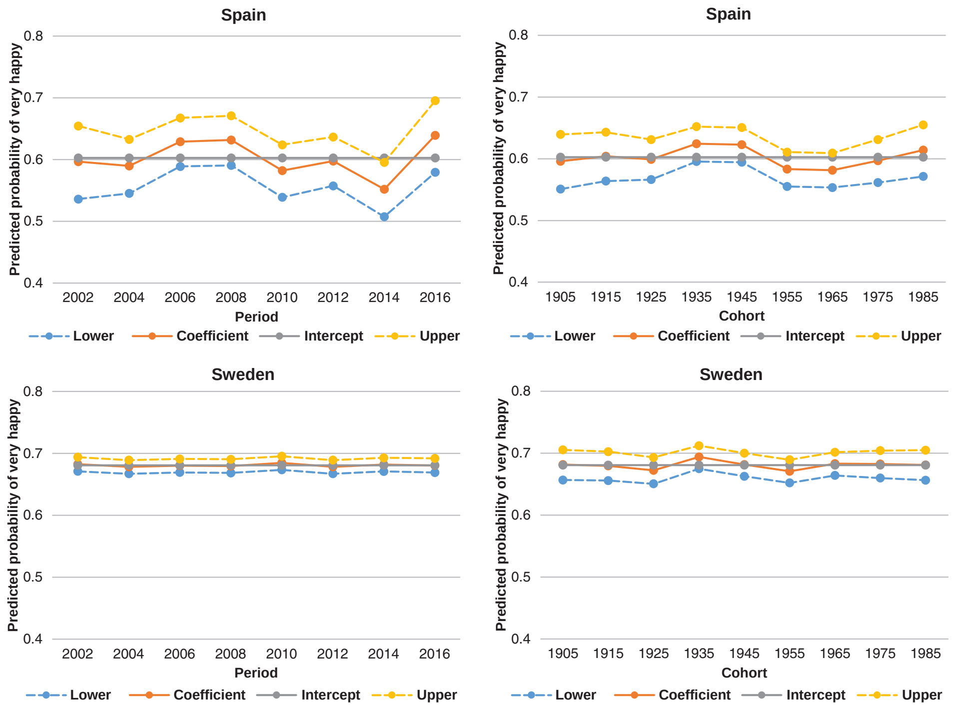


Figure 3:
Continued



Note: The corresponding figures in each country graphically plot the estimated cohorts and time period effects within the 95 percent confidence bounds.

than early baby boomers (i.e., those born in 1945–1954). These relatively high levels of happiness found for pre-baby boomers are in line with our first assumption. Furthermore, even though our models assume no period linear trends, we find lower happiness in the study period (2002–2016), with a decline around 2008 or thereafter in some countries (ES, HU, IE, PT), followed by a recovery (NL, PT).

4.3 Core socio-economic determinants of happiness

We now turn our attention to the core socio-economic determinants shown in model 1 in Table 2. On average, as expected (H2a), women, native people and people with higher levels of educational attainment are more likely to be very happy than men, non-native people and people with lower levels of educational attainment. However, the evidence for Southern European countries is mixed. For some countries, negative effects are found for being female (ES, PT), but positive effects are found for being non-native (PT). The age effects are observed to be larger for more recent cohorts in several countries (DE, HU, IE, NL, SE), but in Portugal, the opposite pattern is found (i.e., age effects are larger for less recent cohorts).

As expected, the effects of socio-economic characteristics change over the life course (Inglehart 2002; Ferraro and Shippee 2009), as presented in Model 2. Here, we find a more mixed picture than we anticipated (H2b), with growing (dis)advantages for certain socio-economic characteristics in some countries, and decreasing gaps for other types of socio-economic characteristics in other countries. First, we observe that the differences in levels of happiness due to gender evolve over the life course. In some countries (DE, HU), women do not remain happy throughout their life course. We also find that women in two Southern European countries have distinct experiences. In Portugal, women are generally not happier than men, but variations are observed at older ages. In Spain, women are generally unhappy, and remain so over their life course. When we look at place of origin, we see that in one Southern European country (PT), non-natives have higher levels of happiness throughout the life course. In other countries, the low levels happiness associated with being non-native disappear (DE, GB, NL), or turn into high levels happiness (ES, NO, SE), showing a strong trend towards convergence. Such cross-over trends are also observed for education. The association of lower levels of educational attainment with lower levels of happiness disappears over the life course in many countries. However, in one Western European country (DE), people with lower levels of educational attainment are less happy when they are middle-aged, but are happier at older ages. In one Southern European country (PT), people with lower levels of educational attainment remain consistently less happy, while lower educated people in one Northern European country (SE) become less happy at older ages. Highly educated people remain happier throughout the life course in non-continental and Western European countries (IE, NL). However, there are crossovers in age effects in Germany, where the highly educated are not happy at older ages. The effects of core socio-economic characteristics change by cohort, but to a much lesser extent than in the case of age.

4.4 Additional socio-economic, demographic and behavioural determinants of happiness, and time variations

In full model 3, additional determinants – income assessment, employment status, partnership status, child status, health status and social interaction behaviour – are introduced. In line with the second hypothesis, all of these determinants produce significant and consistent effects on happiness. As expected, we observe that happiness is highly dependent on socio-economic and demographic characteristics, and our findings are largely in line with previous empirical research (Angeles 2010; Niedzwiedz et al. 2015; Steptoe et al. 2015; Glass et al. 2016; Read et al. 2016; Jenkinson et al. 2020). A negative assessment of income has a consistent negative effect on happiness. Those individuals who are divorced, widowed or never married are less likely to be happy than those who are married. Unemployed people are also less likely to be happy, while retired people are more likely to be happy in selected countries across regions (GB, HU, NO, SE, SI). People in poor or even fair health are less likely to have high levels of happiness than those reporting excellent health. People who are more socially active are more likely to be happy than those who are less active. Being a parent has a positive effect in all countries except Germany, Great Britain and Portugal, although the size of the effect differs across countries.

By including additional covariates, we can show changes in the aforementioned time and core covariates (i.e., gender, native/non-native, levels of educational attainment) and interaction effects found in previous model 2. This implies that core socio-economic covariates are mediated by additional characteristics that are controlled for, in particular income assessment, marital status, employment status and health status. Model 3 shows that for the time factors, the same age patterns are generally observed, albeit with decreased net effects for period and cohort. For the age effect, the same trend is generally found. Even though fixed effects of age and polynomial age do not always show significant effects for all countries, a U-shaped relationship between age and happiness emerges for one Southern European country (ES), while a non-U-shaped pattern is observed for a non-continental Western European country (GB). The negative age trends found for another Southern European country (PT) point to a slight recovery at older ages, although the effects are not significant.

The associations of the core socio-economic covariates vary. Figure A.1 (Appendix) illustrates the probabilities of being very happy from model 3 in Table 2, which compares the age trajectories by the core socio-economic characteristics (i.e., gender, place of origin and levels of education). Female advantages are now observed in all countries except one Southern European country (ES). The negative association with being non-native disappears in most countries, while being non-native remains an advantage in Southern and Eastern European countries only (PT, SI), and becomes an advantage in Eastern and non-continental Western European countries (HU, IE). The disadvantage of being less educated remains for a few countries, extends to more countries (DE, HU, IE, NL, SI), and disappears for Southern European countries (ES, PT), but turns into an advantage for Western-continental and Northern European countries (GB, NO, SE). The advantage of being highly educated remains for countries across regions (DE, HU, PT, SI), but disappears for other non-continental Western and Western European countries (GB, IE, NL), and appears as a disadvantage in still other countries across regions (ES, NO, SE). As we can see, these changes in associations do not reflect the geographical classification of countries.

The interaction effects between age and core socio-economic characteristics in model 3 exhibit substantial country-specific patterns. When controlling for additional covariates, the female disadvantages at older ages disappear for several countries (DE, ES, HU). We find, for example, that Portuguese women are happier, but Swedish women not. Interestingly, non-natives are happier than natives in a few countries across regions (ES, HU, NO, PT, SE), whereas low levels of educational attainment and happiness levels are no longer significant. This educational gap also decreases among the highly educated as changes of coefficients are observed for Western European countries. For instance, highly educated people in Germany are unhappy, whereas highly educated people in the Netherlands and Ireland remain happier than less educated people. Gaps in levels of happiness by core socio-economic characteristics vary by cohort, but to a much lesser extent than by age, as was the case for model 2. The observation that women, native people and highly educated people are not happier than their counterparts over the life course (age), and to some extent in more recent generations (cohort), is partially explained by marital status, partnership and child status, health status and income assessment.

5 Discussion and conclusion

This paper aimed to show the trajectories of subjective well-being (SWB) through measures of happiness, as influenced by time and socio-economic, demographic and behavioural determinants. We approached this research objective by seeking to answer two research questions: 1. What are the effects of different dimensions of time (age, period, cohort) on people's levels of happiness? 2. And, do we observe differences in people's happiness levels depending on their socio-economic, demographic and behavioural characteristics, controlled for time effects?

Our results show a U-shaped relationship between age and happiness for some countries, but a rather flat age pattern for others. Some countries exhibit positive age effects at higher ages (80+), while others show negative effects. Lower happiness levels are generally found for baby boomers, but variations within the baby boom generation (1945–1964) are also observed (i.e., lower levels of happiness for late boomers (1955–1964) than for earlier boomers (1945–1954). No linear trend over time in period residuals is assumed, while period contextual effects are found for some countries (all except Northern European countries) in relation to the economic crisis (2008) and thereafter. Happiness levels are affected by individual characteristics controlled by time: on average, women, higher educated people and native people have higher levels of happiness than men, lower educated people and non-native people, respectively. Moreover, a positive income assessment, having a partner, being a parent, being employed, being in good health and being socially active all have positive effects on levels of happiness. When controlling for these additional covariates (i.e., marital status, partnership and child status, employment and health status and own income assessment), the association between core socio-economic characteristics and happiness weakens. For instance, on average, the advantage of being a women remains, but mixed results are observed for native/non-native status and levels of educational attainment. We find that in some countries, women, native people and highly educated people are not happier than their counterparts over the life course (age), and to some extent in more recent generations (cohort). These variations can be partially explained by marital and child status, health status and own income assessment.

Some findings are in line with our empirical research, while others are not. Some evidence of a U-shaped relationship between age and happiness is found for some countries, but not all countries follow this pattern. Other studies have identified negative happiness by age trends for Eastern European countries (Steptoe et al. 2015), but our study found such results for only one Southern European country (PT). Recent empirical research (Blanchflower 2021) reconfirming a U-shaped age-SWB pattern globally appears to be in contrast with our findings on different age trends of happiness (i.e., non-U-shaped). The happiness variations among older seniors in our findings provide new insights into happiness-age trends, while cohort effects, and particularly low happiness levels among baby boomers, are reconfirmed (Barrett and Toothman 2014), albeit with cross-national differences. Our observations that levels of happiness are higher among women, native people and highly educated people than among their counterparts are supported by other research. However, after controlling for additional covariates (i.e., marital status, partnership and child status, employment and health status, and own income assessment), the relationship between core socio-economic characteristics and happiness levels weakens. Moreover, the degrees of these levels varied across countries. Our finding that time and heterogeneous life course happiness trajectories have country-specific trends, rather than geographical characteristics, supports previous findings (Helliwell et al. 2019). For instance, inequalities in happiness due to levels of education are noted even in Western European countries, where reasonable social welfare provisions are in place. The existing happiness gaps among the highly educated in these countries deserve further attention. Moreover, the decreasing happiness gaps by educational level found in all other countries may simply be due to varying age affects over the life course resulting in less heterogeneity and non-differential happiness trends at older ages.

The key strengths of our analysis include the use of semi-longitudinal and cross-nationally comparable data and the application of (extended) hierarchical age-period-cohort models to examine comparatively the consistency of results on time trends of happiness. Another strength was our use of a wide range of measures, including core socio-economic characteristics (i.e., gender, non-native, educational levels) and additional socio-economic, demographic and behavioural characteristics in the model for each country. The latter allowed us to study comparatively the social differentials of happiness levels and their varying age effects.

There are a few points to consider. The measure of self-reported happiness used in this paper covered only one side of SWB. In order to capture the full multi-dimensional breadth of the concept, future studies should also take into account eudemonic (i.e., human flourishing or prosperity) measures and objective health measures of physical capabilities (i.e., mental health) (Abdallah et al. 2011; OECD 2013). Survival bias may have partially affected our results, in particular the non-U-shaped relation between age and SWB for some countries, leading to variations in age effects among older seniors. In some countries, the survival status of the elderly is likely to also be associated with their higher happiness levels and their attention to emotional goals and experiences, as socioemotional selectivity theory explains (Carstensen et al. 2003). The proportion of the population who are institutionalised is also higher among the elderly, and this proportion differs substantially across countries because of diverse living arrangements cross-nationally, which creates additional biases (Schanze and Levinson 2019). In other words, the socio-economic and demographic differentials among the older age group are considered to be smaller in Northern and Western European countries than in Southern and Eastern European countries. In addition, we do not deny that country-specific cultural factors may affect how people respond to survey questions on happiness. We have not fully explored measurement equivalence across countries (Davidov et al. 2014), and our constructed three-point happiness scales may have discounted small differences in levels of happiness, although this choice was made because of concentrated response choices in the original scales. Moreover, as our findings are based on assumptions by constraining period trends while examining age and cohort effects in the hierarchical age-period-cohort (HAPC) model, the period effects of, for instance, the economic crisis may be undermined for some countries. However, the experiences of people in their earlier years of life are considered to be larger than the general time period effects influencing all ages. Finally, we used a semi-longitudinal survey design to perform an extension of the HAPC analysis, even though true individual changes over time can only be studied through surveys with a longitudinal design with repeated measurements on the same individual.

Our study has important implications for future research. Inequalities in happiness that are associated with people's experiences and accumulation of socio-economic and demographic characteristics during adulthood can widen later in life, and this trend may be stronger among younger generations. Future research should focus on the areas that are negatively influencing the socio-economic and demographic determinants of happiness among various sub-groups of the population in order to improve the chances that these happiness gaps will decrease later in life. This kind of research will be particularly relevant as the late baby boomers – a more socio-economically heterogeneous group than the early boomers – join the young senior group in the coming decade. Insights we obtain about age-happiness trends can provide additional useful lessons for the younger generations, such as Generations X, Y (Millennials) and Z, who are undergoing new and different challenges in adulthood (e.g., COVID-19) that were not experienced by previous generations.

Acknowledgements

An early version of this paper was presented at the UNECE Second International Seminar on Active Ageing Index, held in Bilbao, Spain, on 27–28 September 2018. The research was funded by the University of Leuven Geconcerteerde Onderzoeksactie (GOA) project *New Approaches to the Social Dynamics of Long-term Fertility Change* (Grant GOA/14/001, https://soc.kuleuven.be/ceso/fapos/nasdltfc/index). Finally, we are thankful for the many substantial and insightful comments and suggestions provided by two anonymous reviewers, the editor and the editorial team that improved this article.

References

Aassve, A., L. Mencarini and M. Sironi 2015. Institutional change, happiness and fertility. *European Sociological Review* 31(6): 749–765. https://doi.org/10.1093/esr/jcv073

Abdallah, S., S. Mahony, N. Marks, J. Michaelson, C. Seaford, L. Stoll and S. Thompson 2011. *Measuring our progress: The power of well-being*. Centre for Well-being. The New Economics Foundation. Download from https://neweconomics.org/uploads/files/Measuring_our_Progress.pdf, accessed on 20 June 2016.

Angeles, L. 2010. Children and life satisfaction. *Journal of Happiness Studies* 11: 523–538. https://doi.org/10.1007/s10902-009-9168-z

Barrett, A. E. and E. L. Toothman 2014. Babyboomers' subjective life course and its physical health effects: How distinctive is the "forever young" cohort? *International Journal of Ageing and Human Development* 79(2): 109–129. https://doi.org/10.2190/AG.79.2.b

Bell, A. and K. Jones 2013. The impossibility of separating age, period and cohort effects. *Social Science & Medicine* 93: 163–165. https://doi.org/10.1016/j.socscimed.2013.04.029

Bell, A. 2014. Life course and cohort trajectories of mental health in the UK 1991–2008 – A multilevel age-period-cohort analysis. *Social Science & Medicine* 120: 21–30. https://doi.org/10.1016/j.socscimed.2014.09.008

Bell, A. and K. Jones 2015. Age, period and cohort processes in longitudinal and life course analysis: A multilevel perspective. In *A life course perspective on health trajectories and transitions. Life course research and social policies*, eds C. Burton-Jeangros, S. Cullati, A. Sacker and D. Blane, vol. 4. 197–213. Cham: Springer. https://doi.org/10.1007/978-3-319-20484-0_10

Blanchflower, D. G. and A. J. Oswald 2008. Is well-being U-shaped over the life cycle? *Social Science & Medicine* 66(8): 1733–1749. https://doi.org/10.1016/j.socscimed.2008.01.030

Blanchflower, D. G. 2021. Is happiness U-shaped everywhere? Age and subjective well-being in 145 countries. *Journal of Population Economics* 34: 575–624. https://doi.org/10.1007/s00148-020-00797-z.

Brown, S. L. and Wright, M. R. 2017. Marriage, cohabitation and divorce in later life. *Innovation in Ageing* 1(2): 1–11. https://doi.org/10.1093/geroni/igx015

Carstensen, L. L., H. H. Fung and S. T. Charles 2003. Socioemotional selectivity theory and the regulation of emotion in the second half of life. *Motivation and Emotion* 27(2): 103–123. https://doi.org/10.1023/A:1024569803230

Clark, A. 2007. Born to be mild? Cohort effects don't (fully) explain why well-being is U-shaped in age. Institute for the Study of Labour. *Discussion Paper Series*. No. 3170. Download from http://www.parisschoolofeconomics.com/clark-andrew/CohortNov2007.pdf, accessed on 17 September 2017.

Dannefer, D. 2003. Cumulative advantage/disadvantage and the life course: Cross-fertilizing age and social science theory. *Journal of Gerontology: Series B* 58(6): S327–S337. https://doi.org/10.1093/geronb/58.6.S327

Davidov, E., B. Meuleman, J. Cieciuch, P. Schmidt and J. Billiet 2014. Measurement equivalence in cross-national research. *Annual Review of Sociology* 40: 55–75. https://doi.org/10.1146/annurev-soc-071913-043137

Deeming, C. 2013. Addressing the social determinants of subjective wellbeing: The latest challenge for social policy. *Journal of Social Policy* 42(3): 541–565. https://doi.org/10.1017/S0047279413000202

Diener, E. (ed) 2009. *The science of well-being: The collected works of Ed Diener*. Dordrecht: Springer. https://doi.org/10.1007/978-90-481-2350-6

Easterlin, R. A., C. M. Schaeffer and D. J. Macunovich 1993. Will the baby boomers be less well off than their parents? Income, wealth, and family circumstances over the life cycle in the United States. *Population and Development Review* 19(3): 497–522. https://doi.org/10.2307/2938464

Easterlin, R. A. 2010. *Happiness, growth, and the life cycle*, eds H. Hinte and K. F. Zimmermann. Oxford: Oxford University.

European Commission. 2014. *Population ageing in Europe: Facts, implications and policies.* Directorate General for Research and Innovation. Directorate B Innovation Union and European Research Area Unit B6. Reflective Societies. https://doi.org/10.2777/60452

European Social Survey. 2016. *ESS documentation report.* Norwegian Social Science Data Services. Download from https://www.europeansocialsurvey.org/data/download.html?r=8

European Social Survey. 2018. *ESS1-8 Data Protocol*. European Social Survey Data Archive, Norwegian Social Science Data Services. Download from https://www.europeansocialsurvey.org/data/download.html?r=8

Ferraro, K. F. and T. P. Shippee 2009. Aging and cumulative inequality: How does inequality get under the skin? *The Gerontologist* 49(3): 333–343. https://doi.org/10.1093%2Fgeront%2Fgnp034

Frenk, S. M., Y. C. Yang and K. C. Land 2013. Assessing the significance of cohort and period effects in hierarchical age-period-cohort models: Applications to verbal test scores and voter turnout in U.S. presidential elections. *Social Forces* 92(1): 221–248. https://doi.org/10.1093/sf/sot066

Frijters, P. and T. Beatton 2012. The mystery of the U-shaped relationship between happiness and age. *Journal of Economics Behaviour and Organization* 82(2–3): 525–542. https://doi.org/10.1016/j.jebo.2012.03.008

Fukuda, K. 2013. A happiness study using age-period-cohort framework. *Journal of Happiness Studies* 14: 135–153. https://doi.org/10.1007/s10902-011-9320-4

Giele, J. Z. and G. H. Elder, Jr. (eds) 1998. *Methods of life course research: Qualitative and quantitative approaches*. Thousand Oaks, London, New Delhi: Sage Publications.

Glass, J., R. W. Simon and M. A. Andersson 2016. Parenthood and happiness: Effects of work-family reconciliation policies in 22 OECD countries. *American Journal of Sociology* 122(3): 886–929. https://doi.org/10.1086/688892

Hadjar, A. and S. Backes 2013. Migration background and subjective well-being a multilevel analysis based on the European Social Survey. *Comparative Sociology* 12: 645–676. https://doi.org/10.1163/15691330-12341279

Helliwell, J., R. Layard and Sachs, J. D. eds 2019. *World Happiness report 2019.* New York, Sustainable Development Solutions Network. Download from https://worldhappiness.report/ed/2019/, accessed on 18 April 2019.

Huppert, F. A. 2009. Psychological well-being: Evidence regarding its causes and consequences. *Applied Psychology: Health and Well-being* 1(2): 137–164. https://doi.org/10.1111/j.1758-0854.2009.01008.x

Inglehart, R. 2002. Gender, ageing, and subjective well-being. *International Journal of Comparative Sociology* 43(3–5): 391–408. https://doi.org/10.1177/002071520204300309

Jivraj, S., J. Nazroo, B. Vanhoutte and T. Chandola 2014. Aging and subjective well-being in later life. The *Journals of Gerontology, series B* 69(6): 930–941. https://doi.org/10.1093/geronb/gbu006

Jenkinson, S., H. Matsuo and K. Matthijs 2020. Multi-dimensional Subjective Wellbeing and Lone Parenthood following Divorce in Flanders (Northern Belgium). In *Divorce in Europe: New insights in trends, causes and consequences of relation break-ups*, ed D. Mortelmans, 211–236. European Studies of Population 21, Cham: Springer. https://doi.org/10.1007/978-3-030-25838-2_10

Lesthaeghe, R. 2014. The second demographic transition: A concise overview of its development. *PNAS* 111(51): 18112–18115. https://doi.org/10.1073/pnas.1420441111

Lin, I.-F. and S. L. Brown 2012. Unmarried boomers confront old age: A national portrait. *The Gerontologist* 52(2): 153–165. https://doi.org/10.1093/geront/gnr141

Margolis, R. and M. Myrskylä 2011. A global perspective on happiness and fertility. *Population Development Review* 37(1): 29–56. https://doi.org/10.1111/j.1728-4457.2011.00389.x

Mencarini, L., D. Vignoli, T. Zeydanli and J. Kim 2018. Life satisfaction favours reproduction. The universal positive effect of life satisfaction on childbearing in contemporary low fertility countries. *PLoS ONE* 13(12): e0206202. https://doi.org/10.1371/journal.pone.0206202

Michalos, A. C. 2008. Education, happiness and wellbeing. *Social Indicators Research* 87: 347–366. https://doi.org/10.1007/s11205-007-9144-0

Myrskylä, M. and R. Margolis 2014. Happiness: Before and after the kids. *Demography* 51: 1843–1866. https://doi.org/10.1007/s13524-014-0321-x

Niedzwiedz, C. L., J. P. Pell and R. Mitchell 2015. The relationship between financial distress and life course socio-economic inequalities in well-being: Cross-national analysis of European Welfare States. *American Journal of Public Health* 105(10): 2090–2098. https://doi.org/10.2105/AJPH.2015.302722

OECD 2013. *OECD guidelines on measuring subjective well-being*. OECD Publishing: Paris. https://doi.org/10.1787/9789264191655-en

OECD 2017. *How's life? 2017 Measuring well-being*. OECD Publishing: Paris. https://doi.org/10.1787/23089679

OECD 2018. *Health at a glance: Europe 2018: State of health in the EU cycle*. OECD Publishing: Paris. https://doi.org/10.1787/23056088

OECD 2020. *How's life? 2020 Measuring well-being*. OECD Publishing: Paris. Download from https://doi.org/10.1787/9870c393-en, accessed on 16 October 2020.

Oris, M., R. Gabriel, G. Ritschard and M. Kliegel 2017. Long lives and old age poverty: Social stratification and life course institutionalization in Switzerland. *Research in Human Development* 14(1): 68–87. https://doi.org/10.1080/15427609.2016.1268890.

Ortiz-Ospina, E. and M. Roser 2017. Happiness and Life Satisfaction. Published online at OurWorldInData.org. Download from https://ourworldindata.org/happiness-and-life-satisfaction, accessed on 28 May 2019.

Pacek, A. and B. Radcliff 2008. Assessing the welfare state: The politics of happiness. *Perspectives on Politics* 6(2): 267–277. https://doi.org/10.1017/S1537592708080602

Pinquart, M. and S. Sörensen 2000. Influences of socioeconomic status, social network, and competence on subjective well-being in later life: A meta-analysis. *Psychology and Aging* 15(2): 187–224. https://doi.org/10.1037/0882-7974.15.2.187

Plant, E. A., K. C. Kling and G. L. Smith 2004. The influence of gender and social role on the interpretation of facial expressions. *Sex Roles* 51: 187–196. https://doi.org/10.1023/B:SERS.0000037762.10349.13

Read, S., E. Grundy and E. Foverskov 2016. Socio-economic position and subjective health and well-being among older people in Europe: A systematic narrative review. *Aging & Mental Health* 20(5): 529–542. https://doi.org/10.1080/13607863.2015.1023766

Rowe, J. W. and R. L. Kahn 1997. Successful Ageing. *The Gerontologist* 37(4): 433–440. https://doi.org/10.1093/geront/37.4.433

Ryder, N. B. 1965. The cohort as a concept in the study of social change. *American Sociological Review* 30(6): 843–861. https://doi.org/10.2307/2090964

Sand, G. and S. Gruber 2018. Differences in subjective well-being between older migrants and natives in Europe. *Journal of Immigrant Minority Health* 20: 83–90. https://doi.org/10.1007/s10903-016-0537-5

Schanze, J. and M. Levinson 2019. Report on what persons live in institutions and the most relevant institution types they live in, the availability of data to select them and their ability to be reached in practice. *Deliverable 2.17* of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221. Available at: www.seriss.eu/resources/deliverables

Shanahan, M. J., J. T. Mortimer and M. K. Johnson (eds) 2016. *Handbook of the life course*. vol. II. New York: Springer.

Skevington, S. M. and J. Böhnke 2018. How is subjective well-being related to quality of life? Do we need two concepts and both measures? *Social Science & Medicine* 206: 22–30. https://doi.org/10.1016/j.socscimed.2018.04.005

Snijders, T. A. B. and R. J. Bosker 1999. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Spruk, R. and A. Kešeljević 2016. Institutional origins of subjective well-being: Estimating the effects of economic freedom on national happiness. *Journal of Happiness Studies* 17: 659–712. https://doi.org/10.1007/s10902-015-9616-x

Stepler, R. 2017. Led by Baby Boomers, divorce rates climb for America's 50+ population. Washington, DC: Fact Tank: Pew Research Center. https://www.pewresearch.org/fact-tank/2017/03/09/led-by-baby-boomers-divorce-rates-climb-for-americas-50-population/, accessed on 16 October 2020.

Steptoe, A., A. Deaton and A. A. Stone 2015. Subjective wellbeing, health and ageing. *Lancet* 385: 640–648. https://doi.org/10.1016/S0140-6736(13)61489-0

Stiglitz, J. E., A. Sen and J. P. Fitoussi 2009. Report by the Commission on the measurement of economic performance and social progress. Paris: Commission on the Measurement of Economic Performance and Social Progress.

Thompson, S. and N. Marks 2008. *Measuring well-being in policy: Issues and applications*. London: New Economics Foundation. Download from https://b.3cdn.net/nefoundation/575659b4f333001669_ohm6iiogp.pdf, accessed on 29 April 2015.

UNECE/EC 2019. *Active ageing index 2018: Analytical report*. Report prepared by G. Lamura and A. Principi under contract with United Nations Economic Commission for Europe (Geneva), co-funded by European Commission's Directorate General for Employment, Social Affairs and Inclusion (Brussels). Download from https://www.unece.org/population/aai.html, accessed on 15 July 2019.

Van de Velde, S., P. Bracke and K. Lavecque 2010. Gender differences in depression in 23 European countries. Cross-national variation in the gender gap in depression. *Social Science & Medicine* 71(2): 305–313. https://doi.org/10.1016/j.socscimed.2010.03.035

Vanhoutte, B. 2014. The multidimensional structure of subjective well-being in later life. *Population Ageing* 7: 1–20. https://doi.org/10.1007/s12062-014-9092-9

Welsch, H. and J. Kühling 2016. Macroeconomic performance and institutional change: Evidence from subjective well-being data. *Journal of Applied Economics* 11(2): 193–218. https://doi.org/10.1016/S1514-0326(16)30008-3

WHO 2015. *World report on ageing and health.* World Health Organization. https://apps.who.int/iris/handle/10665/186463

Witter, R. A., M. A. Okun, W. A. Stoc and M. J. Haring 1984. Education and subjective well-being: A meta-analysis. *Educational Evaluation and Policy Analysis* 6(2): 165–173. https://doi.org/10.3102%2F01623737006002165

Yang, Y. 2008. Social inequalities in happiness in the United States, 1972 to 2004: An age-period-cohort analysis. *American Sociological Review* 73(2): 204–226. https://doi.org/10.1177/000312240807300202

Yang, Y. and K. Land 2013. *Age-period-cohort analysis, new models, methods and empirical applications*. Boca Raton: Taylor and Francis group.

Zaidi, A., K. Gasior, E. Zolyomi, A. Schmidt, R. Rodrigues and B. Marin 2017. Measuring active and healthy ageing in Europe. *Journal of European Social Policy* 27(2): 138–157. https://doi.org/10.1177/0958928716676550

Appendix

Figure A.1:
(a) Predicted age variations in gender effects on happiness based on model 3 in ten countries.

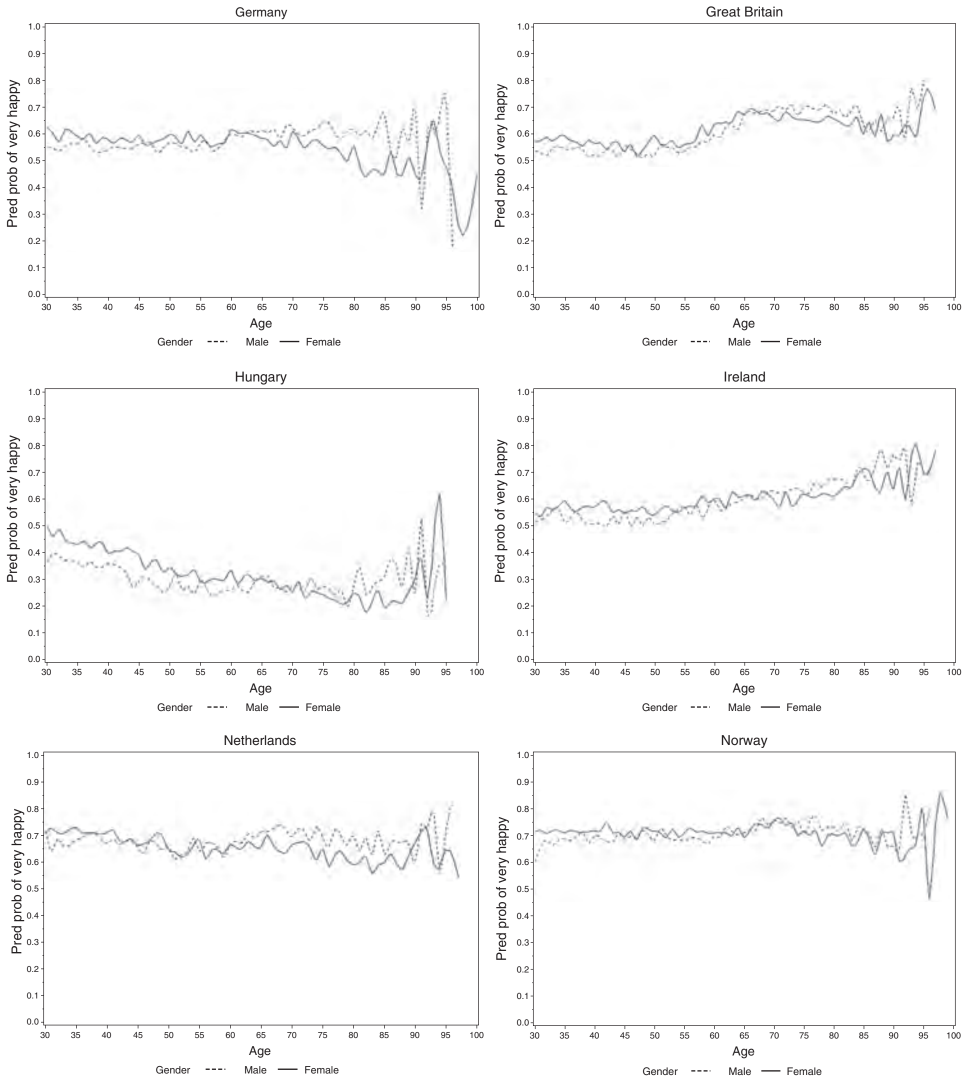


Figure A.1:
(a) Continued

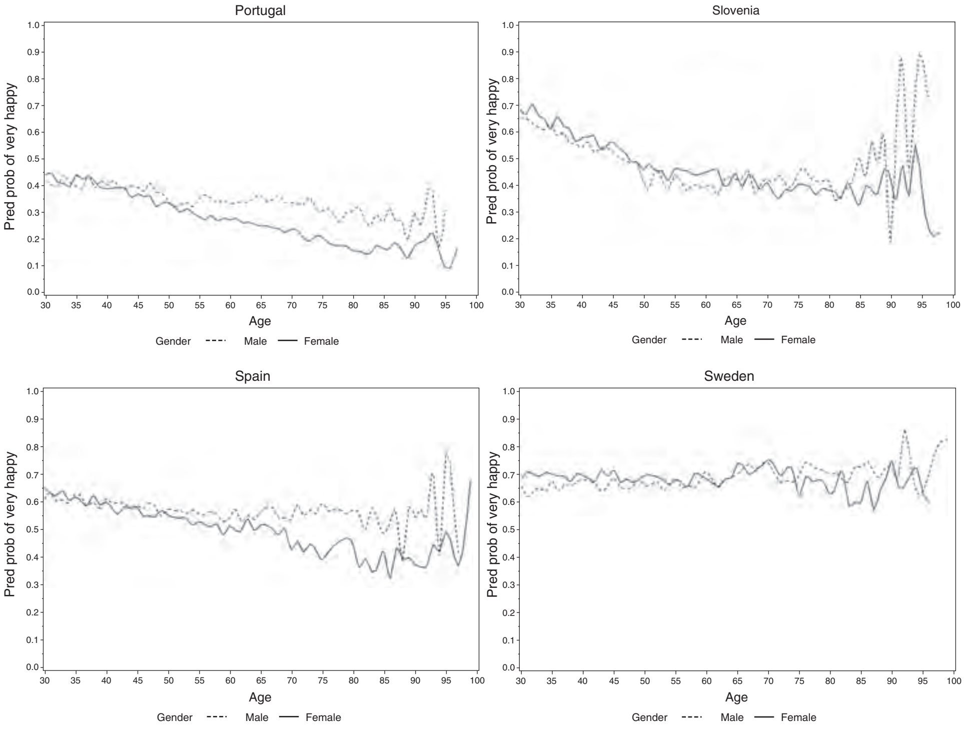


Figure A.1:
(b) Predicted age variations in place of origin effects on happiness based on model 3 in ten countries.

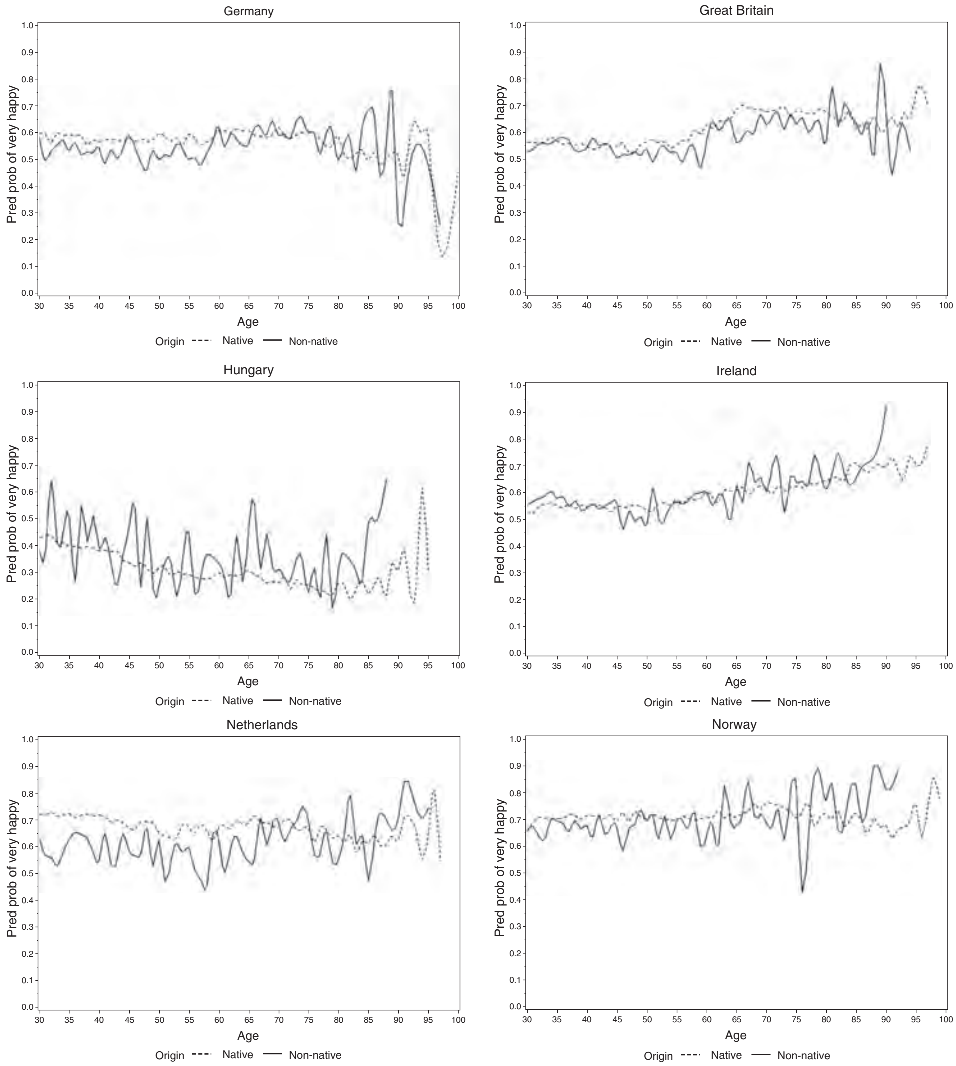


Figure A.1:
(b) Continued

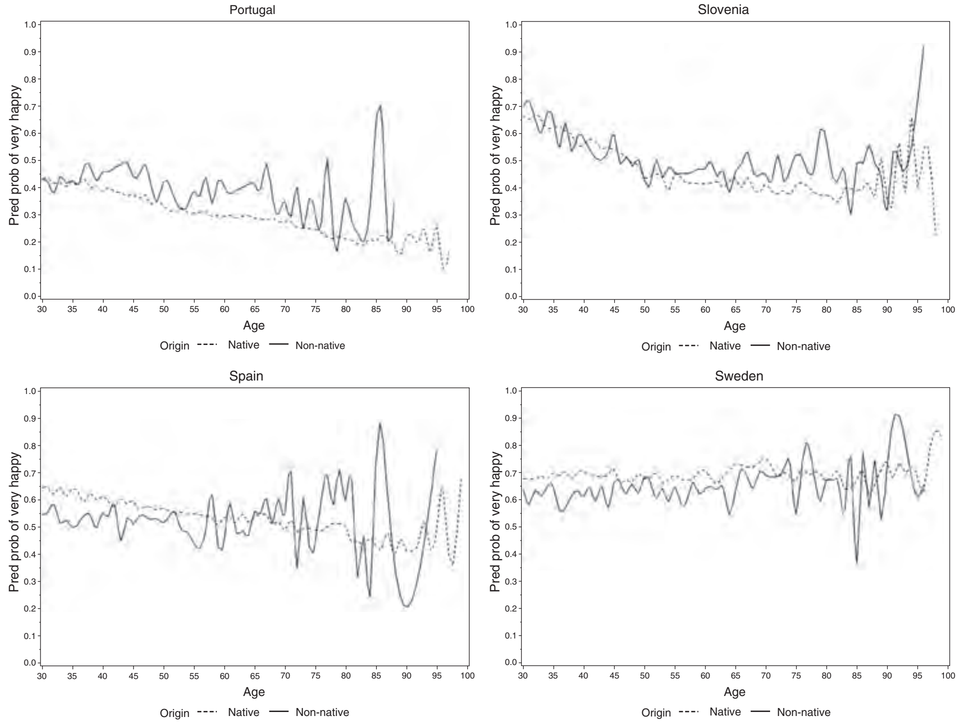


Figure A.1:
(c) Predicted age variations in low-level education effects on happiness based on model 3 in ten countries.

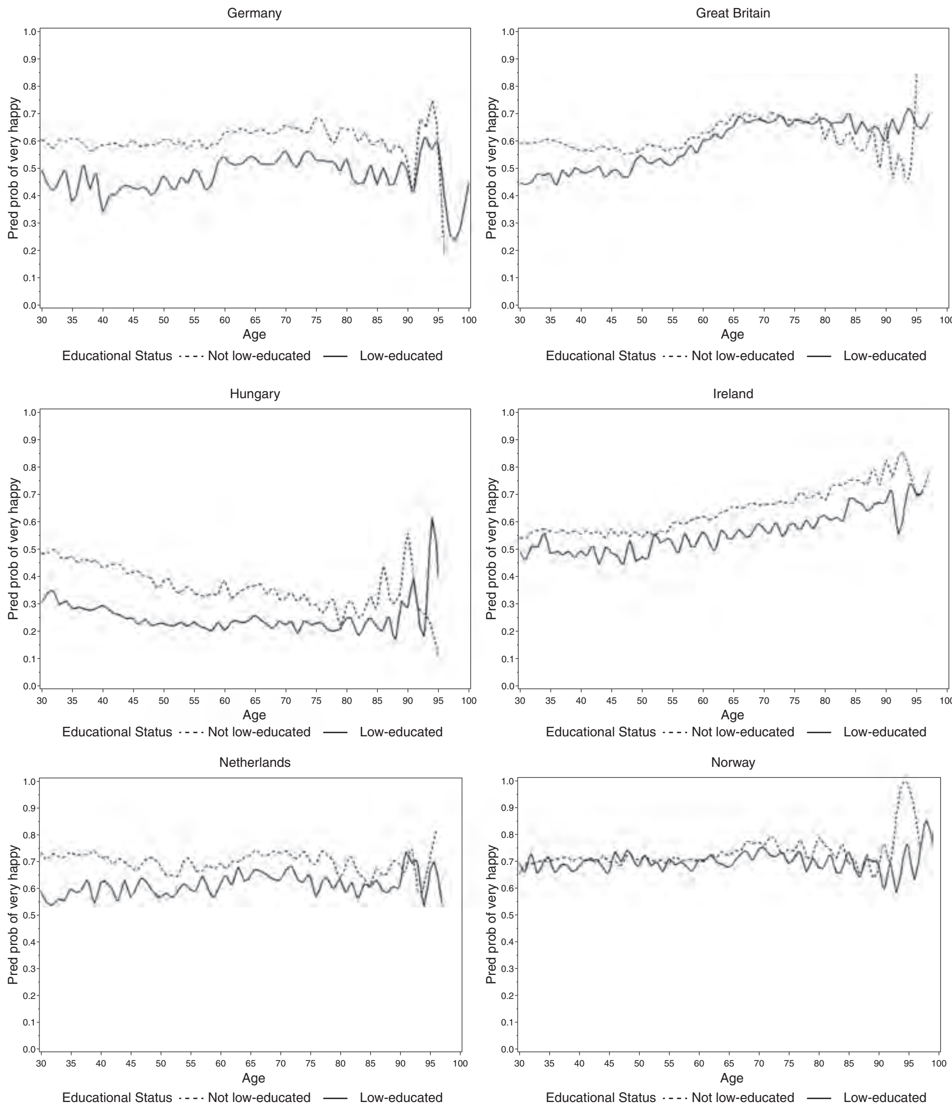


Figure A.1:
(c) Continued

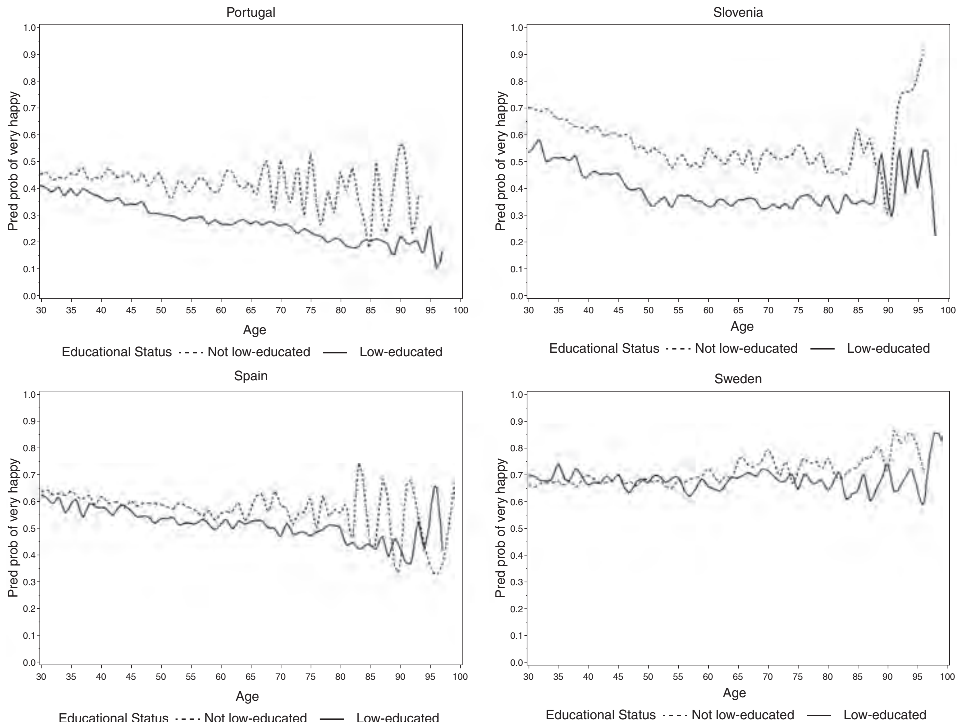


Figure A.1:
(d) Predicted age variations in high level education effects on happiness based on model 3 in ten countries.

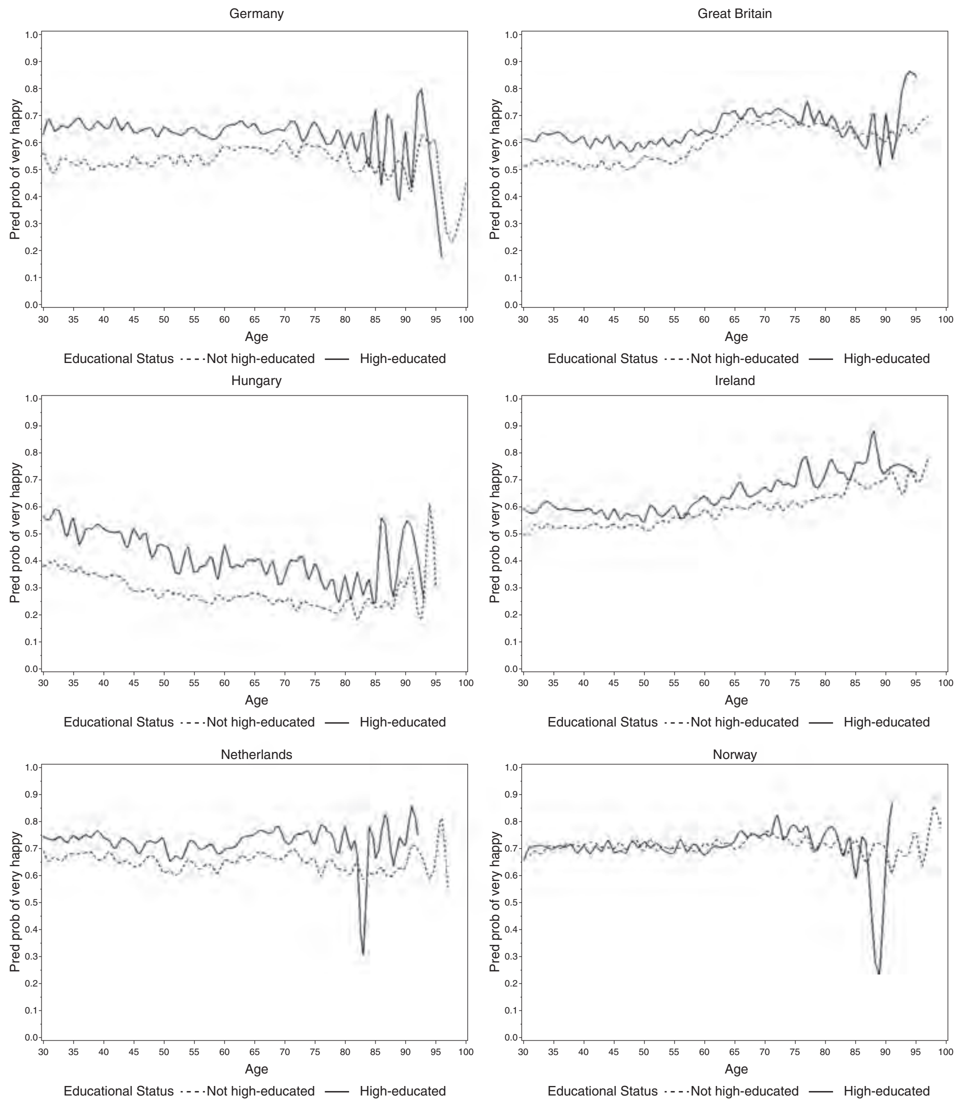
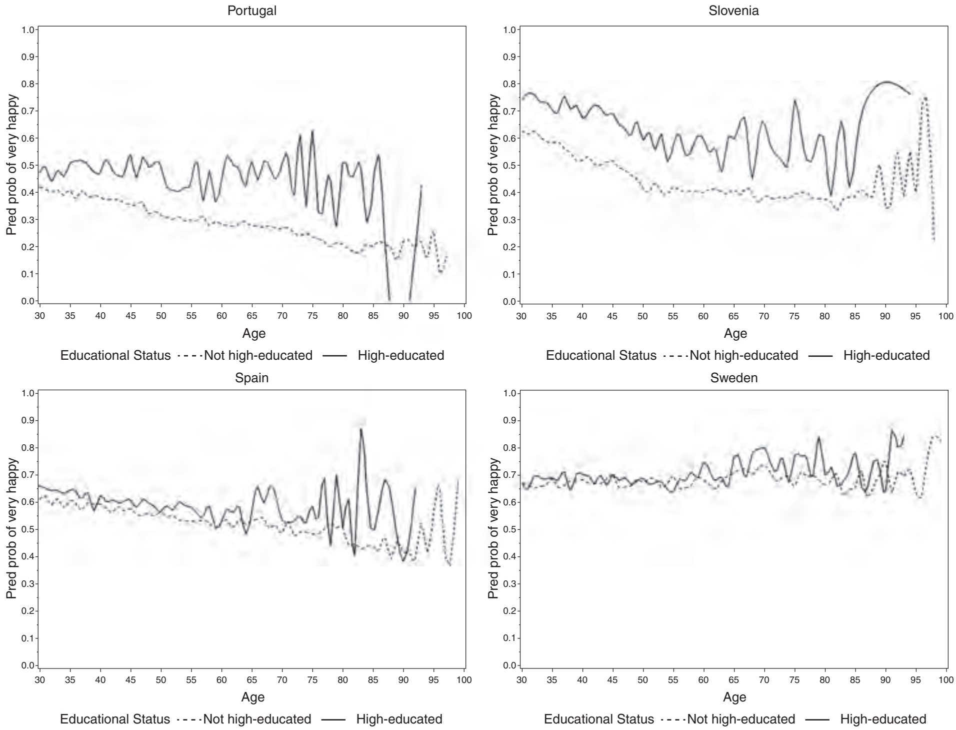


Figure A.1:
(d) Continued



Note: This model includes all independent variables and their interaction effects based on model 3.

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 467–502

Academic track mismatch and the temporal development of well-being and competences in German secondary education

Felix Bittmann[1,*]

Abstract

Formal education is one of the most influential predictors of professional success. As parents in Germany are aware of the importance of education, they often try to enable their children to enrol in the prestigious academic schooling track (*Gymnasium*). This explains why the transition recommendation made by the teacher after the fourth grade is sometimes ignored if the desired track was not recommended for a particular student. How the mismatch between the teacher's recommendation and the parents' choice of schooling for their child affects the child's development is not sufficiently known. It is very likely that such a mismatch can have consequences for the child's well-being, competences and overall academic success. Based on five consecutive panel waves of German National Educational Panel Study (NEPS) data (waves 1 to 5, collected between 2010 and 2016) (n = 2,790 in wave 1), our analyses demonstrate that social background and the probability of ignoring a teacher's recommendation are associated, and that highly educated parents are more likely to overrule the teacher's recommendation. Panel regression models show that pupils who pursued the academic track (*Gymnasium*) despite the absence of a teacher's recommendation were more likely to drop out of the academic schooling track, and were not able to catch up with their peers with respect to both objective and subjective academic competences over the entire observation window. However, the models also show that academic track mismatch did not seem to negatively influence the health and well-being of these pupils.

Keywords: German secondary education; well-being; competences; longitudinal analysis; mediation analysis; mismatch; school tracking; teacher's recommendation

[1]Leibniz Institute for Educational Trajectories (LIfBi), Department: Educational Decisions and Processes, Migration, Returns to Education, Bamberg, Germany

[*]Correspondence to: Felix Bittmann, felix.bittmann@lifbi.de

DOI: 10.1553/populationyearbook2021.res5.1

1 Introduction

Education undoubtedly plays a fundamental role in social and economic development, and the professional success and income of individuals depend to a large extent on the educational qualifications they have obtained (Triventi 2013). This relationship is also known to parents, who usually try to support and encourage their children to perform well academically. Parental involvement in school choice is particularly important in the German educational system, in which the age of first selection is early, and the children undergoing this process are hardly fully aware of the implications of choosing a secondary school (Jonkmann et al. 2010). However, in addition to the desires of the parents, a teacher's recommendation after the completion of the fourth grade of primary school also plays a vital role in school choice. Depending on the federal state, class teachers make either a binding or a non-binding school recommendation regarding the type of secondary school the student should attend. This recommendation reflects the teacher's impression of which school and learning environment is suitable for the student based on the child's previous performance and grades (Füssel et al. 2010; Lohmann and Groh-Samberg 2010). In the literature, this recommendation is considered to be extremely influential, as it can have a significant impact on the further course of the child's education (Billmann-Mahecha and Tiedemann 2006). If the teacher does not recommend enrolment in the academic track, the parents may disagree with this judgement, as they want their child to enrol in the most prestigious track, which leads directly to eligibility for higher education. If there is a mismatch between the teacher's recommendation and the type of school the parents choose for their child, the teacher's recommendation can often be ignored, as the binding character of this recommendation has been abolished in most federal states.[1] Thus, German parents have greater autonomy in the choice of schooling for their child today than they had in the past, but they are also assuming a greater degree of responsibility. The following question therefore arises: When the parents decide against accepting the teacher's recommendation regarding the type of school their child should attend, how is the child's future educational achievement and development affected?

While previous studies have investigated the association between track mismatch and various outcomes in secondary education, these analyses were not complete. For example, it is known that grades are associated with both health status (Rathmann et al. 2017) and well-being (Herzberg 2013). However, since these studies were based on cross-sectional analyses, they could not determine the direction of causality. All in all, it is clear that there is little existing empirical research on the relationship between educational track mismatch and its further consequences.

[1] Note that in this article, only the states where the teacher's recommendation for secondary school track choice is not binding are included in the analysis. Since there are no further entrance tests at the *Gymnasium* level, the children can then transition to the academic track even if their academic performance was poor in primary school.

Nonetheless, analyses of official data have shown that children who lack a teacher's recommendation to enrol in the academic track are less likely to stay on the academic track (retention rate). For example, it has been estimated that just under 70.2% of all pupils with an intermediate secondary track (*Realschule*) recommendation and 55.0% with a lower secondary track (*Hauptschule*) recommendation are still attending the *Gymnasium* by the seventh grade (Tiedemann and Billmann-Mahecha 2010). However, as these figures are reported two years after the start of secondary education, the long-term trajectories of these pupils are not known. Since the final school-leaving certificate (*Abitur*) is obtained after students have spent eight or nine years in secondary education, extending the window of observation would clearly provide more detailed insights into their trajectories. A second major limitation of these studies is that they examined a relatively small number of outcomes. These constraints were often due to the limited scope of the surveys the studies were based on, since carrying out longer and more detailed questionnaires would have increased survey costs. Moreover, even the surveys that collected more extensive data usually surveyed the participants only irregularly. However, the comprehensive NEPS dataset includes a much wider range of relevant outcome variables, many of which are surveyed every year. Thus, the NEPS provides a rich dataset for longitudinal analyses. In summary, the objectives of the present study are to extend the current state of research, and to perform genuine longitudinal analyses of the effects of academic track mismatch. Whereas in the past, only cross-sectional analyses or analyses with two points in time were available, we can now use newly available data to trace the complete school progression of children from grades five to nine. In addition, we can now analyse various other aspects of students' lives that have so far been insufficiently investigated, but can be regarded as valuable indicators of their development, such as well-being and health status (Ravens-Sieberer et al. 2000).

Exploiting longitudinal data drawn from the German National Education Panel Study (NEPS), this paper aims to answer the following research questions:

- How is the decision by a child's parents to overrule the teacher's recommendation, and to enrol the child in the academic track even when another track has been recommended, related to the social origin of the child?
- How do the retention rates, competences (math and German competences tested by the NEPS) and grades of children enrolled in the academic track without a teacher's recommendation develop compared to those of children with a teacher's recommendation?
- How do children who are enrolled in the academic track without a teacher's recommendation fare in terms of other aspects of life – including well-being, enjoyment of reading and general health – compared to children with a teacher's recommendation?
- Can differences in the academic achievement and well-being levels of the two groups of children be explained by the children without a teacher's recommendation having insufficient academic competences?

The rest of the article is structured as follows. After giving a brief overview of the German educational system and the theoretical background of the research, testable hypotheses are formulated. In the empirical analyses, which examine the experiences of pupils who were recently enrolled in the academic track in secondary education in federal states where the binding character of the teacher's recommendation has been abolished, longitudinal panel regressions based on data from the German National Educational Panel Study (NEPS) are performed to investigate these questions, and to test for mediation effects. The findings are then summarised and discussed. The main contribution of the following analyses is to generate a comprehensive picture of the overall development of pupils in secondary education by extending the window of observation using five waves of panel data, while taking into account a large set of relevant outcomes. The inclusion in the analyses of a large set of relevant control variables further strengthens the robustness of the results, and reduces the probability of spurious findings.

2 Background and theoretical framework

This section describes the German educational system with a focus on school and track differentiation. It also discusses the conceptual framework underlying our research hypotheses.

2.1 The German educational system

Here, we provide an overview of the German educational system (primary and secondary education). In grades 1 to 4 (which span an average age range of 6–7 to 10–11 years), pupils attend primary school. Based on their performance in primary school, pupils are then sorted into one of the four qualitatively different tracks (Eckhardt 2017). For each pupil, the grade four teacher makes a recommendation, which is either binding or non-binding depending on the federal state, regarding which secondary education track the pupil should be enrolled in based on his or her general academic achievement and development potential. The five-year lower secondary school track (*Hauptschule*), which prepares pupils for low-skilled service or manual work, is the least demanding. The six-year intermediate track (*Realschule*) provides pupils with vocational training in skilled white-collar or service occupations. The nine-year upper secondary school track (*Gymnasium*) prepares pupils to obtain the qualifications needed to enter university (*Abitur*), and is regarded as the most prestigious. Furthermore, most federal states have introduced comprehensive schools (*Gesamtschulen*) in addition to the tracked tripartite school system. These comprehensive schools can be seen as another approach to making the secondary school system more permeable. While comprehensive schools usually offer all three of the school-leaving certificates issued by the traditional tripartite school system, students can continue from one level to the next level given adequate performance. Thus, a comprehensive school can combine several tracks within one institution, and enable students to move between the tracks. In general, the hierarchy of the schools' performance requirements is as follows, from lowest to highest: *Hauptschule*, *Realschule*, *Gesamtschule*, *Gymnasium*.

After the final year of elementary school (grade four), the class teacher gives a track recommendation for each pupil that is based on the teacher's overall impression of the pupil's abilities and competences, including the child's social and soft skills and potential for further development.

2.2 Teacher recommendations, mismatch and social origin

As was explained above, even when a child is a low performer, the parents may be motivated to overrule the teacher's recommendation and to enrol the child in the academic track; i.e., in the school track that is associated with the best outcomes and the highest wage premiums. It has always been possible for parents to choose to enrol their child in a *less* demanding track (for example, the teacher recommends the academic track, but the parents select the intermediate track instead). Recently, parents have been given the option to overrule the teacher's decision and select a *more* demanding track for their child (for example, the teacher recommends the intermediate track, but the parents choose the academic track instead). In our analyses, we refer to a situation in which a child was enrolled in the academic track (*Gymnasium*) in grade five, and had received a teacher's recommendation to do so, as a *match*. Conversely, we refer to a situation in which a child was enrolled in the academic track, but had received a teacher's recommendation to enrol in a less demanding track in grade four (i.e., the parents overruled the recommendation), as a *mismatch*. Children who were not enrolled in the academic track in grade five are not of interest for our study, and are not considered in the theoretical expectations or analyses.

In order to clarify which social classes tend to accept a mismatch condition, the concept of primary and secondary effects of Boudon (1974) can be applied. Boudon argued that educational inequality is caused primarily by two factors. First, he noted, socially disadvantaged children tend to have lower academic performance than other children, including lower grades and standardised achievement test results (primary effects). Boudon attributed this lower performance to socially disadvantaged families being less able to invest in their children's education by, for example, providing them with individual early support or tutoring. He also identified secondary effects related to educational *decisions*. Boudon argued that compared to their more advantaged counterparts, children from disadvantaged backgrounds are less likely to achieve high educational qualifications even when their academic performance levels are the same, and are therefore more likely to choose the lower schooling tracks. He explained this tendency by observing that socially disadvantaged families often regard attending the *Gymnasium* as a comparatively long and costly educational route with a lower probability of success. This view has been further supported by the concept of relative risk aversion, which posits that children must primarily reproduce the social status of their parents (Breen and Goldthorpe 1997; Esser 1999). Thus, children from the higher social strata are expected to at least secure eligibility to enrol in higher education, as this is seen as the only way for the parents to achieve their desired level of status; whereas children from disadvantaged families are not expected to earn more than a lower diploma (Stocké 2007). Thus, the motivation to attend the *Gymnasium* is often not educational in nature (Schulze et al. 2009; Sewell and Shah 1968). If this is the case, which children belong to the group of pupils who have to ignore the teacher's recommendation to enrol in the *Gymnasium*, rather than in a lower track? Since the recommendation should be based primarily on the child's academic performance, and not on his or her social class affiliation, these individuals are likely to be low-performing students who have not achieved the required grade point average. Based on the primary effects argument, this group should include many socially disadvantaged children, as well a few children who are more socially advantaged. Nevertheless, based on the secondary effects argument, it may be assumed that among the children in this group, the more socially advantaged children (and their parents) in particular are motivated to ignore the teacher's recommendation that the child enrol in a lower secondary track, and thus proceed to enrol in the *Gymnasium*. Furthermore, it is known that compared to their more socially advantaged counterparts, socially disadvantaged families tend to be less informed about the education system, and to rely more on teacher recommendations (Harazd and van Ophuysen 2008). Based on this argumentation, the following research hypothesis can be formulated: parents who overrule the teacher's recommendation that their child enrol in a less demanding track by sending their child to the *Gymnasium* tend to belong to the higher social classes (**Hypothesis 1**).

This expectation is supported by previous research. For example, it is known that in Germany, social origin influences the probability of transitioning to the *Gymnasium*, even when controlling for the actual performance (Ditton and Krüsken 2006; Stubbe and Bos 2008); and that children from socially disadvantaged families are less likely to receive a recommendation to attend the *Gymnasium*, even if their performance is equal to that of their more socially advantaged counterparts (Dumont et al. 2014; Maaz et al. 2008). These effects are also stronger in federal states without a binding teacher recommendation (Gresch et al. 2010), which means that in these states, the parents often overrule the teacher's recommendation. Overall, the secondary effects appear to be responsible for almost 60% of the educational inequality that arises in the transition to upper secondary school (Neugebauer 2010). Other studies have also concluded that parents from the higher social strata are especially like to ignore undesirable recommendations, whereas parents from the lower social strata are more likely to accept them (Ditton et al. 2005; Harazd 2008). This discrepancy can also be attributed to evidence indicating that social position and educational aspirations are positively correlated. Therefore, it may be assumed that socially disadvantaged families are less likely to want their children to attend the *Gymnasium* in the first place (Haunberger and Teubner 2008). In order to go beyond the results of these previous studies, the following analyses will extend the window of observation by several years, as longer running panel data have recently become available. In contrast to either cross-sectional studies or panel analyses with only two waves, including a large number of survey waves allows us to trace the development of children over a longer period of time, which has several advantages. First, the problem of outliers and errors is reduced since the total number of data points is larger. Second, it makes it easier to recognise general trends over time (like up- or downward trends), and thus to draw more robust inferences. For example, it is possible that pupils with a mismatch condition have abilities that are comparable to those of pupils without a mismatch condition at the beginning of secondary education, but that the gap in performance between these two groups widens as the pupils progress. These diverging trends can be studied in detail only when several time points are included. Thus, our study has an advantage over previous studies, which often drew conclusions from a smaller number of points in time. Clearly, the more data that are available, the more precisely we can investigate the general development of pupils.

2.3 Well-being and academic performance

In light of the considerations outlined above, the following question arises: How does a mismatch condition – that is, enrolling in the academic schooling track without a teacher's recommendation – affect a child's subsequent educational achievement and success? Does a child with such a mismatch have lower chances of successfully completing the academic track? How do indicators of a child's success in life, such as his or her well-being and academic performance, as well as other relevant factors, such as the enjoyment of reading, develop over time? First, there is evidence that from the outset, children with a mismatch have lower academic performance than their classmates in the academic track (Klicpera et al. 1993; Stern 2008). This seems logical, as previous differences in the academic performance of these two groups is the main reason for the mismatch. It can, therefore, be assumed that these children do not meet the performance standards of the *Gymnasium*, or at least have below-average performance. If these children are unable to overcome their academic shortcomings, they may have to repeat the grade or even transfer to another type of school. Ultimately, these pupils are generally found at the lower end of the intra-class performance distribution, which is communicated to them by, among other things, the grades they achieve (Pfost et al. 2018). These observations can be further explained by the big-fish-little-pond effect. It is well known that for students, their own class is the most important reference group, and the comparison with their classmates has a particularly large impact on their academic self-concepts (Möller and Trautwein 2015; Zeinz and Köller 2006). Pupils whose performance is poor relative to that of the reference group are, therefore, negatively influenced (Köller 2004). Based on these findings, we can expect to observe that students who attend the *Gymnasium* without a teacher's recommendation will, on average, have worse academic outcomes than their classmates (**Hypothesis 2**). This working hypothesis is supported by previous findings, and serves as a starting point for the following analyses (Klicpera et al. 1993; Stern 2008). Furthermore, it can be assumed that the negative effects of a mismatch are ultimately caused by inadequate school performance, since performance is regularly tested and communicated by grades (**Hypothesis 3**). A pupil with unsatisfactory results will find herself at the lower end of the performance distribution, and will also face the problem of not being able to advance to the next school year, which is an objective measurement of failure that can cause the pupil to experience stress, to develop negative attitudes towards schooling, or to develop low self-esteem.

3 Empirical analyses

3.1 Data and sample

The empirical analyses we use to test our hypotheses are based on data from the German National Education Panel Study (NEPS) (Blossfeld et al. 2011).[2] The NEPS has been collecting longitudinal individual data (panel data) in a multicohort sequence design since 2008. For the present study, we use the starting cohort 3 (data version 8-0-0), which includes secondary school students who have been surveyed annually since the fifth grade; i.e., the first grade of secondary school. The target population for the starting cohort 3 consists of all children in Germany who were enrolled in fifth grade in Germany in the 2010/2011 school year. At the time the analyses were performed, information was available for grades five to nine; i.e., a total of five survey waves were conducted between 2010 and 2016. These prospectively collected individual data cover a range of relevant dimensions, such as competence development, grades, and health; but also socio-demographic variables, such as social origin, as measured by the educational degrees of the parents and their ISEI classification. Thus, the NEPS data are well-suited for the purposes of the present study. The teachers' recommendations in the fourth grade are also available.

The total sample consists of 5,753 students enrolled in the fifth grade of secondary school. This sample is additionally restricted. First, pupils from federal states that had a binding teacher recommendation at the time of the first selection (the transition to secondary education) were removed from the sample, because in such cases, the teacher's decision could not be ignored, and a mismatch was, therefore,

[2] This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Grade 5, https://doi.org/10.5157/NEPS:SC3:8.0.0. From 2008 to 2013, NEPS data were collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, the NEPS was carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

not possible. These federal states were Bavaria, Baden–Württemberg, Bremen, Saxony and Thuringia (i.e., five out of the 16 federal states, which underlines the general trend in Germany towards abolishing the binding character of the teacher's recommendation). This leaves 3,345 pupils. In addition, the states of Berlin and Brandenburg are excluded because in these states, the first selection takes place in the *sixth* grade. Children who attended a special needs school (*Förderschule*) are also excluded, since it can be assumed that our theoretical concepts do not fully apply to them. This leaves a sample of 2,790 children for analyses in wave 1 (raw sample).

3.2 Operationalisation

As explained above, theoretically, students at both the academic track schools (*Gymnasium*) and the comprehensive schools (*Gesamtschulen*) can earn the qualifications needed to enter higher education (*Abitur*). In the data, it is not always possible to distinguish which types of comprehensive schools provide this option, and which do not. Therefore, all analyses were carried out twice. In the first definition, only the academic track schools provide this option; while in the second version, both academic track and comprehensive schools are analysed together, and are considered to be equivalent. By using this approach, very precise effects can be estimated. Note that the second form of operationalisation always contains a larger number of cases. The variable of interest, ***mismatch***, is binary, and is operationalised as follows: if a child in grade five attends either the *Gymnasium* (or a comprehensive school, depending on the operationalisation) despite not having a teacher's recommendation to enrol in this track, this is classified as a *mismatch* (coded 1); whereas if a child attends the *Gymnasium* after having received the corresponding recommendation, there is no mismatch (coded 0).

The **social background** is operationalised in accordance with Boudon's theory of the two effects via the **highest school-leaving qualifications the parents** achieved, as doing so enables us to directly measure whether a particular school type can reproduce the parents' school-leaving qualifications. While there are other options available for operationalising social background (e.g., Erikson–Goldthorpe–Portocarero class scheme (EGP) or International Socio-Economic Index of Occupational Status (ISEI)), in the context of the German system – in which the educational degrees students earn are decisive for their subsequent life course, since these qualifications have a gatekeeping function that allow students to enrol in specific educational tracks – using these qualifications seems most appropriate, as each degree has a clear and inherent meaning. If information on the school-leaving qualifications achieved by both parents is available, the higher of the two degrees is used (principle of dominance). To create categories of approximately equal size, they are operationalised as follows: no degree, lower (*Hauptschulabschluss*) or intermediate degree (*Mittlere Reife*) (1), any higher education eligibility (*Fachabitur* or *Abitur*), (2) and any higher education degree (3). For the sake of simplicity, social background and parental educational qualifications are sometimes considered equivalent in the further analyses, even if the two concepts are not identical. **Parental social status** is measured by the highest ISEI (International Socio-Economic Index of Occupational Status) level, which includes additional information on the occupational status of the parents. These variables are included as control variables in the longitudinal models.

The **objective academic competence** or performance of students is measured by **standardised competence tests**, which the NEPS conducts in grades five, seven and nine. The tests cover competences in mathematics, reading skills and orthography. Reading competence is tested via 32 items in a partial credit model (Pohl et al. 2012). Mathematics competence is calculated from 24 items, which include aspects such as counting, spatial thinking, form and probability (Duchhardt and Gerdes 2012). As these different areas of competence are highly correlated, for the sake of simplicity, a single performance score is generated, which has a high Cronbach's Alpha (an Alpha in each wave greater than 0.86). Since the competences are only measured three times over the five ways, they are imputed in the two missing waves by averaging the two adjacent measurement points, if both were available. Further information on the competence measures of the NEPS can be found in the data manual (Fuß et al. 2019). As a second performance indicator, we use the **school grades** achieved in the subjects of mathematics and German from the annual reports. These grades are, in turn, combined into a single score by means of simple averaging. The German grade system ranges from one (best grade) to six (worst grade), so lower values indicate a higher performance. As both grades and objective competences were z-standardised by wave, the results can be interpreted in standard deviations. This approach also removes any growth and time effects from the data, which is beneficial for the following analyses.

Well-being is measured by general life satisfaction, and is generated from six Likert-scaled individual items (Alpha always greater than 0.76), which have already been tested in previous studies (TNS Infratest Sozialforschung 2009).[3] This variable will be referred to as general well-being, and is available in all five waves (von Collani and Herzberg 2003). The **subjectively perceived health status** of each pupil is measured on a scale of zero ("very poor") to four ("very good"), and is taken from the pupil questionnaires. As another indicator of general well-being and health, each pupil's **self-reported days of absence from school** in the past four weeks are selected. Values above the 99th percentile (corresponding to more than 15 days of absence) have been removed from this item to avoid distortions due to outliers. We assume that pupils with health problems will have a higher number of days absent from school. The **enjoyment of reading** is measured using six different Likert-scaled items (Möller and Bonerad 2007), and indicates how much a child

[3] The items are as follows: "How satisfied are you currently, all in all, with your *?" The asterisk indicates the following areas of life: "life" (1), "standard of living" (2), "health" (3), "family life" (4), "friends and acquaintances" (5), "situation at school" (6).

enjoys reading (Alpha always greater than 0.85). This variable is selected to measure the child's subjective interest in reading. The **age of the children** is measured in years in wave 1. The absolute age of the child is used to capture his or her overall cognitive development, which might be confounding if the parents sent the child to school earlier or later than the other children.

3.3 Strategy of analysis

It is important to note that in the analyses, a hierarchical data structure exists, as up to five data points are available for one person. In the longitudinal analyses, this clustering must be taken into account in order to estimate standard errors correctly. In addition, missing information is imputed to the extent possible in order to increase the effective number of cases. Multiple imputation with chained equations (MICE) is utilised. In each case, 40 imputations are generated after a burn-in of 60 iterations (Allison 2001). Common diagnostic criteria for imputations, such as convergence, were examined and approved. Some statistics about the total number of imputed values and convergence are presented in the appendix (Table A.1 and Figures A.1 and A.2). It should be noted that analyses that do not make use of any imputation procedures report basically the same results and conclusions. Thus, we are confident that the imputation does not disturb or bias the analyses in any way.

First, purely descriptive statistics are reported for all of the relevant variables in order to get a general impression of the distribution of the variables. Then, a first cross-sectional model is used to test how social origin and mismatch are related. The sample includes all children in grade five who did *not* receive an academic track recommendation (n = 1008), and may, therefore, have ignored it. This enables us to test which of these children nevertheless enrolled in the academic track. The dependent variable is binary coded (academic track/comprehensive school = 1, other school type = 0). The key explanatory variable is social origin, measured by the highest parental educational certificate with three levels.

In all of the subsequent analyses, only those children who actually enrolled in the *Gymnasium* in the fifth grade are included (n = 862 in schooling year five). This enables us to test what effects a mismatch had on a child's outcomes. Multilevel growth curve models are utilised to take the data structure into account. The dependent variables are the probability of remaining in the academic track (retention rate), competences and grades, well-being, self-assessed health, days absent from school in the last four weeks, and the enjoyment of reading. Depending on the scaling of the dependent variable, different statistical models are used, which are described in more detail below. All graphs contain 95% confidence bars to test for significant differences between groups. Unlike most other growth curve models, we use a different parametrisation, since the number of waves was low, and all of the pupils were interviewed at virtually the same point in time. This approach allows for a very flexible estimation of effects, and does not require us to make assumptions about the functional form of the temporal developments of outcomes.

In the final mediation analysis, we examine whether the group differences in outcomes between the pupils with and without a mismatch can be explained by their competences or grades. Longitudinal panel regressions with random effects are utilised. By employing a nested model design, the strength of the mediation can be assessed.

Since spurious correlations are always possible in studies using observational data, these should be mitigated or eliminated by including relevant control variables. The following variables were selected as potential confounders: the gender of the child, the child's migration background (both parents born in Germany/one parent born abroad/both parents born abroad),[4] the highest parental educational level, the highest parental ISEI level, the age of the child, whether the parents were living together (including whether the parents were divorced, widowed or single) and the federal state where the school was located. It should be noted that due to the data protection regulations for the NEPS, we are not allowed to publish regression coefficients that were computed for the federal states; thus, these coefficients are not included in the output tables.

All calculations are computed in Stata 16.1, using, among other programs, the user program *mimrgns*, which is needed to calculate the average margin effects for the imputations (Klein 2014).

4 Findings

First, the descriptive statistics are briefly summarised. All results are calculated for pupils who were attending the *Gymnasium* in grade five in the first wave of the survey (Table 1). Continuous variables are presented with the mean and standard deviation. Categorical variables are presented dichotomised in the table.

Of all the pupils in the sample, 49.3% were enrolled in the academic track when the strict operationalisation of "academic" is utilised, while this share increases to about 65% when comprehensive schools are also considered "academic" based on the alternative definition. The data clearly indicate that only about 12% to 18% of all pupils enrolled in the *Gymnasium* displayed a mismatch, depending on the form of operationalisation used. The large share of these parents who had a very high level of education is also striking, as more than 50% had higher education eligibility (*Abitur*). This result is approximately in line with data from official statistics. In 2016, 63.4% of parents with a child in the academic track had higher education eligibility (*Statistisches Jahrbuch Deutschland 2016* 2016: 81).

[4] It is known that immigrants often have high aspirations for their children, which could make them prone to ignore the teacher's recommendation. Thus, migration background might be a confounding factor (Kao and Tienda 1995).

Table 1:
Descriptive statistics for wave 1

	Academic track only				Academic track and comprehensive schools			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Attending academic track	49.3%				64.9%			
Mismatch	12.3%				18.3%			
Female	50.2%				50.2%			
Age in Wave 1	11.191	0.464	9.7	12.9	11.192	0.465	9.6	13.1
Life satisfaction*	8.607	1.629	0.0	10.0	8.608	1.628	0.0	10.0
Enjoy reading*	3.122	0.779	1.0	4.0	3.122	0.779	1.0	4.0
Health*	4.290	0.753	1.0	5.0	4.290	0.753	1.0	5.0
Days absent*	0.964	1.954	0.0	15.0	0.967	1.951	0.0	15.0
Grades (Std.)*	0.000	1.000	−1.7	4.8	0.000	1.000	−1.7	4.8
Competences	0.060	0.807	−2.6	2.6	0.060	0.807	−2.6	2.6
Competences (Std.)*	0.000	1.000	−3.3	3.2	0.000	1.000	−3.3	3.2
Not intact family	22.7%				22.7%			
Both parents born in Germany	78.5%				78.5%			
One parent born abroad	11.1%				11.1%			
Both parents born abroad	10.4%				10.4%			
Max. Intermediate degree	42.1%				42.0%			
Higher education eligibility	25.0%				25.0%			
Tertiary degree	32.9%				33.0%			
Parental ISEI	50.053	18.156	11.0	86.0	50.041	18.157	11.0	86.0

Source: NEPS SC3, wave 1. Imputed data.
Note: Variables marked with an asterisk are time-varying, and can, therefore, change in the subsequent waves.

4.1 Social origin and mismatch

To test how social origin was associated with a mismatch, binary logistic regressions are performed (Table 2). For each operationalisation, two models are computed, one without controls and one with all control variables added. Unlike all of the following analyses, these analyses include children enrolled in the fifth grade in *all* types of schools who did *not* receive a teacher's recommendation to enrol in the academic track. Average marginal effects (AMEs) are reported. It should be noted that parental social status, measured via the ISEI level, is not included in these models, as this variable is strongly correlated with parental education, and its inclusion would undermine the effect of the education variable.

In the models without the controls added, we find a positive and statistically significant effect. For example, the results show that compared to parents with lower levels of education, parents with a university degree whose children were enrolled in the academic track (strict operationalisation) were 8.3 percentage points more likely to enrol their child in the academic track despite not having received a teacher's recommendation to do so. The results for the alternative definition of academic track that includes comprehensive school are similar. However, our conclusions change as soon as controls are added. In the strict operationalisation, a big part of the effect vanishes and the significance is lost, while the trend is still the same. When the alternative definition of the academic track that includes comprehensive schools is applied, the effect becomes smaller, but is still significant. Taken together, we conclude that the effect is reduced somewhat after the controls are added; that there is still a clear trend in both forms of operationalisation; and that highly educated parents are, on average, more prone to enrol their child in the academic track despite not having received a recommendation to do so.

4.2 Longitudinal trajectories

In the following analyses, mismatch status is the central explanatory variable. Multilevel growth curve models are calculated in which an interaction between the wave dummy and the mismatch variable is inserted (Rabe-Hesketh and Skrondal 2012). The educational trajectories can thus be modelled in a very flexible way, and allow for a clear graphical representation. First, we investigate whether students with a mismatch are more likely to drop out of the academic track than students who received a teacher's recommendation to enrol (retention rate). For this purpose, a binary variable that indicates whether or not a student was enrolled in the academic track in a given wave is created. By design, this share was 100% in grade five. The entire trajectories are depicted in Figure 1. Each model includes the previously introduced control variables. Complete regression tables are available in the appendix (Table A.2).

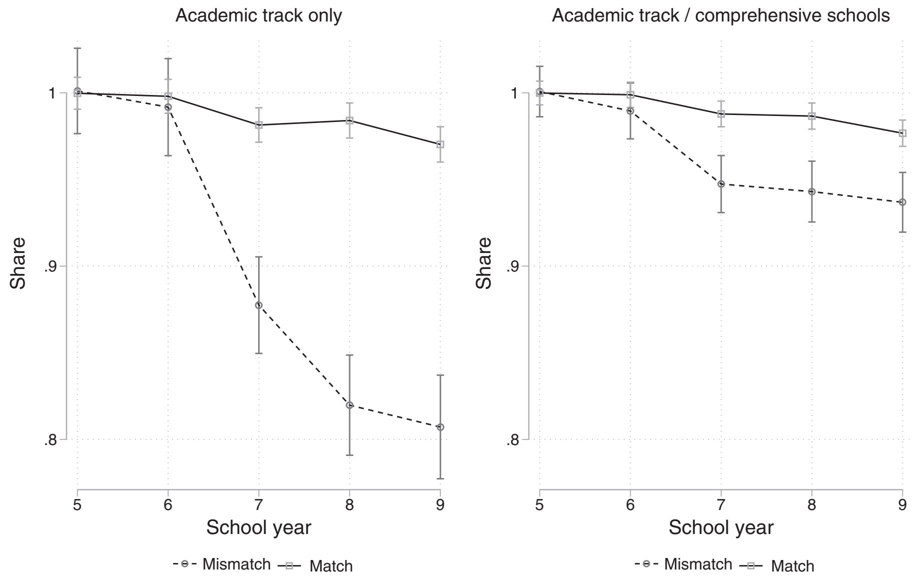
The trend is quite clear. Pupils with a mismatch tended to drop out more often than the pupils who received the recommendation. In the strict operationalisation,

Table 2:
Logistic regression results of enrolling in the academic track without a teacher's recommendation (AMEs)

	Academic track only		Academic track and comprehensive schools	
	Base model	**Controls added**	**Base model**	**Controls added**
Female		–0.026		0.004
		(0.023)		(0.015)
Age (Wave 1)		-0.076^{*}		-0.061^{**}
		(0.031)		(0.019)
Parents' highest educational qualification				
Max. intermediate degree	Ref.	Ref.	Ref.	Ref.
Higher education eligibility	0.119^{***}	0.077^{*}	0.075^{**}	0.063^{**}
	(0.034)	(0.032)	(0.025)	(0.022)
Tertiary degree	0.083^{**}	0.047	0.118^{***}	0.096^{***}
	(0.032)	(0.030)	(0.021)	(0.020)
Both parents born in Germany		Ref.		Ref.
One parent born abroad		–0.038		0.017
		(0.041)		(0.020)
Both parents born abroad		0.017		–0.036
		(0.048)		(0.026)
Parents not living together		-0.081^{*}		-0.042^{*}
		(0.033)		(0.021)
Federal State	No	Yes	No	Yes
Observations	1008	1008	1073	1073

Source: NEPS SC3, imputed data.
Note: Standard errors in parentheses. Only children who did not receive a teacher's recommendation to enrol in the academic track are included. Dependent variable: transition to the academic track (1) or not (0). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

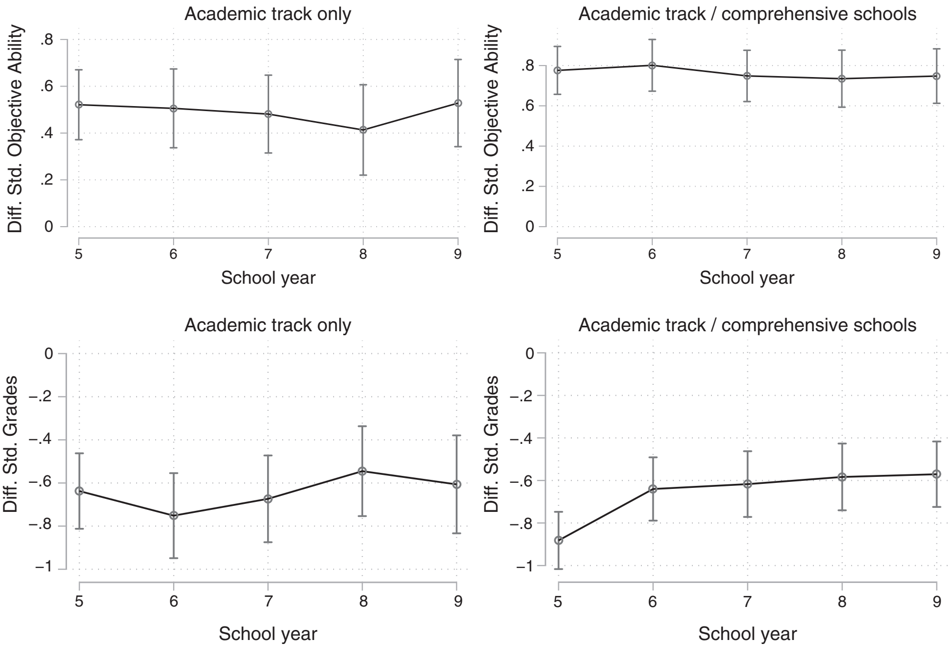
Figure 1:
Retention rates by mismatch status and form of operationalization



this effect is huge, and is more than 15 percentage points in wave nine. When the alternative definition of academic track that includes comprehensive schools is applied, the effect is smaller (more than five percentage points), but is still statistically significant. Additional tests for the difference between the two lines indicate that the difference is actually statistically significant on the 5% level from grade seven onwards (not depicted). These results are in line with previous findings (Tiedemann and Billmann-Mahecha 2010).

In the next model, the trajectories of objective competences and school grades are analysed. Linear models are estimated, since the outcomes are measured continuously and are approximately normally distributed. The models include the same control variables as before to account for spurious correlations. To allow for a clearer interpretation, the model shows not the predicted outcomes, but the differences between the two groups (match/mismatch). This means that for the objective competences, a positive number represents an advantage for the pupils with a match, while for the grades, a negative number indicates an advantage for these pupils (since lower grades indicate better performance in the German system, with one being the best grade and six being the worst). This interpretation makes it convenient to test for statistical significance. As long as the zero line is not touched by the confidence bars, the difference is significant on the 5% level. The results are

Figure 2:
Differences in competences by type of competence and form of operationalization



Note: The *differences* between the mismatch and the match status groups are depicted.

shown in Figure 2, and the complete regression tables are presented in the appendix (Table A.3).

When we look at the results for the objective competences (upper row), we can see that the differences are always positive, and that zero is never touched by the confidence bars. This means that pupils with a match status always display higher objective competences. The gap between the two groups is rather constant over time, and lies between 0.4 and 0.5 standard deviations. The difference between the two forms of operationalisation is rather small, while the effects are stronger in the alternative definition of academic track that includes comprehensive schools. The same overall conclusions can be reached for subjective grades (lower row). While the signs are reversed due to the coding scheme, the meaning is the same. The pupils with a mismatch condition always had worse outcomes; that is, numerically higher grades.

The models for the other four outcomes are built identically. The predicted outcomes are reported for both groups (Figure 3 and 4), and complete regression tables are shown in the appendix (Table A.4). Again, the results are visualised separately for both operationalisations.

Figure 3:
Trajectories of satisfaction and enjoyment of reading by mismatch status

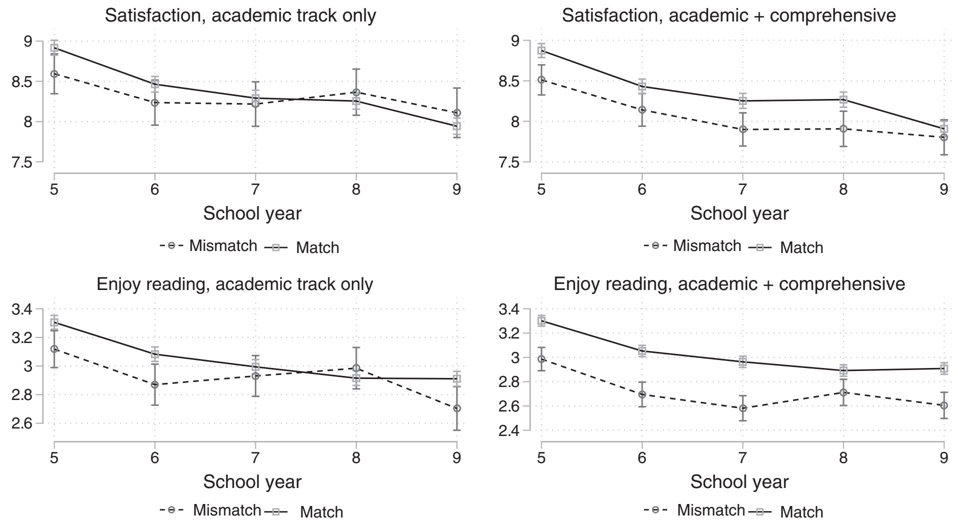
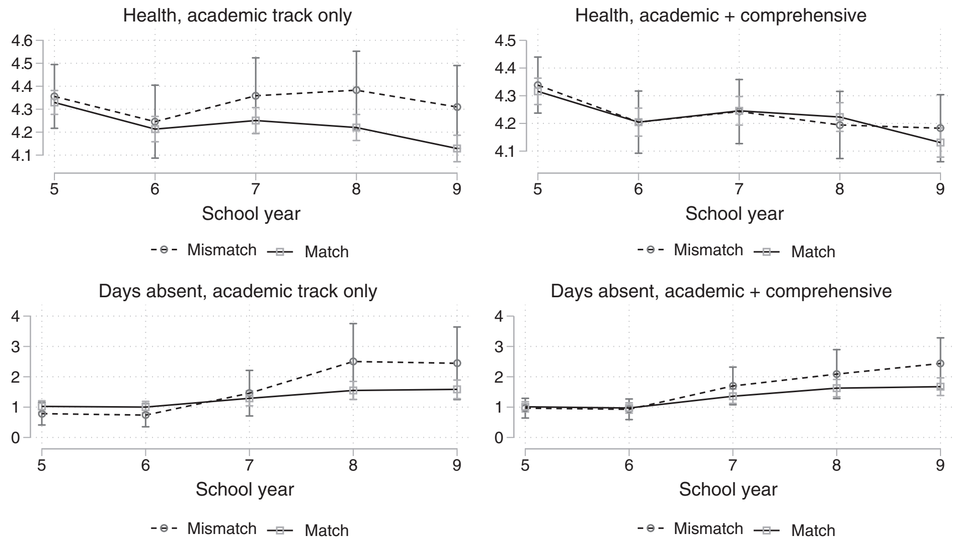


Figure 4:
Trajectories of health and days absent from school by mismatch status



As we can see, there are virtually no differences between the matched and the mismatched pupils using the strict operationalisation. For both outcomes, satisfaction and enjoyment of reading, the confidence intervals clearly overlap. The picture is quite different when the alternative definition of the academic track that includes comprehensive schools (right side of the figure) is employed: i.e., we see clear gaps for both satisfaction and enjoyment, which indicate significant group differences. Here, we observe that pupils in the mismatch condition have lower levels of satisfaction and of enjoyment of reading.

With respect to the two final outcomes, the general health of the child and the days the child was absent from school in the last four weeks, we do not see any group differences at all, regardless of the form of operationalisation. These findings lead us to conclude that there is no association between the mismatch status and these two outcomes after controlling for various potentially confounding variables, which are the same as those used in the previous models.

4.3 Mediation analysis

Since the previous analyses indicate that there are significant group differences in levels of satisfaction and of enjoyment of reading, the question of whether these differences can be attributed to differences in performance, which are clearly present, arises. To investigate this question, mediation models can be applied. Again, both forms of operationalisation are used for the outcomes of satisfaction and enjoyment of reading. The models are built as follows. The first model includes the treatment (mismatch) and all controls. The second model adds both variables of performance (objective performance and grades). If the coefficients of the treatment status change, a mediation effect is present. To test this statistically, the Sobel test is applied to each model (Hayes 2018). If the coefficients are found to be statistically significant, this proves that a mediation effect is present. Results are presented in Table 3.

When interpreting the results from the strict operationalisation, we notice that the models also indicate that there is no effect of mismatch status in these findings. However, we see that the sign of the coefficient changes after the mediators are included. The mediator variables are highly significant (for satisfaction, only grades; for enjoyment, both grades and objective performance). Applying the Sobel test revealed that there is indeed a mediation effect, and that the coefficients are highly significant. However, since the baseline effect of mismatch is not significant, the total share mediated is probably very small, and these findings should, therefore, be interpreted with caution. The outcomes are clearly different when the alternative definition of academic track that includes comprehensive schools is used. Here, the mismatch variable is highly significant and negative, which suggests that there is a negative effect for pupils with a mismatch. This was also clear from the previous analyses. When the mediators are added, the coefficient of mismatch is no longer significant, which indicates that a mediation effect is present. The mediating variables are again highly significant. As the Sobel test results are highly significant

Table 3:
Mediation regression models

	Academic track only				Academic track and comprehensive schools			
	Satisfaction baseline	Satisfaction mediators added	Enjoyment reading baseline	Enjoyment reading mediators added	Satisfaction baseline	Satisfaction mediators added	Enjoyment reading baseline	Enjoyment reading mediators added
Mismatch	−0.076	0.140	−0.071	0.075	−0.199*	0.042	−0.226***	−0.026
	(0.120)	(0.127)	(0.067)	(0.068)	(0.092)	(0.098)	(0.053)	(0.054)
Female	0.190**	0.133	0.294***	0.255***	0.141*	0.110	0.263***	0.243***
	(0.073)	(0.071)	(0.039)	(0.037)	(0.065)	(0.063)	(0.035)	(0.034)
Both parents German	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
One parent born abroad	−0.000	0.054	0.053	0.108	0.048	0.098	0.044	0.093
	(0.130)	(0.126)	(0.066)	(0.064)	(0.112)	(0.109)	(0.057)	(0.056)
Both parents born abroad	0.029	0.116	0.058	0.146*	0.156	0.249	−0.045	0.053
	(0.158)	(0.153)	(0.061)	(0.062)	(0.131)	(0.131)	(0.061)	(0.060)
Max. intermediate degree	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Higher education eligibility	0.134	0.110	0.060	0.056	0.038	0.018	0.102*	0.079
	(0.100)	(0.098)	(0.052)	(0.049)	(0.090)	(0.088)	(0.046)	(0.044)
Tertiary	−0.029	−0.078	0.005	−0.007	−0.072	−0.118	0.049	0.023
	(0.117)	(0.115)	(0.062)	(0.059)	(0.100)	(0.099)	(0.054)	(0.052)
Parental ISEI	0.006	0.004	0.005**	0.003*	0.008**	0.006*	0.004**	0.002
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)

Continued

Table 3:
Continued

	Academic track only				Academic track and comprehensive schools			
	Satisfaction baseline	**Satisfaction mediators added**	**Enjoyment reading baseline**	**Enjoyment reading mediators added**	**Satisfaction baseline**	**Satisfaction mediators added**	**Enjoyment reading baseline**	**Enjoyment reading mediators added**
Age	−0.031	−0.032	−0.015	−0.003	−0.052	−0.035	−0.043	−0.014
	(0.092)	(0.091)	(0.054)	(0.052)	(0.078)	(0.077)	(0.044)	(0.042)
Not intact family	-0.257^{**}	-0.208^{*}	−0.072	−0.054	-0.358^{***}	-0.319^{***}	−0.020	−0.002
	(0.096)	(0.091)	(0.053)	(0.051)	(0.085)	(0.083)	(0.046)	(0.045)
Grades		-0.300^{***}		-0.082^{***}		-0.262^{***}		-0.059^{***}
		(0.034)		(0.017)		(0.031)		(0.015)
Academic performance		0.040		0.195^{***}		0.074		0.207^{***}
		(0.053)		(0.025)		(0.044)		(0.021)
Constant	8.528^{***}	8.569^{***}	2.793^{***}	2.656^{***}	8.800^{***}	8.594^{***}	3.201^{***}	2.906^{***}
	(1.051)	(1.038)	(0.616)	(0.588)	(0.896)	(0.888)	(0.499)	(0.477)
Observations	3396	3396	3365	3365	4345	4345	4301	4301
Sobel-test	–	-6.496^{***}	–	-5.142^{***}	–	-7.125^{***}	–	-7.895^{***}
		(0.030)		(0.018)		(0.026)		(0.020)

Source: NEPS SC3, imputed data.
Note: Standard errors in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

as well, we can conclude that a mediation effect is present, and that the performance differences between the two groups can explain why the pupils with a mismatch status had worse outcomes.

5 Discussion

We start this section by discussing the choice of the secondary education track. The models indicated that higher educated parents were more likely to deviate from a teacher's recommendation to enrol their child in a lower academic track by enrolling their child in the academic track. Since the results were stable even after controls were added, this seems to be a robust association. Even though one coefficient was no longer statistically significant, the trends remained clear. Therefore, hypothesis one is accepted. This means that the probability that a teacher's recommendation that a child be enrolled in a lower schooling track was overruled increased with the social status of the parents. The kind of operationalisation was shown to play a minor role only. This finding is in line with previous studies that also found that there is a correlation between the social origin of a pupil's parents and the choice of school track (Ditton et al. 2005). Taken together, these outcomes further confirm that highly educated parents had a strong tendency to enrol their child in the academic track, even if the child was not well-prepared for it at the time of the transition. As was discussed above, motives of status maintenance might explain this choice. By contrast, less educated parents tended to rely more on the decisions of the teacher, as they were less informed about the school system and all of the available options. These findings show that abolishing the binding character of the recommendation after elementary school appears to affect higher educated parents in particular, as their choice is decisive, and it is their responsibility to select a track for their child.

With respect to retention rates, the results were very clear. As expected, we found that the children with a mismatch status in the academic track had a significantly higher chance of dropping out and of leaving this track. This finding held for both types of operationalisations, while the results were clearly stronger when the strict operationalisation was analysed. This makes sense, since switching from the academic track to the comprehensive school was not coded as a dropout in the alternative definition of academic track that includes comprehensive schools. In short, this result is in line with our expectations, since, on average, the pupils with a mismatch status had lower performance at the start of their secondary education (this being the reason why they did not receive a recommendation to enrol in the academic track). For the academic track, it is clear that there was a selection based on performance; thus, the fact that there were dropouts was not surprising. These findings are in line with those of previous studies (Tiedemann and Billmann-Mahecha 2010). In addition, these results underline that enrolling a child in the academic track without a teacher's recommendation obviously does not guarantee that the child will have academic success or become eligible to enter higher education, even if the child joined this track at the start of her or his secondary education. Parents should be aware of these findings and the implications. Having to leave a track might disturb the social networks a child has built over time. Moreover, it might feel humiliating for a child to have to drop out due to substandard performance. Therefore, parents should consider these implications when overruling the decision of the teacher.

Next, we discuss the academic performance trajectory. The results from our statistical analysis showed that the gap between the pupils with a match and a mismatch was very large, and was almost constant over the entire five-year observation period. On the one hand, this finding was expected since the performance gap was the main reason for the mismatch status. However, it was very interesting to see that the gap persisted, and that there was almost no sign of catching-up processes. Even in the more stimulating learning environment of the academic track with peers who were performing well, the initially disadvantaged pupils were not, on average, able to demonstrate significant improvements and reach the level of their peers. This finding is highly relevant, as it addresses the expectation of many parents that being in the right learning environment could improve their child's academic performance. Since the finding was stable for objective and subjective competences and for both forms of operationalisation, we believe that is a robust conclusion.

However, the trajectories of the four other outcomes that were not related to academic performance were much less clear. In terms of health, no group differences at all were found in health or in days absent from school. This was a positive finding, as it indicated that even if there was a substantial gap in terms of a child's academic performance, the general health of the child was not associated with a mismatch. For the other two outcomes, life satisfaction and enjoyment of reading, the differences were only significant when the alternative definition of academic track that includes comprehensive schools was used. It is, however, not clear why the form of the operationalisation made a difference in this case. Since the finding was not very robust, we assume that the effects were rather weak, and should be interpreted with caution. Seen from the perspective of the children, this outcome was positive, as it indicated that a mismatch did not necessarily negatively affect the well-being of a child. The expectation that having a large academic performance gap compared to their peers could have a large negative influence on the satisfaction or health of mismatched pupils was not confirmed.

With respect to the mediation analysis, we did see that the negative effects were mediated by academic competences. We should be cautious in interpreting these findings, since the effect of the treatment variable mismatch was not stable, which suggests that the total effect was small. The Sobel test indicated that a mediation effect was probably present based on the measurements of performance, and that both subjective and objective performance could be the mediator. These findings are in line with our expectations. Since performance gaps between the two groups were the main reason for the mismatch, and these gaps persisted over time, it is not very surprising that a mediation effect could be observed.

What are the implications of these findings? Did the parents who ignored a teacher's recommendation and forced their child to enrol in the academic track act sensibly, or were they irresponsible? The findings clearly show that the expectation that pupils who had poor performance in the lower grades would catch up with their peers over time did not materialise, and that mismatched pupils had much higher chances of dropping out. For this reason, the success of the shift towards allowing parents to enrol their child in the academic track against the teacher's recommendation is highly questionable. However, if the child is successful in pursuing the academic track and does not drop out, this will enable him or her to proceed directly to higher education. Since we found that other important outcomes like health were clearly not negatively associated with a mismatch, the decision of the parents cannot be framed as irresponsible.

Finally, the limitations of the present study should be discussed. Since only observational data were available, it was not possible to estimate pure causal effects. Even if certain disturbance factors could be ruled out by including control variables, it was impossible to control for *all* conceivable confounding variables. Thus, the results shown here should be understood as associations, or as an attempt to approximate causal effects. With respect to the sample selection, only federal states that allowed parents to ignore the teacher's recommendation were included in the analyses. Therefore, all conclusions are only valid for these states. However, since the majority of states have abolished the binding character of the recommendation, it seems likely that these results will become more relevant in the future as more power is given to the parents to overrule the teacher's decision regarding tracking.

An additional limitation is that the NEPS survey is not perfectly representative of the overall population in Germany due to selection dropout and refusal to participate. The consequence is that highly educated parents are somewhat overrepresented in the sample. However, in this study, our main focus was on examining how mismatched students fared, rather than on investigating who these students were. While we cannot be sure that the percentages of mismatched student will be the same in the overall population, their trajectories can still be described. If we further assume that mismatched pupils from socially advantaged families will do better on average than mismatched pupils from disadvantaged families, we can conclude that the outcomes might be even worse (that is, for example, levels of satisfaction or performance could be even lower) in the general population, since they probably receive less support. Given that highly educated parents are overrepresented in the NEPS data, our estimation of the academic competences and well-being of mismatched students may even be biased downward. Finally, it should be emphasised that the study could not and should not evaluate the usefulness of the (non-) binding character of track recommendations. However, as the results presented here are similar to previous findings, we would argue that, all in all, our results are quite robust and trustworthy.

6 Conclusion

Based on the German NEPS student cohort, the analyses showed that students who enrolled in the academic track after the fourth grade without a teacher's recommendation had significantly lower performance levels, and that their performance disadvantage persisted over time. Therefore, the expectation of many parents that the academic performance of a student who performed poorly in the early grades would converge with the class average over time did not seem to be justified. These outcomes could be explained by the pupils' inadequate performance, which was responsible for their negative educational trajectories. However, since the findings also showed that other outcomes like satisfaction were probably not associated with a mismatch, it appears likely that no great damage was done to the mismatched children.

Overall, the findings reported in this study can serve as a foundation for parents who face the decision of whether to enrol their child in the academic track of secondary education, even if the child did not receive a teacher's recommendation to do so. While attending the *Gymnasium* is the most direct route to earning the *Abitur*, it is important to keep in mind that there may be negative effects if the fit between a child's abilities or interests and the type of track she or he is enrolled in diverge strongly. Accordingly, it seems sensible to consider alternative pathways, which are now numerous. From a research perspective, the question of how the differences between the groups of students who did and did not have a mismatch evolve even further into the future – that is, after they leave school and enter the labour market or tertiary education – appears highly relevant. Do the children with a mismatch experience long-time negative effects, or will the trajectories of the two groups converge over time? Furthermore, even if the children with a mismatch have worse outcomes that their peers *in the academic track* with a match, what are the effects when they are compared to their peers who chose a less demanding educational track based on a teacher's recommendation? Investigating these highly interesting research questions could be very rewarding for future analyses. In addition, replication studies, especially from different countries and other contexts, are very welcome.

References

Allison, P. D. 2001. *Missing data*. Vol. 136. Thousand Oaks: Sage publications.

Billmann-Mahecha, E. and J. Tiedemann 2006. Übergangsempfehlung als kritisches Lebensereignis. Migration, Übergangsempfehlung und Fähigkeitsselbstkonzept. In *Risikofaktoren kindlicher Entwicklung*, ed A. Schründer-Lenzen, 193–207. Wiesbaden: Springer. https://doi.org/10.1007/978-3-531-90075-9_10

Blossfeld, H.-P., H.-G. Roßbach and J. von Maurice 2011. *Education as a lifelong process. The German National Educational Panel Study (NEPS)*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Boudon, R. 1974. Educational growth and economic equality. *Quality and Quantity* 8: 1–10. https://doi.org/10.1007/BF00205861

Breen, R. and J. H. Goldthorpe 1997. Explaining educational differentials. Towards a formal rational action theory. *Rationality and Society* 9(3): 275–305. https://doi.org/10.1177/104346397009003002

Ditton, H. and J. Krüsken 2006. Der Übergang von der Grundschule in die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft* 9: 348–372. https://doi.org/10.1007/s11618-006-0055-7

Ditton, H., J. Krüsken and M. Schauenberg 2005. Bildungsungleichheit – der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft* 8: 285–304. https://doi.org/10.1007/s11618-005-0138-x

Duchhardt, C. and A. Gerdes 2012. NEPS technical report for mathematics. Scaling results of starting cohort 3 in fifth grade. NEPS Working Paper No. 19. Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

Dumont, H., K. Maaz, M. Neumann and M. Becker 2014. Soziale Ungleichheiten beim Übergang von der Grundschule in die Sekundarstufe I. Theorie, Forschungsstand, Interventions- und Fördermöglichkeiten. *Zeitschrift für Erziehungswissenschaft* 17: 141–165. https://doi.org/10.1007/s11618-013-0466-1

Eckhardt, T. 2017. The Education System in the Federal Republic of Germany 2015/2016. A description of the responsibilities, structures and developments in education policy for the exchange of information in Europe. Berlin: Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany. https://www.kmk.org/fileadmin/Dateien/pdf/Eurydice/Bildungswesen-engl-pdfs/dossier_en_ebook.pdf, accessed on 14 January 2021.

Esser, H. 1999. *Die Konstruktion der Gesellschaft*. Vol. 2. Frankfurt am Main: Campus-Verlag.

Fuß, D., T. Gnambs, K. Lockl and M. Attig 2019. Competence data in NEPS. Overview of measures and variable naming conventions (starting cohorts 1 to 6). NEPS Research Data Paper. Bamberg: Leibniz Institute for Educational Trajectories.

Füssel, H.-P. et al. 2010. Der institutionelle Kontext von Übergangsentscheidungen. Rechtliche Regelungen und die Schulformwahl am Ende der Grundschulzeit. In *Der Übergang von der Grundschule in die weiterführende Schule. Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten*, eds K. Maaz et al., 87–106. Bonn: Bundesministerium für Bildung und Forschung.

Gresch, C., J. Baumert and K. Maaz 2010. Empfehlungsstatus, Übergangsempfehlung und der Wechsel in die Sekundarstufe I. Bildungsentscheidungen und soziale Ungleichheit. In *Bildungsentscheidungen*, eds J. Baumert, K. Maaz and U. Trautwein, 230–256. Wiesbaden: Springer. https://doi.org/10.1007/978-3-531-92216-4_10

Harazd, B. 2008. Was kennzeichnet Eltern, die die Grundschulempfehlung nicht wahrnehmen? In *Chancenungleichheit in der Grundschule. Ursachen und Wege aus der Krise*, eds J. Ramseger and M. Wagener, 95–98. Wiesbaden: Springer. https://doi.org/10.1007/978-3-531-91108-3_12

Harazd, B. and S. van Ophuysen 2008. Was bedingt die Wahl eines nicht empfohlenen höheren Bildungsgangs? *Zeitschrift für Erziehungswissenschaft* 11: 626–647. https://doi.org/10.1007/s11618-008-0037-z

Haunberger, S. and M. Teubner 2008. Bildungswünsche von Eltern und Kindern im Vergleich. In *Kinderleben. Individuelle Entwicklungen in sozialen Kontexten*, ed C. Alt, 293–316. Wiesbaden: Springer. https://doi.org/10.1007/978-3-531-91982-9_12

Hayes, A. F. 2018. *Introduction to mediation, moderation, and conditional process analysis. A regression-based approach*. Second edition. New York: Guilford Press.

Herzberg, P. Y. 2013. Determinanten subjektiven Wohlbefindens bei Jugendlichen. In *Systemwechsel zwischen Projekt und Prozeß. Analysen zu den Umbrüchen in Ostdeutschland*. Vol. 13, eds H. Bertram, W. Kreher and I. Müller-Hartmann, 209–246. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-663-11590-8_7

Jonkmann, K., K. Maaz, M. Neumann and C. Gresch 2010. Übergangsquoten und Zusammenhänge zu familiärem Hintergrund und schulischen Leistungen. Deskriptive Befunde. In *Der Übergang von der Grundschule in die weiterführende Schule. Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten*, eds. K. Maaz et al., 123–150. Bonn: Bundesministerium für Bildung und Forschung.

Kao, G. and M. Tienda 1995. Optimism and achievement. The educational performance of immigrant youth. *Social Science Quarterly* 76(1): 1–19.

Klein, D. 2014. MIMRGNS: Stata module to run margins after mi estimate. Statistical Software Components S457795, Boston College Department of Economics.

Klicpera, C., A. Schabmann and B. Gasteiger-Klicpera 1993. Lesen- und Schreibenlernen während der Pflichtschulzeit. Eine Längsschnittuntersuchung über die Häufigkeit und Stabilität von Lese- und Rechtschreibschwierigkeiten in einem Wiener Schulbezirk. *Zeitschrift für Kinder- und Jugendpsychiatrie* 21: 214–225.

Köller, O. 2004. *Konsequenzen von Leistungsgruppierungen*. Münster: Waxmann.

Lohmann, H. and O. Groh-Samberg 2010. Akzeptanz von Grundschulempfehlungen und Auswirkungen auf den weiteren Bildungsverlauf/Acceptance of secondary school track recommendations and their effects on educational achievement. *Zeitschrift für Soziologie* 39(6): 470–492. https://doi.org/10.1515/zfsoz-2010-0603

Maaz, K. et al. 2008. Der Übergang von der Grundschule in die weiterführende Schule. Die Rolle von Schüler- und Klassenmerkmalen beim Einschätzen der individuellen Lernkompetenz durch die Lehrkräfte. *Schweizerische Zeitschrift für Bildungswissenschaften* 30(3): 519–548. https://doi.org/10.1007/978-3-531-92216-4_7

Möller, J. and E.-M. Bonerad 2007. Fragebogen zur habituellen Lesemotivation. *Psychologie in Erziehung und Unterricht* 54(4): 259–267.

Möller, J. and U. Trautwein 2015. Selbstkonzept. In *Pädagogische Psychologie*, eds E. Wild and J. Möller, 177–199. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41291-2_8

Neugebauer, M. 2010. Bildungsungleichheit und Grundschulempfehlung beim Übergang auf das Gymnasium. Eine Dekomposition primärer und sekundärer Herkunftseffekte/Educational inequality and teacher recommendations at the transition to upper secondary school. A decomposition of primary and secondary effects of social origin. *Zeitschrift für Soziologie* 39(3): 202–214. https://doi.org/10.1515/zfsoz-2010-0303

Pearl, J. 2009. *Causality*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Pfost, M. et al. 2018. Zur Entwicklung von Gymnasiastinnen und Gymnasiasten ohne Gymnasialempfehlung. *Zeitschrift für Erziehungswissenschaft* 21(3): 511–534. https://doi.org/10.1007/s11618-017-0787-6

Pohl, S. et al. 2012. NEPS technical report for reading. Scaling results of starting cohort 3 in fifth grade. NEPS Working Paper No. 15, Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

Rabe-Hesketh, S. and A. Skrondal 2012. *Multilevel and longitudinal modeling using Stata*. Third edition. College Station, Texas: Stata Press Publication.

Rathmann, K., M. Herke and M. Richter 2017. Die Entwicklung der selbstberichteten Gesundheit und des Wohlbefindens in der Sekundarstufe. Eine Analyse individueller Wachstumskurven. *Das Gesundheitswesen* 79(8/9): 656–804. https://doi.org/10.1055/s-0037-1606032

Ravens-Sieberer, U., E. Görtler and M. Bullinger 2000. Subjektive Gesundheit und Gesundheitsverhalten von Kindern und Jugendlichen. Eine Befragung Hamburger Schüler im Rahmen der schulärztlichen Untersuchung. *Das Gesundheitswesen* 62(3): 148–155. https://doi.org/10.1055/s-2000-10487

Schulze, A., F. Wolter and R. Unger 2009. Bildungschancen von Grundschülern. Die Bedeutung des Klassen- und Schulkontextes am Übergang auf die Sekundarstufe I. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 61(3): 411–435. https://doi.org/10.1007/s11577-009-0072-7

Sewell, W. H. and V. P. Shah 1968. Social class, parental encouragement, and educational aspirations. *American Journal of Sociology* 73(5): 559–572. https://doi.org/10.1086/224530

Statistisches Bundesamt 2016. Statistisches Jahrbuch Deutschland 2016. Wiesbaden: Statistisches Bundesamt.

Stern, E. 2008. Verpasste Chancen? Was wir aus der LOGIK-Studie über den Mathematikunterricht lernen können. In *Entwicklung von der Kindheit bis zum Erwachsenenalter. Befunde der Münchner Längsschnittstudie LOGIK*, ed. W. Schneider, 187–202. Weinheim: Beltz.

Stocké, V. 2007. Explaining Educational Decision and Effects of Families' Social Class Position. An Empirical Test of the Breen Goldthorpe Model of Educational Attainment. *European Sociological Review* 23(4): 505–519. https://doi.org/10.1093/esr/jcm014

Stubbe, T. C. and W. Bos 2008. Schullaufbahnempfehlungen von Lehrkräften und Schullaufbahnentscheidungen von Eltern am Ende der vierten Jahrgangsstufe. *Empirische Pädagogik* 22(1): 49–63.

Tiedemann, J. and E. Billmann-Mahecha 2010. Wie erfolgreich sind Gymnasiasten ohne Gymnasialempfehlung? Die Kluft zwischen Schullaufbahnempfehlung und Schulformwahl der Eltern. *Zeitschrift für Erziehungswissenschaft* 13(4): 649–660. https://doi.org/10.1007/s11618-010-0146-3

TNS Infratest Sozialforschung 2009. Leben in Deutschland. Befragung 2009 zur sozialen Lage der Haushalte. Personenfragebogen für alle. In SOEP 2009 – Erhebungsinstrumente 2009 (Welle 26) des Sozio-oekonomischen Panels. SOEP Survey Papers 106: Series A. Berlin: DIW/SOEP.

Triventi, M. 2013. The role of higher education stratification in the reproduction of social inequality in the labor market. *Research in Social Stratification and Mobility* 32(1): 45–63. https://doi.org/10.1016/j.rssm.2013.01.003

von Collani, G. and P. Y. Herzberg 2003. Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg. *Zeitschrift für Differentielle und Diagnostische Psychologie* 24: 3–7. https://doi.org/10.1024//0170-1789.24.1.3

Zeinz, H. and O. Köller 2006. Noten, soziale Vergleiche und Selbstkonzepte in der Grundschule. In *Risikofaktoren kindlicher Entwicklung*, ed. A. Schründer-Lenzen, 177–190. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90075-9_9

Appendix

Table A.1:
Imputed data summary

	Academic track only		Academic track and comprehensive	
Variable	**Complete**	**Imputed**	**Complete**	**Imputed**
Competences	7244	197	7244	188
Enjoy reading	7870	175	7883	162
Satisfaction	8071	103	8084	90
Parental ISEI	8092	43	8092	43
Age Wave 1	7658	220	7658	219
Grade Math	7691	163	7703	150
Grade German	7700	165	7711	152
Health	8020	78	8033	65
Parental Education	8212	2	8212	2
Migration	7852	120	7852	120
Days absent	6686	220	6699	207

Note: Variables that are not included in this list were not imputed since either no values are missing or cases with missing information were deleted. Some central variables were not imputed, but all cases with missing information were dropped since the information is too critical to be imputed (e.g., type of track attended).

Figure A.1:
Imputation convergence plots, academic track only

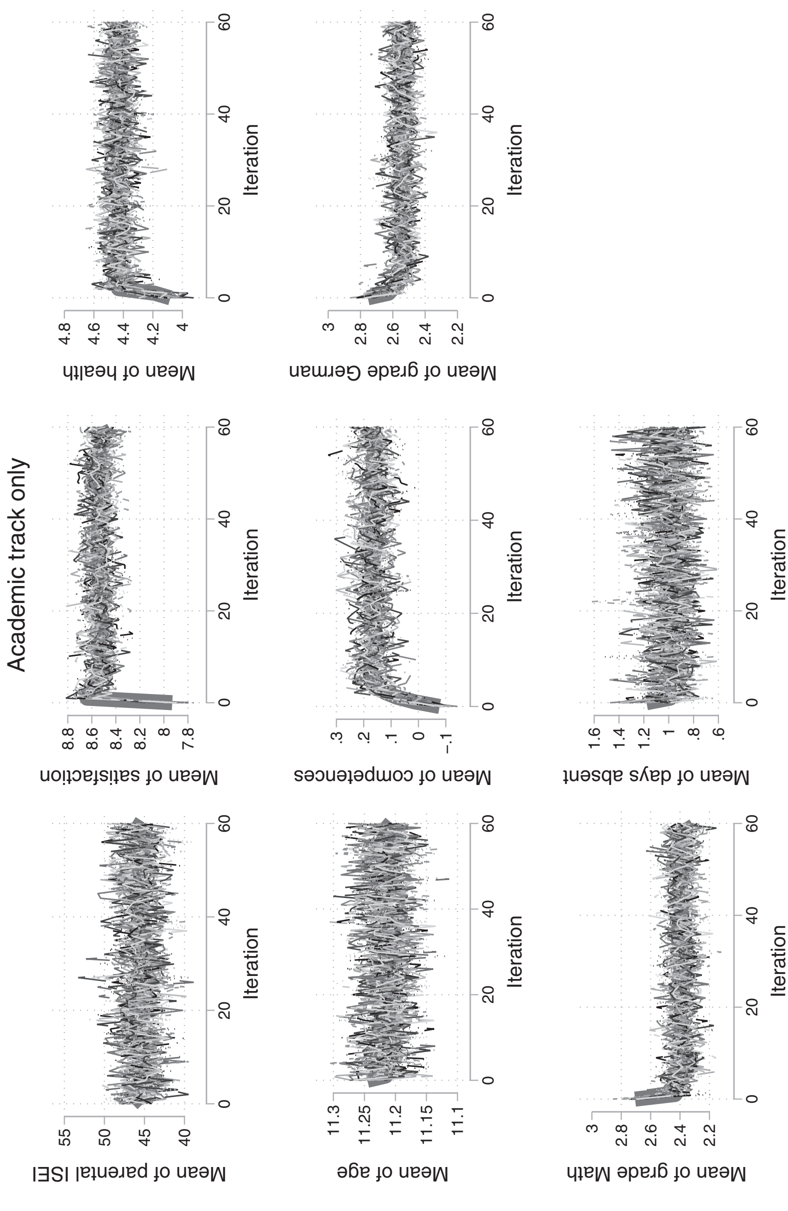


Figure A.2:
Imputation convergence plots, academic track and comprehensive schools

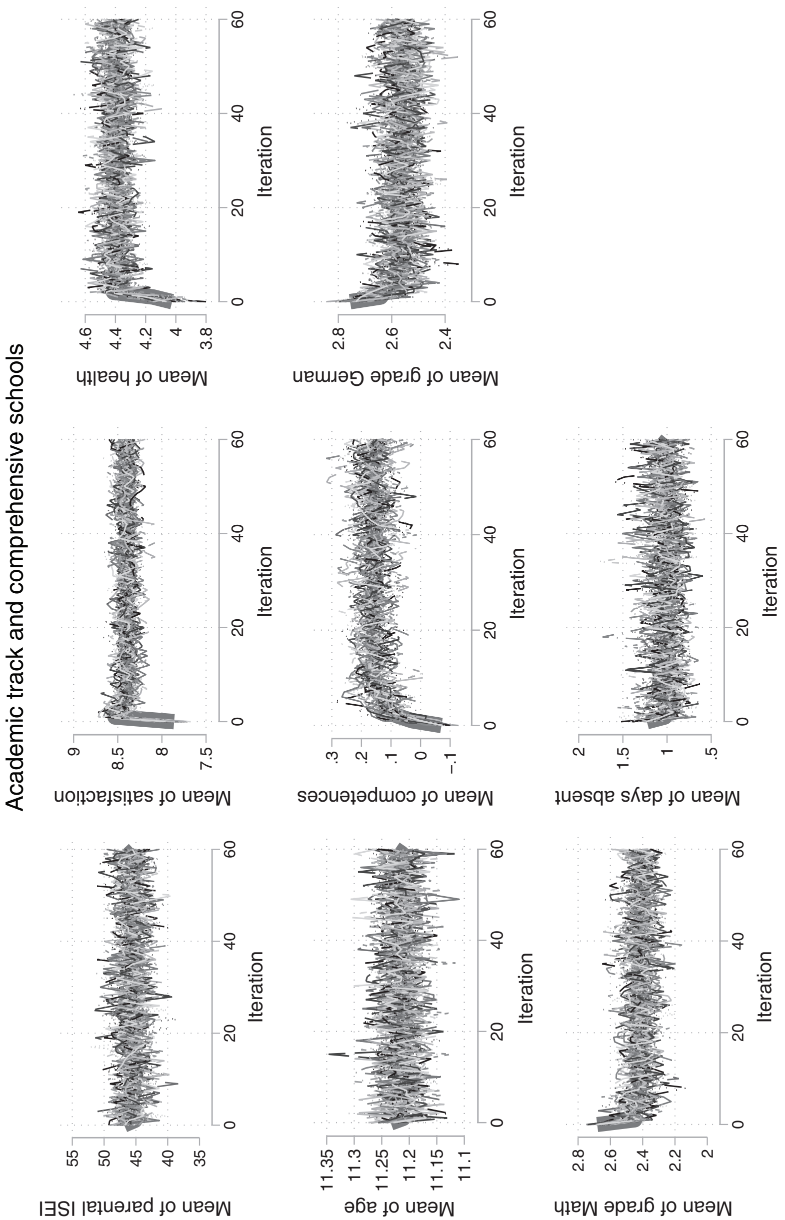


Table A.2:
Numerical regression output for retention rates

	Academic track only		Academic track and comprehensive	
	Base model	**Controls added**	**Base model**	**Controls added**
Recommendation	−0.001	−0.000	−0.001	0.000
	(0.014)	(0.014)	(0.008)	(0.008)
Wave 2	−0.009	−0.009	−0.011	−0.011
	(0.016)	(0.016)	(0.010)	(0.010)
Wave 3	−0.124***	−0.124***	−0.053***	−0.053***
	(0.016)	(0.016)	(0.010)	(0.010)
Wave 4	−0.181***	−0.181***	−0.057***	−0.058***
	(0.017)	(0.017)	(0.010)	(0.010)
Wave 5	−0.195***	−0.194***	−0.064***	−0.064***
	(0.017)	(0.017)	(0.010)	(0.010)
Recommendation ∗ Wave 2	0.008	0.008	0.010	0.010
	(0.017)	(0.017)	(0.010)	(0.010)
Recommendation ∗ Wave 3	0.106***	0.105***	0.041***	0.041***
	(0.017)	(0.017)	(0.011)	(0.011)
Recommendation ∗ Wave 4	0.166***	0.166***	0.044***	0.044***
	(0.018)	(0.018)	(0.011)	(0.011)
Recommendation ∗ Wave 5	0.165***	0.164***	0.041***	0.041***
	(0.018)	(0.018)	(0.011)	(0.011)
Female		0.006		0.006
		(0.006)		(0.004)
Both parents born in Germany		Ref.		Ref.
One parent born abroad		−0.017		−0.001
		(0.010)		(0.007)
Both parents born abroad		0.028*		0.022**
		(0.012)		(0.008)
Max. intermediate degree		Ref.		Ref.
Higher education eligibility		0.001		−0.002
		(0.009)		(0.006)
Tertiary degree		−0.006		−0.007
		(0.010)		(0.007)
Parental ISEI		0.000		0.000
		(0.000)		(0.000)
Age (Wave 1)		−0.019*		−0.004
		(0.008)		(0.006)
Not intact family		0.001		−0.003
		(0.008)		(0.006)
Constant	1.001***	1.207***	1.000***	1.038***
	(0.013)	(0.094)	(0.007)	(0.066)
Observations	3949	3949	5146	5146

Source: NEPS SC3, imputed data.
Note: Standard errors in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A.3:
Numerical regression results for both competence measures

	Competences, academic only	Grades, academic only	Competences, academic/ comprehensive	Grades, academic/ comprehensive
Recommendation	0.521***	−0.637***	0.776***	−0.882***
	(0.076)	(0.089)	(0.061)	(0.069)
Wave 2	−0.037	0.410***	−0.091*	−0.015
	(0.062)	(0.098)	(0.041)	(0.072)
Wave 3	−0.021	0.307**	−0.045	−0.039
	(0.062)	(0.099)	(0.041)	(0.075)
Wave 4	0.007	0.234*	−0.070	−0.026
	(0.079)	(0.103)	(0.052)	(0.077)
Wave 5	−0.105	0.336**	−0.073	0.018
	(0.074)	(0.114)	(0.046)	(0.074)
Recommendation ∗ Wave 2	−0.016	−0.114	0.025	0.242**
	(0.065)	(0.103)	(0.045)	(0.079)
Recommendation ∗ Wave 3	−0.040	−0.036	−0.027	0.265**
	(0.064)	(0.105)	(0.044)	(0.081)
Recommendation ∗ Wave 4	−0.108	0.092	−0.041	0.299***
	(0.082)	(0.109)	(0.055)	(0.083)
Recommendation ∗ Wave 5	0.007	0.031	−0.028	0.311***
	(0.077)	(0.119)	(0.050)	(0.081)
Female	0.124**	−0.171***	0.099*	−0.128**
	(0.046)	(0.045)	(0.042)	(0.039)
Both parents born in Germany	Ref.	Ref.	Ref.	Ref.
One parent born abroad	−0.186*	0.139	−0.177*	0.142*
	(0.076)	(0.075)	(0.071)	(0.065)
Both parents born abroad	−0.344***	0.318***	−0.404***	0.299***
	(0.088)	(0.085)	(0.086)	(0.072)
Max. intermediate degree	Ref.	Ref.	Ref.	Ref.
Higher education eligibility	−0.007	−0.043	0.104	−0.011
	(0.066)	(0.063)	(0.060)	(0.054)
Tertiary degree	0.010	−0.152*	0.105	−0.136*
	(0.074)	(0.072)	(0.068)	(0.062)
Parental ISEI	0.006**	−0.004*	0.007***	−0.003*
	(0.002)	(0.002)	(0.002)	(0.002)
Age (Wave 1)	−0.085	0.027	−0.115*	0.063
	(0.059)	(0.056)	(0.046)	(0.047)
Not intact family	−0.023	0.177**	−0.046	0.145**
	(0.060)	(0.059)	(0.054)	(0.050)
Federal state	Yes	Yes	Yes	Yes
Constant	0.528	0.143	0.465	−0.086
	(0.685)	(0.642)	(0.537)	(0.550)
Observations	3704	3869	4795	4978

Source: NEPS SC3, imputed data.
Note: Standard errors in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A.4:
Numerical regression results for all four outcomes

	Academic track only				Academic track and comprehensive schools			
	Satisfaction	Enjoy reading	Health	Days absent	Satisfaction	Enjoy reading	Health	Days absent
Recommendation	0.289*	0.131	−0.058	0.291	0.290**	0.242***	−0.044	0.059
	(0.137)	(0.071)	(0.077)	(0.257)	(0.107)	(0.055)	(0.059)	(0.189)
Wave 2	−0.358*	−0.248***	−0.110	−0.061	−0.370**	−0.291***	−0.134*	−0.039
	(0.157)	(0.075)	(0.088)	(0.290)	(0.115)	(0.054)	(0.062)	(0.199)
Wave 3	−0.377*	−0.188*	0.003	0.620*	−0.612***	−0.405***	−0.096	0.566**
	(0.155)	(0.074)	(0.091)	(0.309)	(0.117)	(0.055)	(0.064)	(0.214)
Wave 4	−0.229	−0.133	0.027	1.160***	−0.604***	−0.274***	−0.144*	0.772***
	(0.160)	(0.075)	(0.093)	(0.306)	(0.123)	(0.057)	(0.066)	(0.222)
Wave 5	−0.485**	−0.415***	−0.046	1.136***	−0.709***	−0.381***	−0.156*	0.927***
	(0.170)	(0.079)	(0.099)	(0.295)	(0.122)	(0.057)	(0.066)	(0.206)
Recommendation $*$ Wave 2	−0.095	0.026	−0.006	0.040	−0.074	0.043	0.023	−0.001
	(0.166)	(0.079)	(0.094)	(0.305)	(0.127)	(0.059)	(0.068)	(0.218)
Recommendation $*$ Wave 3	−0.249	−0.123	−0.082	−0.388	−0.010	0.067	0.025	−0.270
	(0.165)	(0.079)	(0.096)	(0.323)	(0.128)	(0.060)	(0.070)	(0.232)
Recommendation $*$ Wave 4	−0.433*	−0.257**	−0.136	−0.745*	−0.002	−0.135*	0.051	−0.297
	(0.170)	(0.080)	(0.098)	(0.320)	(0.134)	(0.062)	(0.072)	(0.240)
Recommendation $*$ Wave 5	−0.490**	0.021	−0.155	−0.698*	−0.259	−0.011	−0.029	−0.424
	(0.179)	(0.084)	(0.104)	(0.309)	(0.133)	(0.062)	(0.072)	(0.224)
Female	0.170*	0.297***	0.074*	0.169	0.160**	0.279***	0.045	0.140
	(0.066)	(0.037)	(0.038)	(0.108)	(0.058)	(0.032)	(0.033)	(0.095)

Continued

Table A.4:
Continued

	Academic track only				Academic track and comprehensive schools			
	Satisfaction	Enjoy reading	Health	Days absent	Satisfaction	Enjoy reading	Health	Days absent
Both parents born in Germany	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
One parent born abroad	0.018	0.066	−0.027	0.032	0.051	0.050	−0.034	−0.022
	(0.112)	(0.061)	(0.062)	(0.178)	(0.096)	(0.053)	(0.055)	(0.161)
Both parents born abroad	−0.090	0.046	0.032	0.061	0.034	−0.034	0.046	0.119
	(0.124)	(0.069)	(0.070)	(0.203)	(0.108)	(0.060)	(0.061)	(0.176)
Max. intermediate degree	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Higher education eligibility	0.036	0.054	0.049	0.127	−0.019	0.084	0.005	0.006
	(0.092)	(0.051)	(0.052)	(0.150)	(0.080)	(0.045)	(0.046)	(0.130)
Tertiary degree	−0.053	0.034	0.030	0.020	−0.081	0.071	−0.019	−0.083
	(0.105)	(0.058)	(0.060)	(0.172)	(0.091)	(0.050)	(0.052)	(0.149)
Parental ISEI	0.005^{*}	0.005^{**}	0.002	−0.001	0.006^{**}	0.003^{*}	0.002	0.001
	(0.003)	(0.001)	(0.002)	(0.004)	(0.002)	(0.001)	(0.001)	(0.004)
Age (Wave 1)	−0.014	0.017	0.001	0.047	−0.060	−0.028	−0.016	0.014
	(0.083)	(0.044)	(0.047)	(0.135)	(0.068)	(0.038)	(0.039)	(0.116)
Not intact family	-0.306^{***}	−0.070	0.019	0.133	-0.431^{***}	−0.028	−0.024	0.180
	(0.086)	(0.048)	(0.049)	(0.141)	(0.074)	(0.041)	(0.042)	(0.119)
Federal state	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	8.645^{***}	2.566^{***}	4.183^{***}	−1.437	9.199^{***}	3.089^{***}	4.435^{***}	−1.007
	(0.955)	(0.505)	(0.538)	(1.553)	(0.798)	(0.446)	(0.451)	(1.349)
Observations	3942	3900	3908	3497	5135	5075	5084	4473

Source: NEPS SC3, imputed data.
Note: Standard errors in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 503–543

Is educational wellbeing associated with grade repetition and school dropout rates among Indian students? Evidence from a panel study

***Ronak Paul*[1] *and Rashmi*[1,*]**

Abstract

Despite the Indian government's continuing efforts to encourage children to attend school, levels of educational wellbeing among some groups of children during their elementary schooling remain low. High school dropout and grade repetition rates are among the negative and deleterious outcomes of poor educational wellbeing in children that are rarely discussed as policy issues. Using the panel dataset of the India Human Development Survey (IHDS) conducted in 2005 and 2012, this study explores the effects of educational wellbeing on children's later educational outcomes, as measured by their school dropout and grade repetition rates. Variation in the educational outcomes of children across states was also examined. The results show that the children whose educational wellbeing index was below average during their elementary schooling were more likely to drop out of school or repeat a grade in early adolescence. For policymakers, this study highlights that the experiences of children during their elementary schooling merit more attention.

Keywords: educational wellbeing; school dropout; grade repetition; educational outcome; India Human Development Survey

1 Introduction

Education is a complex process that involves a combination of individual ability and motivation, as well as resources provided by the family, and actions taken by communities and states (Darling-Hammond et al. 2020). Thus, education can be seen as a lifelong process that ends with death, and that improves the lives of

[1]International Institute for Population Sciences, Mumbai, India

[*]Correspondence to: Rashmi, rashmir635@gmail.com

DOI: 10.1553/populationyearbook2021.res5.2

individuals, while supporting the progress of nations. Globally, various programs have been proposed that seek to improve the lives of all people based on a commitment to "Education for All". Despite being a developing country, India has implemented educational initiatives within the framework of the country's Global Goals, including a universal elementary education program in 1992, the Sarva Shiksha Abhiyan program, the Right to Education Act, and a mid-day meal scheme (Govinda and Bandyopadhyay 2008). While the achievements of these initiatives, including improvements in access to education, have been reported in the previous literature, the goal of the universalisation of education has yet to be realised in India (Chudgar 2009; Gouda and Sekher 2014).

A recent report on the National Education Policy (NEP) of India found that the government's commitment to ensuring that all children attend school starting at early ages is likely to be achieved in the near future, as the gross enrolment ratio among children in grades 6–8 recently reached 91% (MHRD 2020). However, the same report showed that the gross enrolment ratio was 79% among children in grades 9–10, and was 56% among children in grades 11–12, which indicates that many children leave education during adolescence. Moreover, while the Annual Status of Education Report (ASER) 2018 found that 97% of Indian children had enrolled in school, it also showed that half of these children were unable to read grade 2-level textbooks at ages 10–11, and only 28% of students were able to divide appropriately in grade 5. Against this background, we seek to answer the question of how the educational wellbeing of children during their elementary schooling contributes to their likelihood of dropping out of school or repeating a grade during adolescence.

Wellbeing is defined as a state in which the physical, cognitive and social-emotional dimensions of an individual are well-integrated (Bornstein et al. 2003). It is a complex concept that includes both objective and subjective measures related to the individual's circumstances. An Indian study published in 2019 listed 13 different domains of wellbeing, among which education was identified as a potential driver of the wellbeing of individuals (Singh et al. 2019). Different approaches to measuring wellbeing may include the use of both objective and subjective criteria that are designed to reflect the quality of life, experiences and needs of the population (see Table 1). Educational wellbeing in children is usually assessed by measuring children's school performance, learning skills, cognitive development, involvement in education and enjoyment of education (Sylva 1994; Wat 2015). When children fail to learn and to perform well in their early years of schooling, they often develop a negative attitude towards education, which may, in turn, further disrupt their education during adolescence. Thus, a lack of educational wellbeing may be associated with higher school dropout and grade repetition rates. A child is considered a school dropout if he or she fails to complete the basic cycle of schooling (Sabates et al. 2010). Grade repetition is the practice of having a child attend the same grade for another year because she or he was unable to fulfil a pre-set academic goal (Haidary 2013).

Table 1:
List of indicators of educational wellbeing and reasons for inclusion

S. No.	Indicators	Reason for inclusion
1. Objective educational wellbeing index	Reading skills, mathematical skills, writing skills	Rather than focusing on the self-reported educational attainment of children, it is more useful to evaluate the extent to which students can use their reading, mathematical and writing skills in real-life situations. These skills are cognitive abilities that students must develop in order to use the information mentioned or implied (OECD 2003).
2. Subjective educational wellbeing index	Type of student, student enjoys school	Wellbeing includes both subjective and objective aspects of life. These subjective educational wellbeing indicators can be used to measure aspects of children's school performance that are directly or indirectly related to their educational attainment (Sylva 1994).

Source: Author's elaboration.

There is, however, little previous research on the role educational wellbeing plays in the likelihood that a child will drop out of school or experience repetitive failures (Falch and Strøm 2008; Glick and Sahn 2010). Most of the existing studies on this topic for India focused on identifying the determinants of high school dropout and grade repetition rates, including poverty (Rao 2000), the gender of the children (Upendranath 1995), the educational status of the parents (Pratinidhi et al. 1992), school facilities (Gouda et al. 2013), private tuition (Dongre and Tewary 2015), whether the schools are public or private (Desai et al. 2008b; Singhal and Das 2019), parental aspirations (Sethi 2018), family size (Choudhury 2006), religion and household characteristics (Bhat and Zavier 2005; Gouda and Sekher 2014). Moreover, it has been shown that the educational outcomes of children vary across the different regions and states of India (Gouda and Sekher 2014).

In addition to the lack of research on the determinants high school dropout and grade repetition rates, only a few studies have investigated how grade repetition affects the stress levels of early adolescents (Jimerson 2001; McGrath 2006); or the long-term effects of educational failures on levels of poverty, child labour, illiteracy, unemployment, low wages, participation in violence and crime, and mental and physical health problems (Beauvais et al. 1996; Desai 1991; UNESCO 2012). While these consequences are well-known, there is a lack of research in the Indian context on the association between educational wellbeing and grade repetition and dropout rates. To help fill this knowledge gap, the present study explores how the educational wellbeing of children affects their later educational development, as reflected in their school dropout and grade repetition rates. The panel dataset of the India Human Development Survey (IHDS) conducted in 2005 and 2012 further strengthens our study, as it measured the contribution of the educational wellbeing of children at ages 8–11 in round one, and again in early adolescence (at ages 15–18) in round two. Our study also takes into account the variation across states in detrimental educational outcomes – namely, dropping out of school and repeating a grade – while examining the hypothesised association.

2 Data, variables and methods

2.1 Data source

This study uses data from round one and round two of the India Human Development Survey (IHDS), which provide information on the health and morbidity, education, employment, economic status, nuptiality, fertility, gender relations and social capital of a nationally representative sample of 41,554 households. Round one of the IHDS was conducted in 2005–2006. In round two of the IHDS, which was carried out in 2011–2012, 83% of the households surveyed in round one were re-interviewed. Unlike other cross-sectional surveys conducted in India, such as the National Family Health Survey (NFHS), the IHDS covered a wide range of topics, and had a panel design. The National Council of Applied Economic Research (NCAER) conducted both rounds of the IHDS in collaboration with the University of Maryland, USA. The IHDS had a multistage, stratified, random sampling survey design, and covered all of the states and the union territories of India. Further details regarding the IHDS' sample frame, sampling procedures and data collection approaches can be found elsewhere (Desai et al. 2010, 2015).

Our study utilises the panel data for 9840 students who were 8–11 years old in 2005 (round one), and who were 15–18 years old in 2012 (round two). Of the 17,061 children who participated in round one (cross-sectional dataset), 104 had died, 3454 had migrated and 3663 were untraceable during round two. Furthermore, we have excluded the data for 18 children for whom the information regarding their school dropout and grade repetition status was missing in round two. Thus, the analytical sample we use to investigate the relationship between educational wellbeing in round one and school dropout and grade retention status in round two consists of 9822 students.

2.2 Ethics statement

The analysis presented in this article is based on publicly available secondary data with no identifiable information. Therefore, no prior ethical approval for our study was necessary. The data can be obtained from the Inter-university Consortium for

Political and Social Research (ICPSR) data repository (Desai and Vanneman 2015; Desai et al. 2008a).

2.3 Outcome variables

The indicators of the school dropout status (no, yes) and the grade repetition status (no, yes) of the students in our analytical sample between 2005 and 2012 are the two outcome variables of our study. The school dropout status of the students was obtained from the binary indicators of their school enrolment status in 2005 and 2012, respectively. Children who were enrolled in school during round one, but who were not enrolled during round two, were categorised as "yes" (school dropout), and were otherwise categorised as "no". Similarly, the grade repetition status of students was obtained from the responses to the following item collected in both rounds: "has the student ever repeated a grade – yes, no". Students who had never repeated a grade in 2005, but who had repeated at least one grade in 2012, were categorised as "yes", and were otherwise categorised as "no".

2.4 Key explanatory variable

The two binary indicators of educational wellbeing are the key independent variables of our study. While wellbeing cannot be measured directly, it can be estimated using closely related subjective and objective measures. To measure the educational wellbeing of students, we created two educational wellbeing indices (the objective educational wellbeing index and the subjective educational wellbeing index) from five educational outcome indicators. These educational outcome indicators were provided only for students aged 8–11 in round one of the IHDS (Desai et al. 2010). Two of these indicators were reported by the child's mother, and are subjective measures of educational wellbeing: namely, what type of student the child is (below average, average, above average), and whether the child enjoys school (no, yes). The other three indicators are objective measures of educational wellbeing: namely, the child's reading, mathematical and writing skills. These measures were developed from short reading, mathematical and writing tests administered to all students aged 8–11 in the households the IHDS surveyed. The reading skills of students are divided into five categories: (i) cannot read at all, (ii) can read alphabets but not words, (iii) can read words but cannot read full sentences, (iv) can read a short paragraph of 2–3 sentences but cannot read a full page and (v) can read a full story. The mathematical skills of students are divided into four categories: (i) cannot read numbers, (ii) can recognise numbers but cannot do any arithmetic operations, (iii) can subtract a two-digit number from another number and (iv) can divide a three-digit number by a one-digit number. The writing skills of students are divided into two categories: (i) cannot write at all and (ii) can write a sentence with two or fewer mistakes. Distribution of students by these variables are presented in Table A.1 and the technical details of the construction of the index are provided in Section 2.6.1.

2.5 Control variables

Previous research has shown that several factors influence the school dropout and grade repetition rates of students (Falch and Strøm 2008; Glick and Sahn 2010; Gouda and Sekher 2014). Accordingly, we have included relevant student-related characteristics, parent-related characteristics and socio-economic characteristics of the household as control variables. The student-related characteristics are as follows: the age of the student (in completed years), the gender of the student (male, female), the type of school the student attends (public school, private school, other) and whether the student receives private tuition (no, yes). The parent-related control variables are as follows: mother's level of education (no formal schooling, less than five years of schooling, six to 10 years of schooling, more than 10 years of schooling), mother's working status (not working, working), father's level of education (no formal schooling, less than five years of schooling, six to 10 years of schooling, more than 10 years of schooling) and father's working status (not working, working). The control variables related to the household's socio-economic characteristics are as follows: household wealth quintile (richest, rich, middle, poor and poorest), household poverty status (not poor, poor), the caste of the household head (Other Backward Class (OBC), Scheduled Castes (SC), Scheduled Tribes (ST), other), the religion of the household head (Hindu, Muslim, other) and the place of residence (rural, urban). Additionally, the region of the country (northern, north-eastern, central, eastern, western, southern) the student came from is included as a state-level characteristic. All of these variables are measured for the panel of children aged 8–11 in 2005.

We have estimated the wealth quintile for all households in round one using data on each household's asset ownership, material type, water source, sanitation facility type and number of bedrooms. This approach is based on globally-defined standard procedures (Filmer and Scott 2008; Rutstein and Johnson 2004). The wealth scores for each household have been generated using principal component analysis (Filmer and Scott 2008). Based on the wealth score, the households have been classified into five categories (richest, rich, middle, poor, poorest), such that the households with the lowest 20th percentile score are assigned to the "poorest" category; the households with the next-lowest 20th percentile score are assigned to the "poor" category; and so on.

During round one, information was collected about the religion and caste of all household heads. Based on their religion, the households have been assigned to nine categories: Hindu, Muslim, Christian, Sikh, Buddhist, Jain, tribal, other and none. Owing to the skewed population distribution, for the purposes of this study, we have recoded the religion of the household head into three categories: Hindu, Muslim and other, whereby the other category consists of all other religious categories except Hindu and Muslim. Similarly, the IHDS classified the caste of the household head into five categories: Brahmin, Other Backward Caste [OBC], Scheduled Caste [SC], Scheduled Tribe [ST] and other. Again, we have recoded the caste variable into four categories: OBC, SC, ST and other, whereby the other category combines the

Brahmin and the other categories from the original variable. The respondents in the ST and the SC categories belong to the most socially backward group of people; i.e., they are members of the lower rungs of the now constitutionally abolished Indian caste system. As the name suggests, the respondents in the OBC category also belong to a socially and economically backward group of people, although they tend to have better conditions than those of the SC/ST people. The other category consists of all respondents who do not belong to any of the three caste groups.

We constructed the country regions by assigning the current and erstwhile 33 states and union territories of India to six categories. The northern region includes Chandigarh, Delhi, Haryana Himachal Pradesh, erstwhile Jammu and Kashmir, Punjab, Uttaranchal and Rajasthan. The north-eastern region includes Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Tripura and Sikkim. The central region consists of Madhya Pradesh and Chhattisgarh. The eastern zone consists of Bihar, Jharkhand, Odisha and West Bengal. The western region comprises Dadra and Nagar Haveli, Daman and Diu, Goa, Gujarat and Maharashtra. The southern region comprises erstwhile Andhra Pradesh, Karnataka, Kerala, Tamil Nadu and Pondicherry (Sanjay 2020).

2.6 Statistical methods

2.6.1 Construction of the objective and the subjective educational wellbeing index

The educational wellbeing index was constructed using principal component factor analysis (PCFA) based on the variables described in Section 2.4. PCFA is a dimension reduction method that is utilised to reduce the information provided by a large group of variables (indicators of educational wellbeing) to form a smaller group (educational wellbeing index) that contains the majority of the information from the larger group (Fabrigar et al. 1999). The use of a composite index, rather than numerous indicators, has the added advantage of being easier to interpret and to compute (Costello and Osborne 2005). The PCFA method transforms a large number of correlated variables into a smaller number of uncorrelated factors. The first principal factor accounts for as much of the variability in the data as possible, with each succeeding factor then accounting for as much of the remaining variability as possible (Costello and Osborne 2005). The relevance of each factor can be determined from the eigenvalue, which indicates how much of the total variance is explained by each of the factors (Acock 2013). Any factor with an eigenvalue of less than or equal to one is ignored. The factor loadings are the correlation between each variable and the factor we are interested in (Acock 2013). Factor loading values of greater than 0.25 indicate that the variable is relevant in defining the dimensionality of the factor, with higher values denoting increasing relevance (Acock 2013). A negative factor loading value indicates an inverse correlation with the factor.

Table 2:
Factor loading and inter-item correlation obtained from the principal component factor analysis (PCFA)

Scale	Items	Factor Loading	KMO Value
Objective educational wellbeing index			
Eigen value = 2.38	Reading skills	0.88	0.70
Explained variance = 47.5%	Mathematical skills	0.88	0.72
	Writing skills	0.84	0.80
Subjective educational wellbeing index			
Eigen value = 1.29	Type of student	0.60	0.84
Explained variance = 25.8%	Student enjoys school	0.89	0.74

Note: (a) Kaiser-Meier-Olkin measure indicates whether there is sufficient correlation among the indicator variables to perform PCFA.
Source: Author's calculation using IHDS data (2005).

The PCFA for the indicators of educational wellbeing result in a two-factor solution. Looking at Table 2, we can observe that both factor 1 and factor 2 have an eigenvalue greater than one, with each factor explaining 48% and 26% of the variance of all of the indicators of educational wellbeing, respectively. Furthermore, all of the indicators have factor loading values of more than 0.60. In addition, all of the indicators have Kaiser-Meier-Olkin (KMO) values greater than 0.70, which justifies our use of PCFA (KMO values greater than 0.50 are necessary for conducting PCFA).

In addition, we have generated objective and subjective educational wellbeing scores based on the first and second factors, respectively. The objective and subjective scores have mean values of zero and median values of 0.38 and 0.14, respectively, thereby denoting a negatively skewed distribution. Thus, we have obtained two binary indexes of objective educational wellbeing and subjective educational wellbeing in which students have a score that is below the median value (below-average score), or a score that is higher than or equal to the median value (average and above score).

As we noted above, we have chosen to use PCFA over other methods of index construction that are based on the reliability coefficient only (Muttarak and Chankrajang 2015). While measures based on reliability coefficients give equal weights to all components of the index, this approach is rarely applicable in practical scenarios (Kline 2000). The advantage of using PCA is that it weights each component proportional to its correlation with the concept that we are trying to measure with the index. Therefore, an index constructed through PCA has been shown to better represent the concept in terms of measurement validity (Acock 2013; Costello and Osborne 2005).

2.6.2 Bivariate and multivariate analysis

We use bivariate and multivariable analysis to explore the effects of children's educational wellbeing on their later educational outcomes, as measured by their school dropout and grade repetition rates. Bivariate analysis is carried out using the chi-square test for association. The outcome variables, school dropout and grade repetition status are binary. As further sampling units in the survey have been chosen using a stratified random sampling approach, observations within the same strata will have a certain level of correlation among them. Thus, to account for the type of the dependent variable and the stratified nature of the data, we have estimated two-level random intercept logistic regression models. In the two-level model, 9822 students are included in level 1, and these students are nested within 21 states in level 2. Note that because of the skewed distribution of students across the 33 states (in the study sample), we have merged them into 21 groups, such that the union territories, the states in the north-eastern region, and Maharashtra and Goa are in singular groups. The grouping of union territories is based on the similarity of the administrative structure, whereas the grouping of the north-eastern states and Maharashtra and Goa is based on geographical proximity. In addition, although the students are nested within households, which are, in turn, nested within states in the data, we do not include the household as a separate level, as the average number of students per household (1.2) is very low (Steele 2013).

The multivariable model gives an odds ratio, which is the odds of a person being in a particular category in comparison to the reference category after adjusting for the effects of all other variables, and for the group-level effects (Steele 2013). In addition, the multilevel model gives the intraclass correlation coefficient (ICC), which measures the expected degree of similarity (homogeneity) of the school dropout (or grade repetition) rate among students in the same state. The ICC at the state level is the ratio of unexplained variation in the risk of dropping out of school (or of repeating a grade) across the states to the sum of the total unexplained variation. The value of the ICC lies within zero and one, such that the higher the value of the ICC, the greater the degree of within-state correlation is (Merlo et al. 2006). The median odds ratio (MOR) measures the expected degree of heterogeneity of the school dropout (or grade repetition) rate among students in the different states. The MOR at the state level gives the median of the ratio of the likelihood of dropping out of school (or of repeating a grade) among all pairs of students belonging to high-risk and low-risk states. The value of the MOR is always greater than or equal to one, such that a higher MOR value denotes a greater level of heterogeneity (Merlo et al. 2006).

We checked for multicollinearity and found that the mean-variance inflation factor (VIF) for all multivariable models was less than 1.28 (Ender 2010). Therefore, multicollinearity does not affect our multivariate estimations. We performed all statistical estimations using the Stata software version 13 (StataCorp 2014).

3 Results

3.1 Sample description

Of the children in our study sample, nearly 57% had average and above objective educational wellbeing scores, and 50% had average and above subjective educational wellbeing scores in 2005. Looking at Table 3, we can see that of the panel of children, 53% were male and 47% were female. Moreover, 70% of the students were attending public school, and 82% were not receiving private tuition in 2005. Turning to parental education, we can see that 55% of the students had a mother and 26% had a father with no formal schooling. In terms of household economic status, 38% of the students came from a household in the poor/poorest wealth quintile, and 29% came from a household below the poverty line (BPL). Moreover, 40% of the students were from OBC households, 79% were from Hindu families and 72% were from rural households. In terms of region of the country, the vast majority of students came from the northern region (38%), distantly followed by the southern regions (18%).

The last column of Table 3 gives the absolute difference in the percentage distribution of students across the cross-sectional and panel populations in 2005. Note that the percentage distribution of students across the characteristics is largely similar in the two datasets, which indicates that the study is unlikely to suffer from a sample selection bias, whereby students with certain selected characteristics had a greater likelihood of dropping out of the survey in the second wave. Only the percentage distribution by the father's education differed by more than 5% between the two datasets.

3.2 Bivariate results

Table 4 displays the bivariate association between the educational wellbeing of students and other control variables in round one with the dropout and grade repetition status of students in round two. We found that among the 9822 students, 2273 (23%) had dropped out of school and 1668 (17%) had repeated a grade between 2005–2012. The bivariate results showed that students with a below-average objective wellbeing score in 2005 were highly likely to have dropped out of school (31%) or repeated a grade (20%) by 2012. Similarly, among the students who had a below-average subjective educational wellbeing score, the proportions who dropped out of school (27%) or repeated a grade (18%) were relatively large. Furthermore, larger shares of public school than of private school students had dropped out (28%) or repeated a grade (20%) by 2012. Moreover, among the students whose mother had no formal schooling, the shares who dropped out of school (32%) or repeated a grade (19%) were disproportionately large. Likewise, among the students whose father had no formal schooling, the shares who dropped out of school (36%) or repeated a grade (19%) were relatively large.

Table 3:
Absolute (N) and percentage (%) distribution of students aged 8–11 years by relevant demographic and socio-economic characteristics across the cross-sectional and panel datasets in round one

	Students aged 8–11 years in 2005				
	Cross-sectional dataset		Panel dataset		Absolute difference
Characteristics	***N***	**%**	***N***	**%**	**%**
Age of student (in years)					
8	4,311	25.3	2,131	21.7	3.6
9	3,714	21.8	2,315	23.6	1.8
10	5,596	32.8	3,552	36.2	3.4
11	3,440	20.2	1,824	18.6	1.6
Gender of student					
Male	8,940	52.4	5,243	53.4	1.0
Female	8,121	47.6	4,579	46.6	1.0
Type of school attended by student					
Public school	10,308	66.6	6,846	69.7	3.1
Private school	4,167	26.9	2,417	24.6	2.3
Others	991	6.4	559	5.7	0.7
Student receives private tuition					
No	9,941	77.6	8,117	82.6	5.0
Yes	2,863	22.4	1,705	17.4	5.0
Mother's level of education					
More than 10 years of schooling	1,081	6.6	596	6.1	0.5
6 to 10 years of schooling	3,717	22.8	2,176	22.2	0.6
Less than 5 years of schooling	2,595	15.9	1,628	16.6	0.7
No formal schooling	8,943	54.7	5,422	55.2	0.5
Mother's working status					
Not working	12,040	73.3	7,303	74.4	1.1
Working	4,377	26.7	2,519	25.6	1.1
Father's level of education					
More than 10 years of schooling	2,357	15.3	1,297	13.2	2.1
6 to 10 years of schooling	5,495	35.6	4,220	43.0	7.4
Less than 5 years of schooling	2,913	18.9	1,732	17.6	1.3
No formal schooling	4,654	30.2	2,573	26.2	4.0
Father's working status					
Not working	5,189	33.5	3,114	31.7	1.8
Working	10,287	66.5	6,708	68.3	1.8

Continued

Table 3:
Continued

Characteristics	Students aged 8–11 years in 2005: Cross-sectional dataset *N*	%	Panel dataset *N*	%	Absolute difference %
Household wealth quintile					
Richest	3,285	19.3	1,909	19.4	0.1
Rich	3,593	21.1	2,099	21.4	0.3
Middle	3,437	20.1	2,077	21.1	1.0
Poor	3,404	20.0	1,932	19.7	0.3
Poorest	3,342	19.6	1,805	18.4	1.2
Household poverty status					
Not poor	12,146	71.2	7,003	71.3	0.1
Poor	4,915	28.8	2,819	28.7	0.1
Caste of household head					
Scheduled Tribes	1,333	7.8	702	7.1	0.7
Scheduled Castes	3,729	21.9	2,242	22.8	0.9
Other Backward Classes	6,889	40.4	3,928	40.0	0.4
Others	5,110	30.0	2,950	30.0	0.0
Religion of household head					
Hindu	13,353	78.3	7,781	79.2	0.9
Muslim	2,532	14.8	1,349	13.7	1.1
Others	1,176	6.9	692	7.0	0.1
Place of residence					
Rural	12,040	70.6	7,034	71.6	1.0
Urban	5,021	29.4	2,788	28.4	1.0
Country regions					
Northern	6,339	37.2	3,767	38.4	1.2
North-eastern	615	3.6	260	2.6	1.0
Central	1,888	11.1	1,096	11.2	0.1
Eastern	2,845	16.7	1,530	15.6	1.1
Western	2,210	13.0	1,437	14.6	1.6
Southern	3,164	18.5	1,732	17.6	0.9
Overall	**17,061**	**100**	**9,822**	**100**	

Source: Author's calculation using IHDS data (2005).

When we looked at the role of socio-economic characteristics, we found that the shares of students who dropped out were especially large among those who came from a household that was in the poor wealth quintile (39%) or below the poverty line (35%). We observed a similar pattern for grade repetition, whereby 23% of students from a household in the poor wealth quintile and 22% of students from

Table 4:
Bivariate association of the educational wellbeing, demographic and socio-economic characteristics in round one with the school dropout and grade repetition status of students in round two

	Students aged 15–18 years in 2012							
	Total	School dropout		Chi-square test	Total	Grade repetition		Chi-square test
Characteristics	*N*	*N*	%		*N*	*N*	%	
Objective educational wellbeing index								
Average and above	5,548	956	17.2	χ^2 = 250.41,	5,548	823	14.8	χ^2 = 41.73,
Below average	4,274	1,317	30.8	p-value = 0.001	4,274	845	19.8	p-value = 0.001
Subjective educational wellbeing index								
Average and above	4,892	959	19.6	χ^2 = 68.61,	4,892	797	16.3	χ^2 = 3.29,
Below average	4,930	1,314	26.7	p-value = 0.001	4,930	871	17.7	p-value = 0.069
Age of student (in years)								
8	2,131	371	17.4	χ^2 = 96.65,	2,131	289	13.6	χ^2 = 37.44,
9	2,315	461	19.9	p-value = 0.001	2,315	360	15.6	p-value = 0.001
10	3,552	926	26.1		3,552	657	18.5	
11	1,824	515	28.2		1,824	362	19.8	
Gender of student								
Male	5,243	1,196	22.8	χ^2 = 0.69,	5,243	948	18.1	χ^2 = 9.63,
Female	4,579	1,077	23.5	p-value = 0.406	4,579	720	15.7	p-value = 0.002
Type of school attended by student								
Government school	6,846	1,943	28.4	χ^2 = 366.64,	6,846	1,338	19.5	χ^2 = 107.84,
Private school	2,417	230	9.5	p-value = 0.001	2,417	255	10.6	p-value = 0.001
Others	559	100	17.9		559	75	13.4	

Continued

Table 4:
Continued

	Students aged 15–18 years in 2012							
	Total	School dropout			Total	Grade repetition		
Characteristics	*N*	*N*	%	Chi-square test	*N*	*N*	%	Chi-square test
Student receives private tuition								
No	8,117	2,058	25.4	$\chi^2 = 128.66$,	8,117	1,411	17.4	$\chi^2 = 5.33$,
Yes	1,705	215	12.6	p-value = 0.001	1,705	257	15.1	p-value = 0.021
Mother's level of education								
More than 10 years of schooling	596	15	2.5	$\chi^2 = 622.94$,	596	33	5.5	$\chi^2 = 107.49$,
6 to 10 years of schooling	2,176	197	9.1	p-value = 0.001	2,176	279	12.8	p-value = 0.001
Less than 5 years of schooling	1,628	334	20.5		1,628	316	19.4	
No formal schooling	5,422	1,727	31.9		5,422	1,040	19.2	
Mother's working status								
Not working	7,303	1,446	19.8	$\chi^2 = 178.79$,	7,303	1,117	15.3	$\chi^2 = 57.49$,
Working	2,519	827	32.8	p-value = 0.001	2,519	551	21.9	p-value = 0.001
Father's level of education								
More than 10 years of schooling	1,297	71	5.5	$\chi^2 = 666.23$,	1,297	125	9.6	$\chi^2 = 66.46$,
6 to 10 years of schooling	4,220	704	16.7	p-value = 0.001	4,220	713	16.9	p-value = 0.001
Less than 5 years of schooling	1,732	558	32.2		1,732	344	19.9	
No formal schooling	2,573	940	36.5		2,573	486	18.9	
Father's working status								
Not working	3,114	604	19.4	$\chi^2 = 35.96$,	3,114	451	14.5	$\chi^2 = 20.20$,
Working	6,708	1,669	24.9	p-value = 0.001	6,708	1,217	18.1	p-value = 0.001

Continued

Table 4: Continued

	Students aged 15–18 years in 2012							
	Total	School dropout		Chi-square test	Total	Grade repetition		Chi-square test
Characteristics	*N*	*N*	%		*N*	*N*	%	
Household wealth quintile								
Richest	1,909	118	6.2	χ^2 = 688.08,	1,909	181	9.5	χ^2 = 126.69,
Rich	2,099	339	16.2	p-value = 0.001	2,099	336	16.0	p-value = 0.001
Middle	2,077	531	25.6		2,077	369	17.8	
Poor	1,932	574	29.7		1,932	374	19.4	
Poorest	1,805	711	39.4		1,805	408	22.6	
Household poverty status								
Not poor	7,003	1,285	18.3	χ^2 = 315.10,	7,003	1,048	15.0	χ^2 = 70.42,
Poor	2,819	988	35.0	p-value = 0.001	2,819	620	22.0	p-value = 0.001
Caste of household head								
Scheduled Tribes	702	249	35.5	χ^2 = 147.73,	702	182	25.9	χ^2 = 106.20,
Scheduled Castes	2,242	590	26.3	p-value = 0.001	2,242	486	21.7	p-value = 0.001
Other Backward Classes	3,928	947	24.1		3,928	594	15.1	
Others	2,950	487	16.5		2,950	406	13.8	
Religion of household head								
Hindu	7,781	1,699	21.8	χ^2 = 84.14,	7,781	1,378	17.7	χ^2 = 16.17,
Muslim	1,349	442	32.8	p-value = 0.001	1,349	180	13.3	p-value = 0.001
Others	692	132	19.1		692	110	15.9	

Continued

Table 4:
Continued

	Students aged 15–18 years in 2012							
	Total	**School dropout**		**Chi-square test**	**Total**	**Grade repetition**		**Chi-square test**
Characteristics	***N***	***N***	**%**		***N***	***N***	**%**	
Place of residence								
Rural	7,034	1,767	25.1	χ^2 = 54.56,	7,034	1,278	18.2	χ^2 = 24.74,
Urban	2,788	506	18.1	p-value = 0.001	2,788	390	14.0	p-value = 0.001
Country regions								
Northern	3,767	744	19.8	χ^2 = 67.38,	3,767	618	16.4	χ^2 = 265.45,
North-eastern	260	58	22.3	p-value = 0.001	260	36	13.8	p-value = 0.001
Central	1,096	328	29.9		1,096	361	32.9	
Eastern	1,530	371	24.2		1,530	281	18.4	
Western	1,437	389	27.1		1,437	179	12.5	
Southern	1,732	383	22.1		1,732	193	11.1	
Overall	**9,822**	**2,273**	**23.1**		**9,822**	**1,668**	**17.0**	

Source: Author's calculation using IHDS data (2005 and 2012).

a BPL household had repeated a grade. We further observed that 36% of students who belonged to the Scheduled Tribes and 25% of students from a rural community had dropped out of school by 2012. Similarly, 26% of ST students and 18% of rural students had repeated a grade by 2012. Moreover, we found that the shares of students who dropped out of school (30%) or repeated a grade (33%) were relatively large in the central region of India.

3.3 Multivariable results

Table 5 provides the parameters of the group-level (state-level) effects from the random intercept logit models for the school dropout and grade repetition rates, respectively. The ICC from the null models showed that the state-level characteristics accounted for 11% and 8% of the unexplained variation among the students in the risk of dropping out of school or of repeating a grade in 2012, respectively. The MOR showed that among students from high-risk states, the risk of dropping out of school was 1.84 times higher, and the risk of repeating a grade was 1.68 times higher in 2012. From the null to the full models, both the ICC and the MOR declined due to the inclusion of explanatory covariates, which reduced the unexplained variance at the state level (Steele 2013).

After adjusting for the effects of the independent variables and the state-level effects, Table 6 shows the associations between the students' educational wellbeing

Table 5:
State-level effects from the random intercept logistic regression models during round two

	Students aged 15–18 years in 2012			
	School dropout		**Grade repetition**	
Random effect parameters	**Null Model**	**Full Model**	**Null Model**	**Full Model**
Level 2: States of India				
Variance	0.406	0.246	0.293	0.162
Intraclass correlation coefficient (ICC in %)	10.99	6.95	8.19	4.69
Median odds ratio (MOR)	1.84	1.60	1.68	1.47
Likelihood ratio test significance	***	***	***	***
Number of states	**21**	**21**	**21**	**21**
Number of students	**9822**	**9822**	**9822**	**9822**

Note: (a) The null model is an empty model without any explanatory and control variables. (b) The full model contains all the explanatory and control variables. (c) The school dropout and grade repetition variable is categorised into no, yes. (d) The likelihood ratio test shows the significance of using a multilevel logistic model over a standard logistic model where *** denotes p-value <0.001.
Source: Author's calculation using IHDS data (2012).

Table 6:
Odds ratios from the random-intercept logistic regression models showing the association of the educational wellbeing and demographic and socio-economic characteristics of students in round one with the school dropout and grade repetition status of students in round two

Characteristics	Students aged 15–18 years in 2012			
	School dropout		Grade repetition	
	OR	95% CI	OR	95% CI
Objective educational wellbeing index				
Average and above®				
Below average	1.65*	(1.48–1.85)	1.33*	(1.18–1.49)
Subjective educational wellbeing index				
Average and above®				
Below average	1.30*	(1.16–1.45)	1.17*	(1.04–1.31)
Age of student (in years)				
8®				
9	1.43*	(1.21–1.69)	1.24*	(1.04–1.48)
10	2.07*	(1.79–2.40)	1.57*	(1.34–1.83)
11	2.72*	(2.30–3.23)	1.82*	(1.53–2.18)
Gender of student				
Male®				
Female	1.07	(0.96–1.19)	0.82*	(0.74–0.92)
Type of school attended by student				
Private school®				
Public school	1.75*	(1.47–2.08)	1.42*	(1.20–1.68)
Others	1.64*	(1.23–2.20)	1.43*	(1.07–1.92)
Student receives private tuition				
Yes®				
No	1.41*	(1.17–1.69)	0.96	(0.81–1.13)
Mother's level of education				
More than 10 years of schooling®				
6 to 10 years of schooling	1.85*	(1.06–3.22)	1.82*	(1.22–2.70)
Less than 5 years of schooling	2.71*	(1.55–4.74)	2.28*	(1.51–3.42)
No formal schooling	3.69*	(2.12–6.41)	2.11*	(1.41–3.16)
Mother's working status				
Not working®				
Working	1.11	(0.97–1.26)	1.24*	(1.07–1.43)
Father's level of education				
More than 10 years of schooling®				
6 to 10 years of schooling	1.79*	(1.36–2.35)	1.34*	(1.08–1.68)
Less than 5 years of schooling	2.71*	(2.04–3.60)	1.29*	(1.01–1.66)
No formal schooling	2.61*	(1.96–3.46)	1.15	(0.90–1.48)

Continued

Table 6:
Continued

Characteristics	Students aged 15–18 years in 2012			
	School dropout		Grade repetition	
	OR	95% CI	OR	95% CI
Father's working status				
Not working®				
Working	0.98	(0.87–1.12)	1.01	(0.89–1.16)
Household wealth quintile				
Richest®				
Rich	1.62*	(1.27–2.06)	1.39*	(1.13–1.72)
Middle	2.42*	(1.89–3.11)	1.39*	(1.11–1.75)
Poor	2.61*	(2.00–3.40)	1.35*	(1.05–1.74)
Poorest	3.50*	(2.63–4.65)	1.34*	(1.02–1.78)
Household poverty status				
Not poor®				
Poor	1.27*	(1.12–1.43)	1.09	(0.95–1.25)
Caste of household head				
Scheduled Tribes®				
Scheduled Castes	1.06	(0.86–1.31)	1.04	(0.83–1.29)
Other Backward Classes	1.00	(0.81–1.22)	0.81	(0.65–1.01)
Others	0.92	(0.73–1.15)	0.77*	(0.61–0.98)
Religion of household head				
Hindu®				
Muslim	2.13*	(1.81–2.51)	0.80*	(0.66–0.98)
Others	0.97	(0.74–1.26)	1.03	(0.79–1.35)
Place of residence				
Rural®				
Urban	1.45*	(1.25–1.69)	1.10	(0.94–1.28)
Country regions				
Northern®				
North-eastern	1.69	(0.57–5.03)	0.66	(0.26–1.67)
Central	1.33	(0.60–2.96)	1.78	(0.93–3.43)
Eastern	1.02	(0.55–1.91)	0.64	(0.38–1.09)
Western	2.90*	(1.51–5.59)	0.46*	(0.25–0.84)
Southern	1.01	(0.56–1.84)	0.50*	(0.31–0.82)
Number of states	**21**		**21**	
Number of students	**9,822**		**9,822**	

Note: (a) OR stands for the odds ratio. (b) The 95% confidence interval (CI) is given in brackets. (c) Statistical significance is denoted by asterisks, where * denotes p-value <0.05. (d) ® denotes the reference category. (e) The school dropout and grade repetition variable is categorised into no, yes.
Source: Author's calculation using IHDS data (2005 and 2012).

and other control variables in 2005 and their school dropout and grade repetition status in 2012. The multivariate analysis indicated that compared to students with above-average wellbeing scores, students with below-average objective wellbeing scores in 2005 were 1.65 [CI: 1.48–1.85] times more likely to have dropped out of school and 1.33 [CI: 1.18–1.49] times more likely to have repeated a grade between 2005–2012. Similarly, the odds of dropping out of school [OR: 1.30, CI: 1.16–1.45] and of repeating a grade [OR: 1.17, CI: 1.04–1.31] were found to be higher among students with below-average subjective educational wellbeing scores. Additionally, we observed that compared to students who were attending a private school, students who were attending a public school in 2005 were 1.75 [CI: 1.47–2.08] times more likely to have dropped out of school and were 1.42 [CI: 1.20–1.68] times more likely to have repeated a grade. Moreover, we found that the likelihood of having dropped out of school was 3.69 [CI: 2.12–6.41] times higher among students whose mother had no formal schooling, and was 2.61 [CI: 1.96–3.46] times higher among students whose father had no formal schooling. Similarly, the odds of grade repetition were shown to be higher for students whose mother and father had no formal schooling. In addition, our results showed that compared to students from the richest quintile households, students from the poorest wealth quintile households were 3.50 [CI: 2.63–4.65] times more likely to have dropped out of school and 1.34 [CI: 1.02–1.78] times more likely to have repeated a grade. We also found that compared to students living in rural communities, students living in urban communities were 1.45 [CI: 1.25–1.69] times more likely to have dropped out of school.

Furthermore, we tested the interaction effects to explore whether the effects of the objective and the subjective wellbeing indices on the dropout and grade repetition status of students varied by the mother's level of education and by household wealth (see Tables A.3 and A.4 in the appendix section). The empirical evidence showed that the mother having higher education significantly reduced students' chances of dropping out of school and repeating a grade, particularly among girls (Cardoso and Verner 2006; Lloyd et al. 2009; Marphatia et al. 2019). A potential explanation for this finding is that the mother's education protects her children from having below-average educational wellbeing, and thus lowers their chances of experiencing unfavourable educational outcomes, like dropping out of school or repeating a grade. Moreover, previous evidence has shown that the traits of households in the poor wealth quintile may negatively influence the schooling outcomes of the students in these households (Marphatia et al. 2019; Paul et al. 2021; Pong and Ju 2000). Thus, it is clear that coming from a higher income household may pave the way for students to receive more educational resources, like higher quality schooling and tailored tuition. Such amenities can support the educational wellbeing of children, which can, in turn, further reduce their chances of dropping out of school or repeating a grade. Therefore, we examined the interaction effects of educational wellbeing with the mother's education and household wealth. However, we found that the interaction coefficients were not statistically significant (see Tables A.3 and A.4 in the appendix section), which indicates that the effects of the objective and the subjective wellbeing indices on the dropout and grade repetition status of students did not vary by either the mother's education or the household wealth.

4 Discussion

Using the IHDS panel dataset conducted in 2005 and 2012, we have explored the association between the educational wellbeing of children at ages 8–11 and their likelihood of dropping out or repeating a grade at ages 15–18. Our study provides evidence that Indian children whose objective and subjective educational wellbeing scores were below average during their elementary schooling were more likely to drop out or repeat a grade during their early adolescence. Estimates from our multilevel model indicated that the school dropout and grade retention rates among children varied across states. In addition, we found that while including covariates further decreased the measure of variance in the outcomes, higher unobserved heterogeneity in the school dropout and grade retention rates of children persisted at the state level.

Thus, the analysis clearly showed that the children in our study sample were at greater risk of dropping out or of repeating a grade if they had below-average scores on the educational wellbeing index (both objective and subjective), which included indicators such as measures of a child's school performance, learning skills, cognitive development, involvement in education and enjoyment of education. The results were consistent with the findings of previous scholars, who argued that children's prior academic performance often explains their probability of dropping out of upper secondary education (Falch and Strøm 2008). In line with our findings, previous research has shown that children who have poor school performance and lower learning and cognitive skill levels are at greater risk of failing a grade and of discontinuing their secondary schooling (Fetler 1989). It is worth noting that in the Indian context, the no detention policy for students up to eighth grade has helped to prevent children from being demotivated by their results (UNESCO 2012). However, this policy has also created a situation in which children may fail to learn basic skills (Taneja 2018). Thus, to shield students from the negative consequences of dropping out of school or repeating a grade, more attention should be paid to elementary schooling and the environments in which children spend most of their time. The present study has provided evidence on the association between different educational wellbeing indicators and the dropout and grade repetition rates among students (see Table A.2 in the appendix section). The analysis has shown the importance of these indicators in shaping the future of Indian students, and an awareness of these indicators can help policymakers to remove the barriers are leading some children to have negative educational outcomes. While various initiatives have helped to ensure that most younger Indian children are attending school, the high school dropout and grade repetition rates among secondary school students have led to questions being raised about the quality of the education children are being provided. Both the objective and subjective indicators of wellbeing presented in this study can be used by policymakers to improve the quality of the education Indian students are receiving.

In addition, our study found that school dropout and grade repetition rates were higher among children attending public school than they were among children attending private school. These results are consistent with those of an Indian study that found that public school students were more likely than private school students to drop out because of a shortage of effective public school teachers, and because the parents of public school students were less motivated to support their children's education (Desai et al. 2008b; Kundu 2019). In addition, our analysis showed that the children who did not receive private tuition and who were living in urban areas were more likely to drop out of school during adolescence. This evidence again points to the importance of effective and skilled school teachers, and of parents guiding their children in their lives and careers. It appears that when parents continuously motivate their children to do their best, and to excel in whatever activities they choose, their children have higher levels of achievement (Stegelin 2002). Moreover, it has been observed that rural area parents who are illiterate or did not get the chance to have an education themselves tend to be more enthusiastic about their children receiving an education, and that these positive attitudes have helped to improve the educational systems in rural areas. An Indian study has also identified different determinants that may play a role in the lower levels of education among children experiencing urban poverty (Singh 2013). In line with our findings, a study from rural India showed that children whose parents were illiterate had higher school dropout and grade retention rates (Drèze and Kingdon 2003). These results indicate that the involvement and encouragement of parents play important roles in the educational achievement of their children. Consistent with our findings, several studies have confirmed that a range of factors influence children's school performance, including the socio-economic conditions of the family, the quality of the school facilities, the level of support for learning in the home, and whether the child belongs to a Scheduled Caste or Tribe (Chauhan 2006; Gouda and Sekher 2014; Hunt 2008).

Our observation that the educational wellbeing of young children was associated with their future educational outcomes sheds light on significant shortcomings in India's educational system. With this study, we have sought to strengthen the educational literature for the Indian context by examining the factors associated with high rates of school dropout and grade repetition among secondary school students. Most of the past research on this topic focused on socio-economic conditions, parental characteristics, the school environment and household characteristics as the major determinants of educational outcomes. By contrast, this study is the first to show that children's school performance, learning skills, cognitive development, involvement in education and enjoyment of education during their elementary schooling affected their likelihood of dropping out of school or repeating a grade in secondary school. Much of the previous evidence on these associations was based on cross-sectional studies, which were able to capture only the immediate effects of the factors that may have contributed to the school dropout and grade repetition rates among children. However, by using a panel dataset, we were better able to capture the long-term consequences of educational wellbeing in young children.

Despite these advantages, our study also has some limitations. Our study was not able to establish any causal effects in the association between educational wellbeing and school dropout and grade repetition rates among students. Moreover, previous evidence has shown that the characteristics of the school children attend can affect their educational wellbeing, as their school is the only place other than their home where children routinely spend large amounts of time. Although a few factors related to school characteristics were provided in the study, a three-level random intercept model (in which individuals are nested within schools, which are, in turn, nested within states) would enable us to better control for the unobserved effects of the school environment.

5 Conclusion

The present study has provided conclusive evidence that low levels of both objective and subjective educational wellbeing while in elementary school can negatively affect children's longer-term prospects. It is worth noting that although the government of India had succeeded in encouraging younger children to attend school, many of these children are failing to acquire basic skills. Although India's educational policy is focused on the basic development of children, the educational wellbeing of elementary school children is often overlooked. Thus, the results of the current study indicate that policymakers should pay more attention to the experiences of elementary school students, as many of these children have low levels of educational wellbeing. We also recommend that the existing educational system be modernised to focus on the process of learning, rather than on educational outcomes only.

Acknowledgements

We would like to thank the two anonymous reviewers and the editors for their invaluable comments, which significantly improved this paper. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

References

Acock, A. C. 2013. *Discovering structural equation modeling using Stata.* College Station, TX: Stata Press Books.

Beauvais, F., E. L. Chavez, E. R. Oetting, J. L. Deffenbacher and G. R. Cornell 1996. Drug use, violence, and victimization among White American, Mexican American, and American Indian dropouts, students with academic problems, and students in good academic standing. *Journal of Counseling Psychology* 43(3): 292–299. https://doi.org/10.1037/0022-0167.43.3.292

Bhat, P. N. M. and A. J. F. Zavier 2005. Role of religion in fertility decline: The case of Indian muslims. *Economic and Political Weekly* 40(5): 385–402.

Bornstein, M. H., L. Davidson, C. L. M. Keyes and K. A. Moore (eds) 2003. *Well-being: Positive development across the life course*. Psychology Press.

Cardoso, A. R. and D. Verner 2006. School drop-out and push-out factors in Brazil: The role of early parenthood, child labor, and poverty. IZA Discussion Paper 2515. Bonn: Institute for the Study of Labor (IZA).

Chauhan, C. 2006. SC/ST Dropout rates high. Hindustan Times. 22 October 2006.

Choudhury, A. 2006. Revisiting dropouts: Old issues, fresh perspectives. *Economic and Political Weekly* 41(51): 5257–5263.

Chudgar, A. 2009. The challenge of universal elementary education in rural India: Can adult literacy play a role? *Comparative Education Review* 53(3): 403–433. https://doi.org/10.1086/599572

Costello, A. B. and J. Osborne 2005. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation* 10(1): 7. https://doi.org/10.7275/jyj1-4868

Darling-Hammond, L., L. Flook, C. Cook-Harvey, B. Barron and D. Osher 2020. Implications for educational practice of the science of learning and development. *Applied Developmental Science* 24(2): 97–140. https://doi.org/10.1080/10888691.2018.1537791

Desai, S., A. Dubey, B. L. Joshi, M. Sen, A. Sharif and R. Vanneman. 2010. India human development survey users' guide release 03. University of Maryland and National Council of Applied Economic Research, New Delhi. https://www.icpsr.umich.edu/web/pages/DSDR/idhs-data-guide.html, accessed 23 December 2019.

Desai, S., A. Dubey and R. Vanneman 2015. India human development survey-II users' guide release 01. University of Maryland and National Council of Applied Economic Research, New Delhi. https://www.icpsr.umich.edu/web/pages/DSDR/idhs-II-data-guide.html, accessed 23 December 2019.

Desai, S. and R. Vanneman 2015. India human development survey-II (IHDS-II): 2011–12: Version 6. Inter-University Consortium for Political and Social Research. https://www.icpsr.umich.edu/web/DSDR/studies/36151/versions/V6, accessed 23 December 2019.

Desai, S., R. Vanneman and New Delhi National Council of Applied Economic Research. 2008a. India Human Development Survey (IHDS): 2005: Version 12. Inter-University Consortium for Political and Social Research. https://www.icpsr.umich.edu/web/DSDR/studies/22626, accessed 23 December 2019.

Desai, S., A. Dubey, R. Vanneman and R. Banerji 2008b. *Private schooling in India: A new educational landscape*. Working Paper No. 11. India Human Development Survey.

Desai, U. 1991. Determinants of educational performance in India: Role of home and family. *International Review of Education* 37(2): 245–65. https://doi.org/10.1007/BF00599682

Dongre, A. A. and V. Tewary 2015. Impact of private tutoring on learning levels: Evidence from India. SSRN. https://doi.org/10.2139/ssrn.2401475

Drèze, J. and G. G. Kingdon 2003. School participation in Rural India. *Review of Development Economics* 5(1): 1–24. https://doi.org/10.1111/1467-9361.00103

Ender, P. 2010. Collin: Stata command to compute collinearity diagnostics.

Fabrigar, L. R., D. T. Wegener, R. C. MacCallum and E. J. Strahan 1999. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods* 4(3): 272. https://doi.org/10.1037/1082-989X.4.3.272

Falch, T. and B. Strøm 2008. *Student progression in upper secondary education: The effect of academic ability, gender, and schools*. Working Paper 9708, Trondheim: Department of Economics, Norwegian University of Science and Technology.

Fetler, M. 1989. School dropout rates, academic performance, size, and poverty: Correlates of educational reform. *Educational Evaluation and Policy Analysis* 11(2): 109–16. https://doi.org/10.3102/01623737011002109

Filmer, D. and K. Scott 2008. *Assessing asset indices*. The World Bank.

Glick, P. and D. E. Sahn 2010. Early academic performance, grade repetition, and school attainment in Senegal: A panel data analysis. *The World Bank Economic Review* 24(1): 93–120. https://doi.org/10.1093/wber/lhp023

Gouda, J., K. Chandra Das, S. Goli and L. Maikho Apollo Pou 2013. Government versus private primary schools in India. *International Journal of Sociology and Social Policy* 3(11/12): 708–724. https://doi.org/10.1108/IJSSP-12-2012-0105

Gouda, S. M. and T. V. Sekher 2014. Factors leading to school dropouts in India: An analysis of National Family Health Survey-3 data. *IOSR Journal of Research & Method in Education* 4(6): 75–83. http://doi.org/10.9790/7388-04637583

Govinda, R. and M. Bandyopadhyay 2008. *Access to elementary education in India. CREATE Country Analytic Review*. Delhi/Brighton: NUEPA/University of Sussex.

Haidary, A. 2013. *Controvercy over grade repetition: Afghan teachers' view on grade repetition*. Thesis, Karlstad University.

Hunt, F. M. 2008. Dropping out from school: A cross country review of literature. CREATE Pathways to Access, Research Monographs No. 16. Delhi/Brighton: NUEPA/University of Sussex.

Jimerson, S. R. 2001. Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review* 30(3): 420–437. https://doi.org/10.1080/02796015.2001.12086124

Kline, P. 2000. Reliability of tests: Practical issues. In *Handbook of psychological testing*, ed P. Kline, 7–17. Second Edition. London: Routledge.

Kundu, P. 2019. Deteriorating quality of education in schools: Are teachers responsible? Economic and Political Weekly. *Economic & Political Weekly* 54(24).

Lloyd, C. B., C. Mete and M. J. Grant 2009. The implications of changing educational and family circumstances for children's grade progression in rural Pakistan: 1997–2004. *Economics of Education Review* 28(1): 152–160. https://doi.org/10.1016/j.econedurev.2007.04.005

Marphatia, A. A., A. M. Reid and C. S. Yajnik 2019. Developmental origins of secondary school dropout in rural India and its differential consequences by sex: A biosocial life course analysis. *International Journal of Educational Development* 66: 8–23. https://doi.org/10.1016/j.ijedudev.2018.12.001

McGrath, H. 2006. To repeat, or not to repeat. *Words* 26(2): 39–46.

Merlo, J., B. Chaix, H. Ohlsson, A. Beckman, K. Johnell, P. Hjerpe, L. Råstam and K. Larsen 2006. A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology & Community Health* 60(4): 290–297.

MHRD. 2020. National education policy 2020. Government of India.

Muttarak, R. and T. Chankrajang 2015. Who is concerned about and takes action on climate change? Gender and education divides among Thais. *Vienna Yearbook of Population Research* 13: 193–220. https://doi.org/10.1553/populationyearbook2015s193

OECD 2003. *Financing education: Investments and returns: Analysis of the world education indicators*. vol. 3. OECD Publishing.

Paul, R., R. Rashmi and S. Srivastava 2021. Does lack of parental involvement affect school dropout among Indian adolescents? Evidence from a panel study. *PLoS ONE* 16(5): e0251520. https://doi.org/10.1371/journal.pone.0251520

Pong, S.-L. and D.-B. Ju 2000. The effects of change in family structure and income on dropping out of middle and high school. *Journal of Family Issues* 21(2): 147–169. https://doi.org/10.1177/019251300021002001

Pratinidhi, A. K., S. V. Warerkar and S. G. Garad 1992. A study of school dropouts in an urban slum community. *Demography India* 21(2): 301–305.

Rao, M. J. M. 2000. Migration of labour and school dropouts. *Social Welfare-Delhi* 47(6): 26–31.

Rutstein, S. O. and K. Johnson 2004. The DHS Wealth Index. DHS Comparative Reports No. 6. Calverton: ORC Macro.

Sabates, R., J. Westbrook, K. Akyeampong and F. Hunt. 2010. School drop out: Patterns, causes, changes and policies. issues. background paper for the education for all global monitoring report 2011, The hidden crisis: Armed conflict and education. UNESCO.

Sanjay, A. 2020. The creation and mobilization of anti-China sentiment by interest groups in Indian society (2012–2018). CURA Paper No. 1. Qatar: Center for International and Regional Studies. Georgetown University in Qatar.

Sethi, I. K. 2018. *Parental aspirations and educational outcomes: Evidence from Andhra Pradesh, India*. Thesis, University of Oxford.

Singh, B. G. 2013. A study of the reasons of urban deprived children repeatedly becoming out-of-school even after their mainstreaming in the regular schooling system. Research Report. School of Studies in Psychology. Pt. Ravishankar Shukla University.

Singh, R., R. Kesarwani and P. Mukherjee 2019. *Exploring well-being among 22-year-old youth in India*. Working Paper 184. Oxford: Young Lives, Oxford Department of International Development (ODID): University of Oxford.

Singhal, K. and U. Das 2019. Revisiting the role of private schooling on children learning outcomes: Evidence from rural India. *South Asia Economic Journal* 20(2): 274–302. https://doi.org/10.1177/1391561419853678

StataCorp, L. P. 2014. Stata 14. College Station, TX: StataCorp LP.

Steele, F. 2013. Module 10: Single-level and multilevel models for nominal responses: MLwiN Practical. Centre for Multilevel Modelling.

Stegelin, D. A. 2002. *Early literacy education: First steps toward dropout prevention. Effective strategies for school improvement and dropout prevention*. Clemson, SC: National Dropout Prevention Center.

Sylva, K. 1994. School influences on children's development. *Journal of Child Psychology and Psychiatry* 35(1): 135–70. https://doi.org/10.1111/j.1469-7610.1994.tb01135.x

Taneja, A. 2018. *No detention under the RTE Act: The policy options*. Education Policy Brief New Delhi: CARE India.

UNESCO. 2012. Global education digest 2012. Opportunities lost: The impact of grade repetition and early school leaving. UNESCO Institute for Statistics Quebec.

Upendranath, C. 1995. Education of girls in India: The daunting task ahead. *Journal of Educational Planning and Administration* 9: 81–92.

Wat, A. 2015. *Transforming the workforce for children birth through age 8: A unifying foundation*. Washington, DC: The National Academies Press. https://doi.org/10.17226/19401

Appendix

Table A.1:
Distribution of students by variables used for preparing the educational wellbeing index

Characteristics	*N*	%
Type of student		
Below average	982	10.0
Average	7,645	77.8
Above average	1,195	12.2
Student enjoys school		
No	525	5.3
Yes	9,297	94.7
Reading skills		
Cannot read letters	592	6.0
Can read alphabets	932	9.5
Can read words	1,467	14.9
Can read a short paragraph	4,066	41.4
Can read a full story	2,765	28.2
Mathematical skills		
Cannot recognise numbers	1,159	11.8
Can recognise numbers	2,326	23.7
Can do subtraction	4,558	46.4
Can do division	1,779	18.1
Writing skills		
Cannot write	2,161	22.0
Can write	7,661	78.0
Overall	**9,822**	**100**

Table A.2:
Odds ratios from the random-intercept logistic regression models showing the association of the indicators of educational wellbeing in round one with the school dropout and grade repetition status of students in round two

	Students aged 15–18 years in 2012			
	School dropout		Grade repetition	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**
Type of student				
Above average®				
Average	1.19	(0.98–1.44)	1.11	(0.91–1.34)
Below average	1.54*	(1.21–1.97)	1.26	(0.98–1.62)
Student enjoys school				
Yes®				
No	1.06	(0.85–1.32)	1.08	(0.85–1.37)
Reading skills				
Can read story®				
Can read a paragraph	1.34*	(1.14–1.58)	1.06	(0.90–1.24)
Can read words	1.29*	(1.05–1.59)	1.13	(0.91–1.40)
Can read alphabets	1.73*	(1.36–2.19)	1.06	(0.82–1.37)
Cannot read letters	1.48*	(1.10–1.99)	0.68*	(0.48–0.96)
Mathematical skills				
Can do division®				
Can do subtraction	1.17	(0.96–1.44)	1.32*	(1.09–1.61)
Can recognise numbers	1.49*	(1.19–1.87)	1.31*	(1.04–1.64)
Cannot recognise numbers	1.44*	(1.08–1.90)	1.44*	(1.07–1.92)
Writing skills				
Can write®				
Cannot write	1.08	(0.92–1.27)	1.28*	(1.08–1.53)
Age of student (in years)				
8®				
9	1.48*	(1.25–1.74)	1.24*	(1.04–1.48)
10	2.17*	(1.87–2.52)	1.59*	(1.36–1.86)
11	2.89*	(2.44–3.44)	1.86*	(1.56–2.23)
Gender of student				
Male®				
Female	1.06	(0.95–1.17)	0.82*	(0.73–0.92)
Type of school attended by student				
Private school®				
Public school	1.70*	(1.42–2.02)	1.42*	(1.19–1.68)
Others	1.58*	(1.18–2.12)	1.42*	(1.06–1.91)

Continued

Table A.2:
Continued

	Students aged 15–18 years in 2012			
	School dropout		Grade repetition	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**
Student receives private tuition				
Yes®				
No	1.38*	(1.15–1.66)	0.95	(0.80–1.13)
Mother's level of education				
More than 10 years of schooling®				
6 to 10 years of schooling	1.86*	(1.06–3.24)	1.81*	(1.22–2.68)
Less than 5 years of schooling	2.68*	(1.53–4.69)	2.24*	(1.49–3.37)
No formal schooling	3.61*	(2.08–6.29)	2.07*	(1.38–3.10)
Mother's working status				
Not working®				
Working	1.11	(0.97–1.27)	1.24*	(1.07–1.43)
Father's level of education				
More than 10 years of schooling®				
6 to 10 years of schooling	1.77*	(1.35–2.32)	1.33*	(1.07–1.66)
Less than 5 years of schooling	2.63*	(1.98–3.50)	1.26	(0.98–1.62)
No formal schooling	2.51*	(1.89–3.34)	1.13	(0.88–1.45)
Father's working status				
Not working®				
Working	0.97	(0.85–1.10)	1.00	(0.88–1.15)
Household wealth quintile				
Richest®				
Rich	1.62*	(1.27–2.06)	1.39*	(1.13–1.72)
Middle	2.41*	(1.88–3.10)	1.40*	(1.11–1.76)
Poor	2.60*	(1.99–3.39)	1.36*	(1.06–1.75)
Poorest	3.44*	(2.59–4.58)	1.36*	(1.03–1.79)
Household poverty status				
Not poor®				
Poor	1.25*	(1.11–1.41)	1.08	(0.95–1.24)
Caste of household head				
Scheduled Tribes®				
Scheduled Castes	1.05	(0.85–1.29)	1.04	(0.83–1.30)
Other Backward Classes	0.99	(0.81–1.21)	0.82	(0.66–1.02)
Others	0.92	(0.73–1.15)	0.78*	(0.62–0.99)
Religion of household head				
Hindu®				
Muslim	2.10*	(1.78–2.48)	0.79*	(0.65–0.96)
Others	0.96	(0.73–1.25)	1.03	(0.79–1.35)

Continued

Table A.2: Continued

Characteristics	Students aged 15–18 years in 2012			
	School dropout		Grade repetition	
	OR	95% CI	OR	95% CI
Place of residence				
Rural®				
Urban	1.45*	(1.25–1.69)	1.10	(0.94–1.28)
Country regions				
Northern®				
North-eastern	1.74	(0.59–5.13)	0.69	(0.27–1.73)
Central	1.30	(0.59–2.87)	1.70	(0.88–3.26)
Eastern	1.03	(0.55–1.91)	0.65	(0.38–1.11)
Western	2.95*	(1.54–5.67)	0.46*	(0.25–0.84)
Southern	1.02	(0.56–1.84)	0.50*	(0.31–0.82)
Number of states	**21**		**21**	
Number of students	**9,822**		**9,822**	

Note: (a) OR stands for the odds ratio. (b) The 95% confidence interval (CI) is given in brackets. (c) Statistical significance is denoted by asterisks where * denotes p-value <0.05. (d) ® denotes the reference category. (e) School dropout and grade repetition variable categorised into no, yes.

Table A.3:
Random-intercept logistic regression models showing the association of the interaction effects of the educational wellbeing variables by the mother's level of education and the household wealth quintile in round one with the school dropout status of the students in round two

	School dropout among students aged 15–18 years in 2012							
	Model-I		Model-II		Model-III		Model-IV	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Objective educational wellbeing index								
Average and above®								
Below average	1.81	(0.55–5.97)	1.74*	(1.15–2.63)	1.66*	(1.48–1.85)	1.65*	(1.48–1.84)
Subjective educational wellbeing index								
Average and above®								
Below average	1.29*	(1.16–1.44)	1.30*	(1.16–1.45)	0.42	(0.12–1.53)	1.45	(0.98–2.15)
Age of student (in years)								
8®								
9	1.43*	(1.22–1.69)	1.44*	(1.22–1.69)	1.43*	(1.21–1.69)	1.44*	(1.22–1.70)
10	2.07*	(1.79–2.40)	2.08*	(1.79–2.41)	2.07*	(1.79–2.40)	2.08*	(1.79–2.41)
11	2.73*	(2.30–3.23)	2.72*	(2.30–3.23)	2.73*	(2.30–3.23)	2.74*	(2.31–3.25)
Gender of student								
Male®								
Female	1.07	(0.96–1.19)	1.07	(0.96–1.19)	1.07	(0.96–1.19)	1.07	(0.97–1.19)
Type of school attended by student								
Private school®								
Public school	1.75*	(1.47–2.08)	1.75*	(1.47–2.08)	1.75*	(1.47–2.08)	1.75*	(1.47–2.08)
Others	1.65*	(1.23–2.20)	1.64*	(1.23–2.20)	1.64*	(1.22–2.19)	1.65*	(1.23–2.20)

Continued

Table A.3:
Continued

	School dropout among students aged 15–18 years in 2012							
	Model-I		Model-II		Model-III		Model-IV	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Student receives private tuition								
Yes®								
No	1.40*	(1.17–1.68)	1.41*	(1.17–1.69)	1.41*	(1.17–1.69)	1.41*	(1.17–1.69)
Mother's level of education								
More than 10 years of schooling®								
6 to 10 years of schooling	1.80	(0.94–3.45)	1.84*	(1.05–3.20)	1.31	(0.69–2.48)	1.85*	(1.06–3.23)
Less than 5 years of schooling	2.73*	(1.42–5.24)	2.69*	(1.54–4.70)	1.97*	(1.04–3.73)	2.69*	(1.54–4.71)
No formal schooling	3.85*	(2.03–7.30)	3.66*	(2.11–6.37)	2.57*	(1.38–4.78)	3.68*	(2.12–6.39)
Mother's working status								
Not working®								
Working	1.11	(0.97–1.27)	1.11	(0.97–1.27)	1.11	(0.97–1.26)	1.11	(0.97–1.26)
Father's level of education								
More than 10 years of schooling®								
6 to 10 years of schooling	1.79*	(1.36–2.34)	1.78*	(1.36–2.34)	1.80*	(1.37–2.37)	1.78*	(1.36–2.34)
Less than 5 years of schooling	2.70*	(2.03–3.59)	2.69*	(2.03–3.58)	2.73*	(2.05–3.63)	2.71*	(2.03–3.60)
No formal schooling	2.60*	(1.96–3.46)	2.59*	(1.95–3.44)	2.62*	(1.98–3.49)	2.60*	(1.96–3.45)
Father's working status								
Not working®								
Working	0.98	(0.87–1.11)	0.98	(0.87–1.12)	0.98	(0.87–1.12)	0.98	(0.87–1.12)

Continued

Table A.3:
Continued

	School dropout among students aged 15–18 years in 2012							
	Model-I		Model-II		Model-III		Model-IV	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Household wealth quintile								
Richest®								
Rich	1.61*	(1.26–2.05)	1.57*	(1.15–2.14)	1.62*	(1.27–2.07)	1.54*	(1.10–2.16)
Middle	2.41*	(1.87–3.09)	2.47*	(1.81–3.38)	2.42*	(1.89–3.11)	2.81*	(2.01–3.93)
Poor	2.60*	(1.99–3.39)	2.92*	(2.10–4.05)	2.61*	(2.00–3.41)	2.63*	(1.85–3.73)
Poorest	3.48*	(2.62–4.63)	3.41*	(2.41–4.84)	3.50*	(2.63–4.65)	3.88*	(2.69–5.61)
Household poverty status								
Not poor®								
Poor	1.27*	(1.12–1.43)	1.27*	(1.12–1.43)	1.27*	(1.12–1.43)	1.27*	(1.12–1.43)
Caste of household head								
Scheduled Tribes®								
Scheduled Castes	1.06	(0.86–1.31)	1.06	(0.86–1.31)	1.06	(0.86–1.32)	1.06	(0.86–1.32)
Other Backward Classes	1.00	(0.81–1.22)	1.00	(0.81–1.22)	1.00	(0.81–1.22)	1.00	(0.81–1.22)
Others	0.92	(0.73–1.16)	0.92	(0.73–1.15)	0.92	(0.73–1.15)	0.92	(0.73–1.16)
Religion of household head								
Hindu®								
Muslim	2.12*	(1.80–2.51)	2.13*	(1.81–2.51)	2.13*	(1.81–2.51)	2.13*	(1.80–2.51)
Others	0.96	(0.74–1.26)	0.97	(0.74–1.26)	0.97	(0.74–1.26)	0.97	(0.74–1.26)
Place of residence								
Rural®								
Urban	1.45*	(1.25–1.69)	1.45*	(1.25–1.69)	1.45*	(1.25–1.69)	1.46*	(1.25–1.70)

Continued

Table A.3:
Continued

	School dropout among students aged 15–18 years in 2012							
	Model-I		Model-II		Model-III		Model-IV	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Country regions								
Northern®								
North-eastern	1.69	(0.57–5.04)	1.68	(0.56–4.99)	1.69	(0.57–5.02)	1.68	(0.56–5.03)
Central	1.33	(0.60–2.96)	1.33	(0.60–2.95)	1.34	(0.60–2.96)	1.33	(0.60–2.96)
Eastern	1.02	(0.55–1.91)	1.02	(0.54–1.90)	1.02	(0.55–1.91)	1.02	(0.55–1.92)
Western	2.89*	(1.50–5.56)	2.90*	(1.51–5.58)	2.90*	(1.50–5.57)	2.88*	(1.49–5.56)
Southern	1.01	(0.56–1.84)	1.01	(0.56–1.84)	1.01	(0.56–1.84)	1.01	(0.56–1.85)
Objective Wellbeing#Mother's Education								
Below average#More than 10 years of schooling®								
Below average#6 to 10 years of schooling	1.04	(0.30–3.55)						
Below average#Less than 5 years of schooling	0.95	(0.28–3.23)						
Below average#No formal schooling	0.89	(0.27–2.94)						
Objective Wellbeing#Wealth Quintile								
Below average#Richest®								
Below average#Rich			1.05	(0.65–1.70)				
Below average#Middle			0.95	(0.60–1.50)				
Below average#Poor			0.81	(0.51–1.28)				
Below average#Poorest			1.02	(0.64–1.61)				

Continued

Table A.3:
Continued

	School dropout among students aged 15–18 years in 2012							
	Model-I		Model-II		Model-III		Model-IV	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Subjective Wellbeing#Mother's Education								
Below average#More than 10 years of schooling®								
Below average#6 to 10 years of schooling					3.07	(0.81–11.6)		
Below average#Less than 5 years of schooling					2.93	(0.78–11.0)		
Below average#No formal schooling					3.16	(0.86–11.6)		
Subjective Wellbeing#Wealth Quintile								
Below average#Richest®								
Below average#Rich							1.08	(0.68–1.72)
Below average#Middle							0.75	(0.48–1.18)
Below average#Poor							0.97	(0.62–1.52)
Below average#Poorest							0.83	(0.53–1.29)
Number of states	**21**		**21**		**21**		**21**	
Number of students	**9,822**		**9,822**		**9,822**		**9,822**	

Note: (a) OR stands for the odds ratio. (b) The 95% confidence interval (CI) is given in brackets. (c) Statistical significance is denoted by asterisks where * denotes p-value <0.05. (d) ® denotes the reference category. (e) Models I, II, III and IV contain interaction terms between objective wellbeing-mother's education, objective wellbeing-wealth quintile, subjective wellbeing-mother's education and subjective wellbeing-wealth quintile, respectively.

Table A.4:
Random-intercept logistic regression models showing the association of the interaction effects of the educational wellbeing variables by the mother's level of education and the household wealth quintile in round one with the grade repetition status of the students in round two

	Grade repetition among students aged 15–18 years in 2012							
	Model-I		Model-II		Model-III		Model-IV	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Objective educational wellbeing index								
Average and above®								
Below average	1.61	(0.70–3.74)	2.08*	(1.48–2.90)	1.32*	(1.17–1.49)	1.33*	(1.18–1.49)
Subjective educational wellbeing index								
Average and above®								
Below average	1.16*	(1.04–1.31)	1.16*	(1.03–1.30)	1.00	(0.48–2.10)	1.09	(0.79–1.51)
Age of student (in years)								
8®								
9	1.24*	(1.05–1.48)	1.24*	(1.05–1.48)	1.24*	(1.05–1.48)	1.24*	(1.04–1.47)
10	1.57*	(1.34–1.84)	1.57*	(1.35–1.84)	1.57*	(1.34–1.84)	1.57*	(1.34–1.83)
11	1.83*	(1.53–2.18)	1.83*	(1.53–2.19)	1.83*	(1.53–2.18)	1.82*	(1.52–2.17)
Gender of student								
Male®								
Female	0.82*	(0.74–0.92)	0.83*	(0.74–0.92)	0.83*	(0.74–0.92)	0.82*	(0.74–0.92)
Type of school attended by student								
Private school®								
Public school	1.42*	(1.20–1.68)	1.42*	(1.20–1.68)	1.42*	(1.20–1.68)	1.42*	(1.20–1.68)
Others	1.43*	(1.07–1.93)	1.43*	(1.06–1.92)	1.43*	(1.06–1.92)	1.43*	(1.06–1.92)

Continued

Table A.4:
Continued

	Grade repetition among students aged 15–18 years in 2012							
	Model-I		Model-II		Model-III		Model-IV	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Student receives private tuition								
Yes®								
No	0.95	(0.80–1.13)	0.95	(0.80–1.13)	0.96	(0.81–1.14)	0.96	(0.81–1.14)
Mother's level of education								
More than 10 years of schooling®								
6 to 10 years of schooling	1.85*	(1.18–2.91)	1.77*	(1.19–2.64)	1.61	(0.98–2.64)	1.81*	(1.22–2.69)
Less than 5 years of schooling	2.38*	(1.49–3.80)	2.19*	(1.45–3.30)	2.06*	(1.24–3.42)	2.27*	(1.51–3.42)
No formal schooling	2.25*	(1.42–3.56)	2.04*	(1.36–3.06)	2.05*	(1.25–3.36)	2.10*	(1.40–3.16)
Mother's working status								
Not working®								
Working	1.24*	(1.08–1.43)	1.24*	(1.08–1.43)	1.24*	(1.07–1.43)	1.24*	(1.07–1.43)
Father's level of education								
More than 10 years of schooling®								
6 to 10 years of schooling	1.34*	(1.07–1.67)	1.33*	(1.07–1.66)	1.35*	(1.08–1.68)	1.35*	(1.08–1.68)
Less than 5 years of schooling	1.29*	(1.00–1.65)	1.28	(1.00–1.64)	1.29*	(1.01–1.66)	1.29*	(1.01–1.66)
No formal schooling	1.15	(0.90–1.48)	1.14	(0.89–1.46)	1.16	(0.90–1.49)	1.16	(0.90–1.48)
Father's working status								
Not working®								
Working	1.01	(0.89–1.16)	1.01	(0.89–1.16)	1.01	(0.89–1.15)	1.01	(0.89–1.15)

Continued

Table A.4:
Continued

	Grade repetition among students aged 15–18 years in 2012							
	Model-I		Model-II		Model-III		Model-IV	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Household wealth quintile								
Richest®								
Rich	1.39*	(1.12–1.72)	1.49*	(1.14–1.94)	1.39*	(1.13–1.72)	1.37*	(1.04–1.80)
Middle	1.38*	(1.10–1.74)	1.72*	(1.30–2.28)	1.39*	(1.10–1.75)	1.27	(0.95–1.70)
Poor	1.35*	(1.05–1.74)	1.62*	(1.19–2.21)	1.35*	(1.05–1.74)	1.40*	(1.03–1.91)
Poorest	1.34*	(1.02–1.77)	1.78*	(1.28–2.49)	1.35*	(1.02–1.78)	1.28	(0.91–1.80)
Household poverty status								
Not poor®								
Poor	1.09	(0.95–1.25)	1.09	(0.96–1.25)	1.09	(0.96–1.25)	1.09	(0.95–1.25)
Caste of household head								
Scheduled Tribes®								
Scheduled Castes	1.04	(0.83–1.29)	1.03	(0.82–1.28)	1.04	(0.83–1.30)	1.04	(0.83–1.29)
Other Backward Classes	0.81	(0.65–1.01)	0.81	(0.65–1.01)	0.81	(0.65–1.01)	0.81	(0.65–1.01)
Others	0.78*	(0.61–0.99)	0.78*	(0.61–0.98)	0.78*	(0.61–0.99)	0.77*	(0.61–0.98)
Religion of household head								
Hindu®								
Muslim	0.80*	(0.66–0.97)	0.79*	(0.65–0.96)	0.80*	(0.66–0.98)	0.80*	(0.66–0.98)
Others	1.03	(0.79–1.35)	1.04	(0.79–1.36)	1.03	(0.79–1.35)	1.03	(0.79–1.35)
Place of residence								
Rural®								
Urban	1.10	(0.94–1.28)	1.10	(0.94–1.28)	1.09	(0.94–1.28)	1.09	(0.94–1.28)

Continued

Table A.4:
Continued

	Grade repetition among students aged 15–18 years in 2012							
	Model-I		Model-II		Model-III		Model-IV	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Country regions								
Northern®								
North-eastern	0.67	(0.26–1.67)	0.66	(0.26–1.66)	0.66	(0.26–1.65)	0.67	(0.26–1.67)
Central	1.78	(0.93–3.42)	1.76	(0.92–3.38)	1.78	(0.93–3.41)	1.78	(0.92–3.42)
Eastern	0.64	(0.38–1.09)	0.64	(0.38–1.08)	0.64	(0.38–1.09)	0.64	(0.38–1.09)
Western	0.46*	(0.25–0.84)	0.45*	(0.25–0.82)	0.46*	(0.25–0.84)	0.46*	(0.25–0.84)
Southern	0.50*	(0.31–0.82)	0.50*	(0.30–0.81)	0.50*	(0.31–0.82)	0.50*	(0.31–0.82)
Objective Wellbeing#Mother's Education								
Below average#More than 10 years of schooling®								
Below average#6 to 10 years of schooling	0.90	(0.37–2.17)						
Below average#Less than 5 years of schooling	0.83	(0.34–1.99)						
Below average#No formal schooling	0.80	(0.34–1.87)						
Objective Wellbeing#Wealth Quintile								
Below average#Richest®								
Below average#Rich			0.78	(0.51–1.18)				
Below average#Middle			0.56*	(0.37–0.84)				
Below average#Poor			0.61*	(0.40–0.91)				
Below average#Poorest			0.52*	(0.34–0.78)				

Continued

Table A.4:
Continued

	Grade repetition among students aged 15–18 years in 2012							
	Model-I		**Model-II**		**Model-III**		**Model-IV**	
Characteristics	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Subjective Wellbeing#Mother's Education								
Below average#More than 10 years of schooling®								
Below average#6 to 10 years of schooling					1.33	(0.61–2.92)		
Below average#Less than 5 years of schooling					1.27	(0.58–2.77)		
Below average#No formal schooling					1.10	(0.52–2.34)		
Subjective Wellbeing#Wealth Quintile								
Below average#Richest®								
Below average#Rich							1.05	(0.71–1.57)
Below average#Middle							1.21	(0.81–1.79)
Below average#Poor							0.95	(0.64–1.42)
Below average#Poorest							1.11	(0.75–1.66)
Number of states	**21**		**21**		**21**		**21**	
Number of students	**9,822**		**9,822**		**9,822**		**9,822**	

Note: (a) OR stands for the odds ratio. (b) The 95% confidence interval (CI) is given in brackets. (c) Statistical significance is denoted by asterisks where * denotes p-value <0.05. (d) ® denotes the reference category. (e) Models I, II, III and IV contain interaction terms between objective wellbeing-mother's education, objective wellbeing-wealth quintile, subjective wellbeing-mother's education and subjective wellbeing-wealth quintile, respectively.

DATA & TRENDS

Vienna Yearbook of Population Research 2021 (Vol. 19), pp. 547–583

Years of Good Life: An illustration of a new well-being indicator using data for Thailand

***Thananon Buathong*[1], *Anna Dimitrova*[2,*], *Paolo Miguel M. Vicerra*[3] *and Montakarn Chimmamee*[4]**

Abstract

While Thailand has achieved high levels of economic growth in recent decades, poverty at the local level has been increasing. Indicators of human development at the national level often mask the differences in well-being across communities. When responding to the need for sustainable development research, the heterogeneity of a population should be emphasised to ensure that no one is left behind. The Years of Good Life (YoGL) is a well-being indicator that demonstrates the similarities and differences between subpopulations in a given sociocultural context over time. The data used in this analysis were collected from Chiang Rai and Kalasin, which are provinces located in regions of Thailand with high poverty rates. Our main results indicate that the remaining years of good life (free from physical and cognitive limitations, out of poverty and satisfied with life) at age 20 among the sample population were 26 years for women and 28 years for men. The results varied depending on the indicators applied in each dimension of YoGL. Our analysis of the YoGL constituents indicated that cognitive functioning was the dimension that decreased the years of good life the most in the main specification. This study demonstrates the applicability of the YoGL methodology in investigating the well-being of subpopulations.

Keywords: well-being; Thailand; survey design; data collection

[1]Faculty of Sociology and Anthropology, Thammasat University, Thailand

[2]Scripps Institution of Oceanography, University of California, San Diego

[3]College of Population Studies, Chulalongkorn University, Thailand

[4]Social Research Institute, Chulalongkorn University, Thailand

[*]Correspondence to: Anna Dimitrova, adimitrova@ucsd.edu

DOI: 10.1553/populationyearbook2021.dat.1

1 Introduction

Thailand has experienced strong economic growth in recent decades. Indeed, by 2011, Thailand had gained the status of an upper-middle income country (Gil Sander and Burgard 2011). In 2019, the gross domestic product (GDP) of Thailand amounted to US$543.7 billion, which made it the second-largest economy in Southeast Asia, and the eighth-largest economy in Asia (World Bank 2019). Despite these positive changes in the national economy, gains in the levels of human development in Thailand have been inconsistent over time and across population subgroups. Moreover, the levels of development in different locations of the country have varied (Yang et al. 2020). In particular, the provinces in the north and north-eastern regions are notably poorer than other parts of the country. For example, from 2011 to 2013, Bangkok experienced annual economic growth of 10%, while the north and north-eastern regions had growth rates below 5%. In addition, while consumption growth has been increasing in the poorer regions, income growth has remained unchanged (Yang et al. 2020). When assessing growth and development in a given society, GDP offers some information, but it has limitations, as it only measures national aggregates, while failing to account for non-economic determinants of human development (Stiglitz et al. 2010).

Over the past 50 years, a wide range of well-being indices have been introduced to assist policymakers seeking to improve the quality of human life and contribute to the broader discussion of what constitutes sustainable development. These indices include the Better Life Index from the Organization for Economic Co-operation and Development (OECD), the Ecological Footprint (EF), the Global Well-Being Index (GWI), the Happy Planet Index (HPI), the Human Development Index (HDI) and the Sustainable Society Index (SSI), to mention a few (Mclean 2014; Strezov et al. 2017). This wide variety of indices points to differences in the understanding of quality of life based on the country context and/or the agency's objectives (Mclean 2014), and signals the need for a more holistic approach to studying well-being.

Most of these indices look beyond the measurement of GDP, and devote more attention to a broader spectrum of social and ecological issues, including the sustainability of social and natural capital and good governance (i.e., OECD Better Life Index, Social Progress Index, Happy Planet Index). The OECD Better Life Index, for example, integrates multiple dimensions of well-being through an interactive online interface, which allows the user to choose from 11 domains, ranging from current conditions in housing to life satisfaction and work-life balance. Although these indices cover dimensions relevant for human well-being beyond simple economic growth, using them to measure well-being has certain disadvantages. As various commodities and technological regimes can change across places and over time, the nature of the determinants included in human development indices can also change, and may thus become difficult to compare (Dasgupta 2004).

There are also indices that completely ignore economic factors, such as the Happy Planet Index, which combines information on mortality and life satisfaction in different countries with data on their respective ecological footprints. While mortality and stated life satisfaction are measures of current well-being, a country's ecological footprint is not directly reflected in its current conditions, but instead measures possible impacts on future well-being. Therefore, this index has a dual function that makes its direct interpretation difficult.

While each of these proposed indices covers various aspects of human well-being, there is an increasing recognition that quality of life measures should place more emphasis on the conditions experienced by people at the individual and household levels, rather than on the performance of economic systems at the macro level (Veneri and Edzes 2017). Highly aggregated indicators conceal large inequalities in the distribution of well-being within a society. Utilising individual characteristics aggregated at the subpopulation level can help to overcome the aforementioned limitations of national-level indicators (Mascarenhas et al. 2010). This is particularly important when considering well-being from a sustainability science perspective. Thus, the objective of "leaving no one behind" can be realised by recognising the social gradients in well-being determined by gender and urban-rural residence, among other factors.

In addition, these improved indices should acknowledge the multidimensional nature of human well-being. The academic literature has recognised that no single measure can capture all aspects of human well-being (Chakravarty 2017), and has yet to agree on its most important dimensions. A related point is that different cultures may have different understandings of what constitutes a good life.

Efforts have been made to apply the different aspects of the aforementioned human development indices in the context of Thailand, and thus to move beyond an exclusive focus on economic growth. A study that aimed to describe well-being at the local level was conducted in the north-eastern and southern regions of Thailand (McGregor 2008). In the study, several qualitative and quantitative techniques were used to assess the profile of each community, including its quality of life, expenditures, resources, health and well-being regime. However, a major limitation of the study's approach was its lack of applicability to other communities, or even to countries. Another study aimed to formulate indicators of development based on focus group discussions with local community leaders in the north-eastern region of Thailand (Weeranakin and Promphakping 2018). The result was the identification of a set of themes, such as community trust, well-being, security and strength. However, no further attempts were made to achieve universal applicability. In addition, the methods of analysis used in both studies were intrinsically subjective because they involved qualitative assessments made by researchers.

Thus, several issues concerning the use of the existing human development measures have been raised. To gather information about the various aspects that contribute to human well-being, indices must be constructed to ensure that they reflect how the lives of the people in a given society are improving. However, indices that use aggregated indicators to capture a complex reality face limitations and challenges. From a methodological perspective, the weights assigned to indicators in some of these indices are arbitrary (Lorenz et al. 2017). Thus, the values reported for different populations can be difficult to compare. Moreover, when looking at the temporal comparability of these indices, it is clear that the selected indicators and how they are computed may change over time, which can cause the values of a given index to be different in each period (Ghislandi et al. 2019).

The preference for using objective or subjective indicators constitutes another rift in the academic literature on well-being (Easterlin 1974; Diener et al. 1999, 2017; Kahneman 1999). While some scholars have pointed out the cultural and other sources of bias in self-reported measures, such as life satisfaction and happiness, proponents of subjective measures have argued that individuals are best equipped to evaluate their own life circumstances (Frey and Stutzer 2002). Moreover, subjective measures can capture aspects of well-being that are important to individuals, but that may not be seen as such by external evaluators. However, the use of objective or subjective measures need not be mutually exclusive (Lutz et al. 2021).

The rest of this paper focuses on a new and improved well-being indicator, Years of Good Life (YoGL), which was recently introduced by researchers at the International Institute for Applied Systems Analysis (Lutz et al. 2018, 2021). This indicator addresses the shortcomings of the existing well-being measures discussed above, as it is based on individual-level data that can be flexibly aggregated at the subpopulation level; it is designed to be comparable over time; and it is based on universally shared values. Another advantage of the YoGL indicator is that unlike more abstract indices, it can be interpreted directly. It also allows researchers to assess objective and subjective dimensions of well-being, without involving arbitrary weights.

The YoGL indicator rests on the notion that while being alive is a prerequisite for having any quality of life, mere survival is not enough. Using a demographic life table approach, the indicator adjusts overall life expectancy by counting only the number of years in which individuals have a positive score in each of the following four dimensions: (1) being physically healthy; (2) being cognitively able; (3) being out of poverty; and (4) being subjectively satisfied with life. Each of the four dimensions is captured by a single indicator. The YoGL project builds on the existing work on "Healthy Life Expectancy" (Salomon et al. 2012), and provides a more holistic representation of well-being over the life span. A detailed discussion of YoGL and its dimensions is available in Lutz et al. (2021).

In this paper, we present a practical example of how YoGL can be calculated in the context of two Thai provinces, Chiang Rai and Kalasin, which are located in the northern and the north-eastern regions, respectively. Previous studies have shown that these two locations have experienced above-average rates of poverty in recent decades (McGregor 2008; Weeranakin and Promphakping 2018; Yang et al. 2020). Given the level of development Thailand has achieved based on national-level measures, it is important to understand how progress varies at the subnational level, especially in areas where poverty and other social challenges persist. To this end, data were collected for each of the four YoGL dimensions, and were then applied to a life table by adjusting the age- and sex-specific person-years lived within the population by the share of people who had positive scores in all four dimensions. We also discuss alternative measures for each of the four dimensions of well-being, and we present the YoGL results based on a set of preferred indicators.

As was mentioned above, the YoGL indicator has four dimensions that can be subsumed under two main themes: capable longevity and years with positive life satisfaction (Lutz et al. 2021). Capable longevity is measured based on objectively assessable criteria of what constitutes a good life that reflect the three capabilities of basic health, basic material subsistence and cognitive functioning (Desai et al. 1992). These three components are in line with the general approach that was employed in the Human Development Index. However, for the purposes of measuring YoGL, the conceptualisation and operationalisation of these components were refined (Lutz et al. 2021). One of the core dimensions of YoGL is physical health, and, in particular, having no severe activity limitations. Asking a person about his/her difficulties in activities of daily living (ADLs) or testing his/her physical performance limitations, such as getting up from a chair or walking, can provide an objective approximation of the person's overall health (Weber 2016). Cognitive abilities represent another dimension of YoGL that, like health, declines with age. Cognitive abilities can be assessed through standard numeracy and literacy tests. It should be emphasised that a person's cognitive abilities are not the same as his/her educational attainment. Being out of poverty is the third objective dimension of YoGL, which measures the economic aspects of well-being. An absolute measure of poverty is preferred to a relative measure that reflects social inequalities. Household assets can be used as a proxy for material living conditions, particularly in low- and middle-income countries, where individual reporting on income is less reliable (Lok-Dessallien 2000). The final dimension of YoGL is life satisfaction, which is assessed through an individual's perception of his/her own life. Whether the four proposed indicators are adequate to measure human well-being has yet to be determined in a series of simultaneous studies.

The rest of this paper is organised as follows: In Section 2, we describe the survey design and data collection in the context of two Thai provinces. The results of the survey are presented in Section 3, along with a practical example of how YoGL can be calculated. Section 4 concludes.

2 Data and method

2.1 Dataset

The dataset used for this study came from the project *Developing a measure of human well-being and understanding drivers of sustainable livelihoods under global environmental change.* This dataset was compiled by the Chulalongkorn University College of Population Studies and the International Institute for Applied Systems Analysis (IIASA). The selection of the provinces was based on the prevalence of poverty and the socio-economic vulnerability of the population.

The survey was conducted accordingly, and provincially representative data were collected in two locations: Chiang Rai and Kalasin, which are located in the northern and the north-eastern part of Thailand, respectively. The project was approved by the ethics board of Chulalongkorn University (COA No. 160/2561).

2.2 Sampling

To identify provincially representative households, we adopted the sampling frame developed for the 2016 national survey called the Population Change and Well-being in an Ageing Society (PCWAS) by the College of Population Studies, Chulalongkorn University. Two districts (Amphoe) were selected from the original provinces included in the PCWAS survey: the Muang district was purposefully selected in order to ensure the inclusion of respondents from a highly urbanised area; while the second district was randomly selected. Furthermore, one municipal sub-district and one non-municipal sub-district (Tambon) were chosen within each sampled district (total sub-districts = 4).

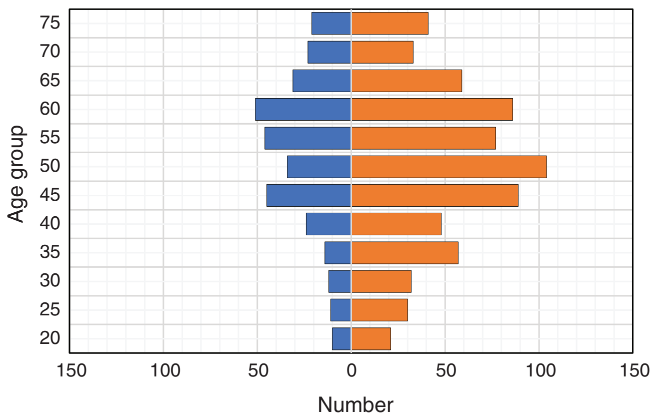
One advantage of adopting the sampling frame established by the PCWAS was that household listings and location maps were already available. The household listings provided the total number of households within the sampled sub-districts and a short description of each household. The location maps, which were produced during the household listing process, depicted the locations of the sampled sub-districts and their entire boundaries, as well as of the households within each sampled sub-district. All listings were updated once the project team entered the field, and the sampled households were approached with the help of the community/village leaders.

To allow for a meaningful statistical analysis, our target was to have at least 500 respondents from each province. To cover non-response or absence, we increased the sample by roughly 10%, to 552. Based on the most recent information from the United Nations' World Urbanization Report (United Nations 2019b), we specified an equal share of respondents from the selected municipal and non-municipal sub-districts (50/50). Since the number of respondents varied depending on the household size, as we explained above, approximately 75% of households within every sampled sub-district were likely to be approached. At least one or two Thai persons aged 20 years or older who had been living in the sampled household for at least three months were approached with the assistance of community/village leaders, as previously mentioned, and were asked if they were willing to participate in the project. Those individuals who were not Thai, were not willing, or were unable to give consent to participate in the project were excluded from the study. Household re-visits were also possible if the interviews were not completed. Due to the design and the scale of the data collection, there was no calculation of statistical weights, because information was gathered from the target households in the communities.

Interviews were conducted face-to-face with one or more household members, depending on the number of adults living in the household. In order to obtain information on the economic activities and status of the household (e.g., income and occupation), the survey focused only on the household member(s) aged 20 years and older. The list enumerating the household members was arranged with the head of the household first, then the partner, then the children, and finally the youngest members. The household head was usually a male adult, though if the parents of a male adult were present, they were listed as the head. Thus, under this procedure, the respondents were selected in a specific manner. In households with between one and three adult members, the second person on the list was asked to participate in the interview. In households with four or more adults, the fourth person was also included in the survey. Due to the selection procedure described above, women were more likely to be interviewed (i.e., the second person listed in the household was usually female). Since the resulting sample was not representative of the overall population (women were overrepresented relative to men), in the subsequent analysis, we presented the results separately for men and women. The small sample size was a major limitation of this study, particularly for certain age groups (e.g., ages 20–35).

Figure 1 shows a population pyramid based on the sample of survey participants, in which 68% were female and 32% were male. The median age was 53; the youngest participant was 18 and the oldest was 100 years old. When we compared this distribution with the figures from the PCWAS, we found that the current sample was similar. Non-response from the selected respondents was low, at below 1%. A total of 999 participants were interviewed.

Figure 1:
Distribution of total sample by age and gender



Notes: Median age = 53, STD. = 14.43.

2.3 Fieldwork and data collection procedures

About 40 college students from two local universities, Chiang Rai Rajabhat University and Mahasarakam University, were hired as survey interviewers. The recruitment was done through personal connections with local researchers. All of the college students underwent training on the following activities:

2.3.1 Training session

All field interviewers were trained thoroughly by the project team on how to explain the project goals. They were also instructed on: (1) how to approach potential respondents; (2) how to ask questions, especially sensitive or complicated ones; (3) how to record answers; (4) what to do in certain situations, for example, if the respondent refuses to take part in the survey; and (5) how to perform consistency checks during the interview or right after the interview is completed. Case-scenario and role-play exercises were carried out during the training sessions to increase the interviewers' familiarity with the questionnaire.

2.3.2 Field editing

The editing work was carried out twice during the data collection process. The first round of editing work was done by the interviewers themselves, just after the completion of the interview, to make sure that the entire questionnaire was filled out. The second round of editing work was carried out by the field supervisors to clarify responses; i.e., to assign the respondents' answers to specific categories.

2.3.3 Daily debrief

A group debriefing was held at the end of each day in the field. During the meeting, the interviewers were encouraged to reflect on and share their field experiences and impressions, as well as any issues they encountered during the fieldwork. They were also asked for their opinions on how such issues could be resolved. If a specific problem had not yet been resolved, the field supervisor and the interviewers discussed it in depth, and looked for a solution.

2.3.4 Office editing

This editing was performed by the project team at the central office (CPS or SRI) after the entire fieldwork had been completed. The editing work included (1) data scrutiny, verification and correction; and (2) the classification of responses (e.g., text or non pre-coded answers).

2.4 Measurement

In this section, we describe in detail the statistical data and procedures we used to measure the Years of Good Life in the context of the two Thai provinces. We propose one main indicator in each of the four YoGL dimensions, and we discuss alternative indicators, which are later used in a sensitivity analysis (Appendix A.2, Table A.1). A notable criterion for the selection of an indicator is that its distribution has to be on the tail ends of the distribution in order to identify individuals in dire conditions (Lutz et al. 2021).

The items were based on standard international surveys, such as the Demographic and Health Survey, to ensure the reliability of the information collected. The translation of the items from English to Thai was carried out by research scientists involved in the project, and crosschecked by the principal investigator. A subsequent back-translation was performed with an equally stringent process to check the validity of the initial translation to the Thai language. The Thai language version of the questionnaire was pre-tested three times to ensure that the wording of the questions was appropriate for the Thai context. Note that when applicable, we also provide the corresponding question number in the Thai version of the questionnaire for reference.

2.4.1 Physical limitations

Due to bias in self-reported measures of health (Spitzer and Weber 2019), objective measures should ideally be used to assess the health status of individuals. This can be done through physical tests, such as chair stand and walking speed tests, both of which have been extensively used in surveys on health and ageing (SAGE and SHARE, for example). However, conducting physical tests can be time-consuming, and requires the additional training of interviewers. For the purposes of this project, data on both objective and subjective measures of health were collected.

The ability of each respondent to get into and out of a chair without assistance was assessed objectively by the interviewer. After confirming that they felt safe performing physical tasks, respondents were asked to perform the action of getting into and out of a chair from a sitting position without assistance, and their performance was assessed by the interviewer. In this context, "assistance" refers to the respondents using their arms when performing the test. This chair stand test is commonly used to assess age-related decline in physical functioning and power, since rising from a chair requires both lower limb strength and power, as well as balance and coordination. Figure 2 shows the types of chairs used to perform the chair stand test during the field work.

In addition to the chair stand test, we collected information on the subjective health status of individuals, which is also widely used in the literature. One direct approach is to simply ask the respondents to describe their perceived level of disability as measured by the concept of general activity limitations. In the survey, we used the following question from the Global Activity Limitation

Figure 2:
Types of chairs used to test physical health in the Thai survey context

Instrument (GALI): "*For at least the past 6 months, to what extent have you been limited because of a health problem in activities people usually do?*" (Jagger et al. 2010).

Another approach to measuring disability is to assess the difficulties people have in performing activities of daily living (ADLs). There are six basic ADLs: eating, bathing, getting dressed, toileting, transferring and continence. The six dimensions follow a hierarchical structure. See Section B of the questionnaire in Appendix A.1 for a list of the items used to assess GALI and ADLs in the survey. More comprehensive lists of ADLs are also available. For example, the World Health Organization's Assessment Schedule was developed through an international collaboration in order to assess health and disability status in a way that would be applicable across cultures and in all adult populations. It is a tool that produces standardised disability levels directly linked with the concepts of the International Classification of Functioning, Disability and Health (ICF). It is applicable to a range of diseases, including mental, neurological and addictive disorders (Üstün et al. 2010).

The extended ADL lists include between 12 and 36 items. The 12-item version explains 81% of the variance of the 36-item version, and it is short, simple and easy to administer (average time of five minutes). Moreover, the 12-item version is applicable in both clinical and general population settings (Üstün et al. 2010). It covers six domains of functioning: cognition, mobility, self-care, getting along, life activities and participation.

While collecting subjective health information is easier, subjective assessments are not considered to be as reliable as objective health measures. We have constructed the YoGL indicator based on both objective and subjective health measures, and compared the differences in the outcomes (see the sensitivity analysis in the results section).

2.4.2 Cognitive limitations

The second dimension of the YoGL indicator captures the cognitive functioning of the respondents. Similar to physical health, it should ideally be assessed through objective tests, such as through tests of literacy, numeracy and memory. In the survey, we included three tests of cognitive functioning with different levels of difficulty. We also included a question on self-reported literacy for comparison.

There are many demographic and health surveys around the world that include questions about the participants' literacy, numeracy and/or memory in the context of healthy ageing (e.g., SHARE and SAGE). In addition, a range of instruments have been specifically developed to capture a more holistic picture of the participants' cognitive skills (e.g., Skills Towards Employment and Productivity – STEP) (Pierre et al. 2013). Given the time restrictions of our survey, and given that the scope of the survey was not limited to testing cognitive function, but included other dimensions of well-being as well, we attempted to design a series of questions that, while short, were relatively comprehensive. Specifically, the questions covered the ability to read a simple sentence (C1) (based on DHS), the ability to recognise print vocabulary (C4) (based on PIAAC/STEP), the ability to process a sentence (C5) (PIAAC/STEP) and the ability to correctly identify the day of the week (C6) (based on SHARE) as a proxy for cognitive functioning. By including questions from DHS, SHARE and PIAAC/STEP, we were not only building on existing knowledge; we were ensuring the comparability of the results. The detailed questions are included in Section C of the questionnaire (See Appendix A.1).

2.4.3 Being out of poverty

Economic items were also included in the questionnaire in order to capture the respondents' capability to fulfil their basic needs. Since this project was concerned with the severe deprivation of basic needs, it relied on absolute rather than relative measures of poverty. Absolute poverty is traditionally measured in monetary terms, such as the share of the population below a certain level of income or consumption expenditure. For example, the World Bank has set the poverty threshold at $1 a day (Ferreira et al. 2016). The use of such measures in the context of developing countries has been criticised because collecting reliable data can be difficult, and defining an internationally comparable poverty line can be a challenge (Jolliffe and Prydz 2016). Moreover, income and expenditure data are usually collected at the household level, and disaggregation to the individual level is problematic. More recent measures of poverty go beyond monetary considerations, and thus incorporate multiple indicators of different dimensions of poverty, such as housing, unemployment, nutrition status and access to services, to mention a few (e.g., the Multidimensional Poverty Index and the DHS wealth index).

In light of the above considerations, a set of questions on household living conditions and personal items were selected to capture absolute poverty among the survey participants. Information was collected on the following items:

A. Ownership of basic household items/living conditions:

- Type of toilet facility (G5)
- Main material of outside walls of dwelling (G1)
- Asset ownership: The household owns a refrigerator (G6)

B. Individual items:

- Skipping meals because there is not enough food in the house (D1)
- Limiting variety of foods due to a lack of resources (D3)

The wall type and toilet facility have been used in both the MPI and the DHS as economic indicators, and have been found to strongly correlate with household wealth (OPHI 2009). Ensuring adequate sanitation was also part of the Millennium Development Goals (MDGs), and is currently included in SDG 6 ("Access to safe water and sanitation"). While the living conditions applied to the whole household, this project was particularly interested in assessing individual deprivation, which might not have been the same among household members. Individual dimensions of poverty were captured through questions on food consumption and asset ownership. The questions related to the measurement of poverty are presented in Sections D and G in the questionnaire (see Appendix A.1).

2.4.4 Life satisfaction

The final dimension was concerned with the respondents' satisfaction with life. Considerable attention has been paid to this aspect of subjective well-being (SWB), with several studies on this topic appearing each year (Diener et al. 2017). Life satisfaction is considered a more accurate measure of SWB than happiness, because it requires a respondent to engage in a more reflective review of his/her entire life. By contrast, questions about happiness can yield volatile responses due to the emotional state it can elicit in individuals (Diener et al. 2017). In the current study, we used the Satisfaction with Life Scale (SWLS), which consists of five questions that were developed to assess the respondents' satisfaction with their life as a whole (Pavot and Diener 2008). Since its initial development as a brief assessment tool indicating satisfaction with one's life as a whole (Larsen et al. 1985), it has been demonstrated that the SWLS has good psychometric characteristics, and can be used in various contexts and cultures (Pavot and Diener 1993, 2008). While the scale does not assess satisfaction with individual life domains, such as health or finances, it allows the respondents to integrate and weight these domains in whatever way they choose. The following items comprise the SWLS, each of which can be ranked from one (lowest score) to five (highest score) by the respondent:

- In most ways my life is close to my ideal (F1);
- The conditions of my life are excellent (F2);
- I am satisfied with my life (F3);
- So far, I have gotten the important things I want in life (F4);
- If I could live my life over again, I would change almost nothing (F5).

In addition, a single item of the life satisfaction scale asking the respondent to rate his/her satisfaction with life as a whole on a scale from one to 10 (F6) was introduced and compared with the five questions in F1–F5. The literature has shown that this single question is reliable and consistent (Bonikowska et al. 2014). Details on each item included in the Life Satisfaction domain of the questionnaire are provided in Appendix A.1 (Part F).

2.4.5 Years of good life

From the indices referring to the four constituent variables discussed above, a binary variable was created for each that indicated whether an individual was below the critical threshold. The cut-offs for the critical threshold are presented in the results section. The binary variable was then aggregated by age and gender. The proportion generated from this dichotomy was then applied to the number of person-years lived in the group, which indicated what portion of remaining life was, on average, spent in good physical and cognitive health, out of absolute poverty and with positive life satisfaction. The life table we used was from the UN World Population Prospects for Thailand, which was disaggregated by gender and 10-year age groups (United Nations 2019a). This method for determining morbidity prevalence was similar to the Sullivan method in (Sullivan 1971).

3 Results

3.1 Descriptive results

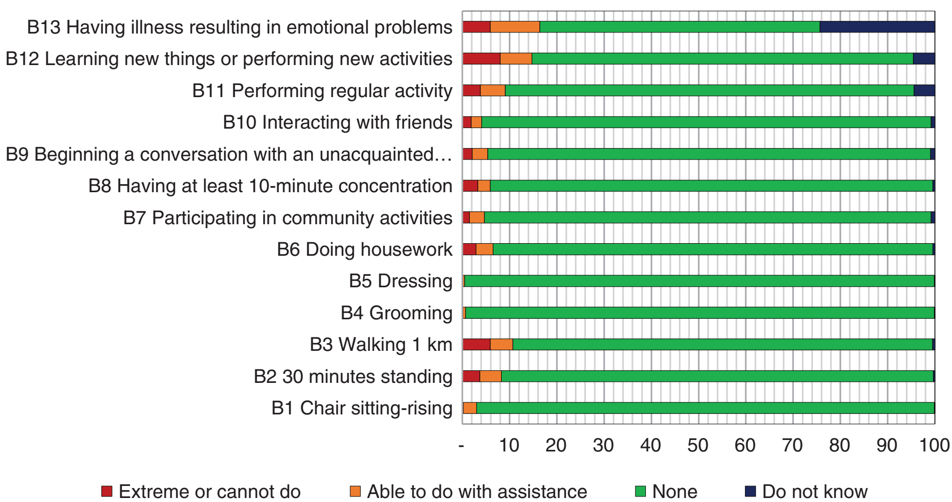
This section presents the distribution of the sample by selected characteristics relevant to YoGL. The results of the current survey were in lieu of observations from the PCWAS survey, which was conducted in 2016 across all the regions of Thailand.

3.1.1 Physical limitations

In terms of physical limitations, most participants were able to perform the chair stand test, either alone or with assistance (see Figure 3). Only two participants had extreme difficulties or could not perform the test at all, while 28 were able to perform it with some assistance. The participants who were not able to get up from the chair, needed assistance or refused to perform the test were classified as not free from physical limitations in the subsequent YoGL analysis. Less than 5% of both males and females fell into this category. In terms of demographic heterogeneities, the small sample size did not permit the meaningful disaggregation of the results by age and sex (see Table 1).

In terms of other activities of daily living, Figure 3 shows a breakdown of 12 ADLs by level of difficulty. Most of the participants did not report having difficulties performing basic activities, such as grooming and dressing. Larger shares of participants reported having difficulties performing more strenuous activities, such as walking 1 km, standing for 30 minutes and performing new activities.

Figure 3:
Prevalence of limitations in activities of daily living (ADLs) by level of difficulty



Notes: Item B1 (chair sitting-rising) was objectively assessed by the interviewer. Items in B2 to B12 were assessed based on self-reports by the survey participants.

Table 1:
Proportion with severe physical limitations by age group and sex

	Chair stand (tested)		ADLs (5 items)		ADLs (12 items)		GALI	
Gender	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
<= 29	4.76	0.00	4.76	1.96	14.29	5.88	0.00	0.00
30–39	3.85	0.00	3.85	3.37	26.92	15.73	0.00	0.00
40–49	0.00	0.00	4.35	3.65	21.74	24.09	1.45	4.38
50–59	5.00	2.21	10.00	10.50	26.25	35.91	8.75	6.63
60–69	0.00	4.14	12.20	28.28	32.93	51.03	6.10	4.83
70+	4.55	17.57	18.18	44.59	36.36	70.27	4.55	10.81
Total	2.48	3.40	9.63	15.07	27.64	35.60	4.66	4.87

Note: Individuals who used assistance to perform the chair stand test were classified as having severe physical limitations. The ADLs were calculated based on items B1-B5 or B1-B12 in the questionnaire. The ADL shares reported in the table reflect the share of respondents who reported having extreme difficulties or who could not perform at least one of the activities without assistance. The GALI was calculated based on item B14 in the questionnaire. The GALI shares reported in the table reflect the share of respondents who reported having severe limitations only.

Table 1 shows a comparison of activity limitations by age and sex based on different indicators. The ADL and GALI (item B14 in the questionnaire) measures were constructed based on being able to perform two or more activities without any assistance. In the sensitivity analysis, the ability to perform at least half of the activities for each composite measure was used to determine whether individuals had physical limitations. This latter composite index has been shown to be statistically similar to the default index.

We can see that there were substantial differences in disability prevalence depending on the choice of indicators. The results of the chair stand test were comparable to those of the five-item list of ADLs among the younger age groups, but not among the older adults. The results of the self-reported GALI question seemed to come closer to those of the chair stand test. As expected, activity limitations were found to increase with age, and more strongly for females than for males.

3.1.2 Cognitive limitations

To assess the respondents' cognitive abilities, we performed a series of cognitive tests with different levels of difficulty, and we also asked the respondents to evaluate their own abilities. Table 2 shows that about 63% of the interviewed individuals reported that they could read/write without any problem, but only 50% were able to pass the simple reading test. As this means that the respondents had a tendency to overestimate their abilities, the self-reported measures should be treated with caution. Both males and females of all age groups seemed to overestimate their reading abilities (Table 4). For example, only 5% of males under the age of 30

Table 2:
Self-reported ability to read (A7) and write (A8) and tested ability to read simple text (C1)

	Self-reported ability to read (A7)	Self-reported ability to write (A8)	Tested ability to read (C1)
Cannot at all	13.21	9.91	19.62
Can with difficulty	23.42	27.63	30.13
Can without any problems	63.36	62.46	50.25
Total	100	100	100

Note: Individuals who were visually impaired or refused to perform the reading test are included in the category "cannot at all". Individuals who were only able to read part of a sentence or almost the whole sentence are included in the category "can with difficulty".

reported that they had difficulties reading, but nearly 20% could not pass the simple reading test.

Nearly all respondents were able to recognise simple print vocabulary (C4) (99.6%), and to correctly identify the day of the week (C6) (96.8%) (Table 3). These two tests measured basic cognitive functioning. A much larger share of respondents were not able to pass the advanced test of reading with understanding (items C5-1 to C5-5 in the questionnaire). Even among the younger individuals (under age 30), 62% of males and 51% of females failed this test. These shares increased to 70% of males and 92% of females among the sample of older individuals (70+); see the last two columns of Table 4.

3.1.3 Poverty

Different dimensions of absolute poverty were considered in the survey. Two indicators focused on food consumption and dietary diversity (items D1 and D3 in the questionnaire), given that insufficient caloric intake is a key issue in low-

Table 3:
Tested ability to recognise print vocabulary (C4), to read with understanding (C5-1 to C5-5) and to correctly identify the day of the week (C6)

	C4	C5-1	C5-2	C5-3	C5-4	C5-5	C6
Correct answer	99.63	87.83	57.27	93.17	96.15	84.22	96.77
Incorrect answer	0.25	8.94	39.63	3.48	0.75	12.67	3.23
Do not know	0.12	3.23	3.11	3.35	3.11	3.11	0.00
Total	100	100	100	100	100	100	100

Table 4:
Proportion with cognitive limitations by age group and sex – self-reported and tested

	Self-reported ability to read (A7)		Self-reported ability to write (A8)		Tested ability to read (C1)		Tested ability to read with understanding (C5)	
Gender	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
<= 29	4.76	7.84	4.76	9.80	19.05	17.65	61.90	50.98
30–39	7.69	11.24	11.54	12.36	19.23	22.47	50.00	51.69
40–49	30.43	27.01	30.43	27.74	42.03	43.07	53.62	51.82
50–59	21.25	40.88	25.00	42.54	37.50	58.56	55.00	63.54
60–69	46.34	59.31	48.78	58.62	56.10	74.48	60.98	75.86
70+	45.45	75.68	45.45	72.97	59.09	74.32	70.45	91.89
Total	30.75	39.44	32.61	39.88	43.48	52.73	58.39	64.40

Note: Share of respondents who: A7 = Cannot read at all; A8 = Cannot write at all; C1 = Cannot read at all. Tested ability to read with understanding was assessed based on having the correct answers to all included items (C5-1 to C5-5).

and middle-income countries. The results showed that 17% of respondents reported skipping meals, and 27% reported limiting the variety of foods they consumed due to a lack of resources (Table 5).

Beyond inadequate food consumption, income-related indicators were also included in the survey, such as ownership of a welfare card (D7), and the subjective evaluation of the individual's financial situation (D8, D9). One-third of the respondents described their financial situation as bad or very bad, and 40% indicated that their financial situation had become worse over the previous five years. It should be noted that these indicators were subjective, as they relied on the respondent's own evaluations. Around 60% of the respondents reported receiving financial aid from the government.

Other poverty indicators focused on ownership of specific household assets and living conditions (items G1, G5 and G6 in the questionnaire). Such indicators are often used to assess poverty in low- and middle-income countries, where reporting on income is less reliable (Jolliffe and Prydz 2016). Depending on the specific indicator for household assets or living conditions, the poverty levels varied substantially. Only 4% of the respondents reported having unfinished walls, and just 4.3% said they do not own a fridge (Table 5). However, 64% of the households indicated that do not have a flush toilet, which could be considered a sign of poverty. However, it may also reflect a lack of infrastructure in remote rural areas.

The breakdown by age and gender uncovered no clear pattern of poverty among the population subgroups (Table 6). Based on whether they were skipping meals and had dietary diversity, younger and older individuals were equally likely to be

Table 5:
Distribution of sample population by poverty measure

Items	%
Skipping meals in a day because there was not enough food (D1)	
Never	83.18
Yes	16.82
Limiting variety of foods due to a lack of resources (D3)	
Never	73.37
Yes	26.63
Owns a welfare card (low-income card) (D7)	
No	39.44
Yes	60.26
Current financial situation (D8)	
Very bad	4.40
Bad	27.43
Neither good nor bad	61.76
Good	6.21
Very good	0.20
Financial situation compared to the past 5 years (D9)	
Much worse	8.51
Worse	31.53
Stayed the same	27.53
Better	30.43
Much better	2.00
Wall or building material (G1)	
Brick, block, stone or cement	44.14
Wood	15.62
Half brick, half wood	35.94
Makeshift, salvaged or improvised materials from the local area (e.g., plant-based material)	0.10
Reused materials (e.g, cardboard, scrap)	4.00
Type of toilet facility (G5)	
Flush toilet	24.22
Squat toilet	63.96
Flush toilet and squat toilet	11.61
No toilet facility	0.10
Owns a fridge (G6)	
No	4.30
Yes	95.60

classified as poor. The exception was among respondents under age 30: in this age group, only 5% of males, but 14% of females, reported skipping meals. The likelihood of owning a welfare card was generally higher among the older age groups. Interestingly, younger individuals were more likely than older people to report living in a house without a flush toilet.

Table 6:
Proportion living in absolute poverty by age group, sex and poverty measure

	Skipping meals (D1)		Dietary diversity (D3)		Welfare card (D7)		Flush toilet (G5)	
Gender	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
<29	4.76	13.73	14.29	27.45	47.62	45.10	85.71	76.47
30–39	19.23	17.98	23.08	31.46	46.15	59.55	46.15	70.79
40–49	23.19	17.52	24.64	24.82	49.28	64.96	71.01	69.34
50–59	13.75	19.34	25.00	34.25	58.75	59.12	63.75	63.54
60–69	18.29	13.79	26.83	24.14	74.39	64.83	60.98	57.93
70+	20.45	12.16	13.64	25.68	65.91	62.16	63.64	50.00
Total	17.70	16.40	22.98	28.36	59.94	60.86	64.60	63.96

3.1.4 Life satisfaction

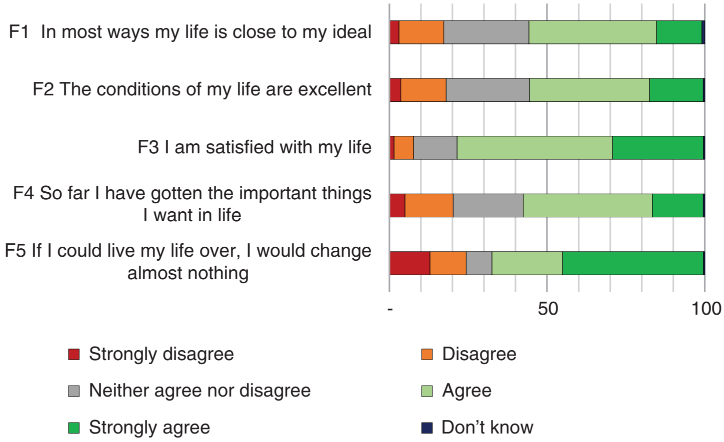
The final dimension of YoGL concerns the respondents' subjective evaluation of their own life. One way to capture this dimension was by asking the respondents about their overall life satisfaction on a scale from one to 10 (item F6 in the questionnaire). Depending on the cut-off point at which life satisfaction was assessed, the overall share of individuals reporting that they are not satisfied with life ranged from 2% to nearly 21% for males and from 3% to 20% for females (Table 7). No clear trend across age groups could be observed.

Alternative questions concerning life satisfaction are presented in Figure 4. Between 8% and 24% of the respondents reported that they are not satisfied with certain aspects of their life (strongly disagreeing or disagreeing with the

Table 7:
Proportion unsatisfied with life by sex

	Life satisfaction (F6 score < 4)		Life satisfaction (F6 score < 5)		Life satisfaction (F6 score < 6)	
Gender	**M**	**F**	**M**	**F**	**M**	**F**
<= 29	0	3.92	0.00	3.92	19.05	21.57
30–39	0	2.15	0.00	5.62	15.38	19.10
40–49	0	0.73	8.70	2.92	21.74	21.90
50–59	3.75	3.87	6.25	7.18	20.00	20.44
60–69	3.66	3.45	4.88	5.52	23.17	20.00
70+	0	2.70	2.27	5.41	18.18	17.57
Total	1.86	2.81	4.97	5.32	20.50	20.24

Figure 4:
Life satisfaction by level of agreement



statements). A large share of individuals (24%) reported that they would like to change something if they could live their life over again.

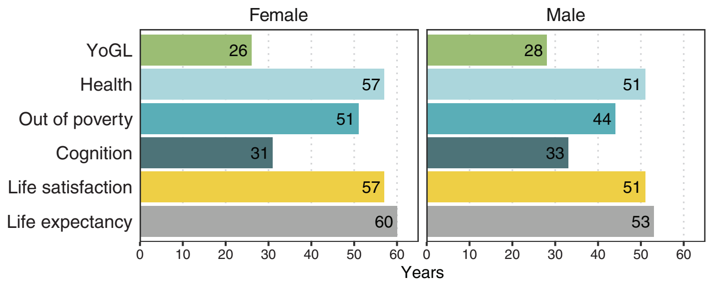
3.2 Years of Good Life (YoGL)

The main results presented in this section are based on a set of preferred indicators for each dimension of YoGL. The alternative measures described in the previous section were used in a series of sensitivity analyses (see Section 3.3), and demonstrate how YoGL can be adapted for different contexts. For the main results, the following measures (and cut-off points) were selected in each dimension:

- Physical health: Can stand up from a chair without assistance.
- Cognitive health: Can read a full sentence (tested).
- Out of poverty: Does not skip meals.
- Life satisfaction: Reports a life satisfaction score of five or higher.

In the physical health dimension, the chair stand test was selected because its distribution was observed to represent the tail ends, or those who were in relatively poor condition. In addition, the chair stand test was utilised in the seminal study that demonstrated YoGL as a well-being indicator (Lutz et al. 2021). For testing the respondents' cognitive health, the ability to read was selected as the measure for YoGL because it did not suffer from reporting error. Poverty was assessed based on the measure of not skipping meals. This measure was preferred to measures of household assets and living conditions because it reflected individual poverty, and allowed for differences between males and females within the same household.

Figure 5:
YoGL and individual dimensions at age 20 by sex



Notes: Life expectancy by five-year age group and sex for Thailand was retrieved from the United Nations World Population Prospects (https://population.un.org/wpp/Download/Standard/Population/).

Finally, to measure life satisfaction, the item with a 10-point scale response category was selected because of its comparability with the measures used in previous studies on YoGL (Lutz et al. 2021). Detailed information about the share of individuals with positive scores on each YoGL dimension by age and sex is provided in Appendix A.2, Table A.2.

When the four constituent dimensions of YoGL were integrated into life expectancy at age 20, the years of good life among women and men were similar: 26 years for women and 28 years for men (Figure 5). Compared to the overall life expectancy at age 20, which was 60 years for women and 53 years for men, we can see that years of good life were lost in each YoGL dimension. Cognitive health was the dimension that reduced the years of good life the most for both sexes (by 29 years for females, and by 20 years for males). In contrast, relatively few years of good life were lost to poor health, low life satisfaction and poverty.

In the context of Thailand, this is the first study to present an integrated view of years of good life in terms of physical health, cognition, life satisfaction and poverty. Previous studies have estimated healthy life expectancy (HLE) in Thailand using comparative characteristics, such as living in an urban or a rural area (Karcharnubarn et al. 2013), or betweeen men and women (Apinonkul et al. 2015). While HLE is useful for determining years without health limitations, it does not explicitly account for other sources of health gradients, such as poverty.

Other studies focused on the older Thai population have tried to identify health determinants and their contributions to healthy years. It has, for example, been observed that individuals with higher levels of income and education have advantages in delaying the onset of physical limitations (Pothisiri et al. 2020) and the decline in cognitive performance (Vicerra and Pothisiri 2020). The approach these studies employed was different, as they compared the health performance of a set of socio-economic subgroups. For example, the years of advantage gained by individuals with higher educational attainment levels were comparable to those gained by individuals with lower levels of education. Individual indices within YoGL also allow for the comparison of years of good life between subpopulations. The difference with previous methods is that those respective indices of physical and cognitive health measures can be constructed into a complete indicator which also accounts for poverty and life satisfaction.

3.3 Sensitivity analysis

We tested the sensitivity of the main YoGL results presented above using different specifications in each YoGL dimension. First, we changed the cut-off points used in the dichotomisation of the health, cognition and life satisfaction measures. Second, we replaced the indicators used in each dimension with alternative indicators, which were discussed in Section 3.1 above. The results of the sensitivity analyses are presented below. Details about the indicators and cut-off points used in each sensitivity analysis are presented in Appendix A.2.

3.3.1 Alternative cut-offs

In a first robustness check, we changed the health measure specification by considering people who used assistance to get out of a chair as healthy. The results for women did not change, and men gained one year of good life (Figure 6, Sensitivity 1). In a second sensitivity test, we changed the specification of the cognitive ability measure by considering people who could read only parts of a sentence as cognitively able. This time, the results changed substantially, with women gaining 12 additional years of good life, and men gaining 10 years (Figure 6, Sensitivity 2). Finally, we changed the cut-off points for the life satisfaction measure by considering individuals who reported a score of four and above (Sensitivity 3) or of six and above (Sensitivity 4) as satisfied. The YoGL results were not noticeably affected in Sensitivity 3, and were only slightly reduced in Sensitivity 4 (Figure 6).

3.3.2 Alternative indicators

We also tested the robustness of the results by replacing the measures used in individual YoGL dimensions with alternative measures. For the physical health dimension, we considered five-item and 12-item ADLs, as well as GALI measures. It should be noted that these three measures were based on self-reported health, rather than on tested health. While no notable change in YoGL was observed when using the GALI and the five-item ADLs, the 12-item ADLs reduced YoGL by six years for women and by five years for men (Figure 7, Sensitivity 5–7). These findings reflect the fact that older adults, and especially women, had a tendency to report more limitations in performing activities of daily living (Table 1).

Figure 6:
YoGL at age 20 by sex calculated with alternative cut-offs for the indicators in the individual dimensions

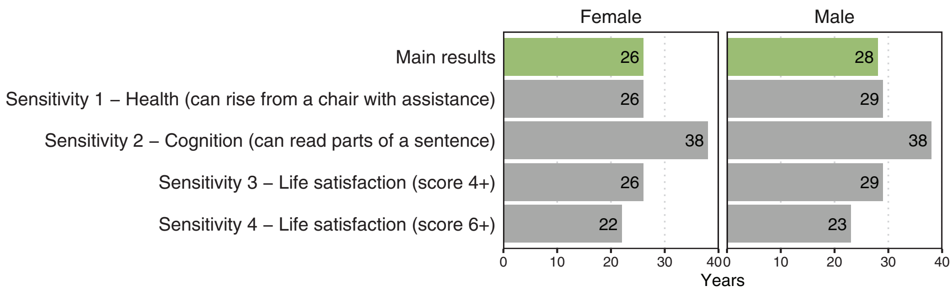
Notes: Bars show YoGL calculated with the preferred cut-offs for each indicator (green bar), and by replacing the cut-offs for one indicator at a time (grey bars).

In addition, we found that the respondents' self-assessments of their ability to read overestimated their true ability to read, as some of the individuals who indicated that they could read were unable to perform the reading test when presented with the opportunity. When we used the self-reported reading and writing measures for calculating YoGL, years of good life improved by four years for women and by six years for men (Figure 7, Sensitivity 8–9). This indicates that men tended to overestimate their reading and writing skills more than women. When we used the more complex test of reading with understanding instead of the simple reading test, years of good life at age 20 were reduced to only 14 years for women and to 15 years for men (Sensitivity 10).

Different measures for poverty were also tested. Using a measure for dietary diversity instead of meal frequency did not noticeably change the results. However, when we applied measures based on ownership of a welfare card and having a flush toilet, years of good life were reduced substantially for both men and women (Figure 7, Sensitivity 11–13).

Finally, we tested the robustness of our main YoGL results by using different life satisfaction measures (Figure 7, Sensitivity 14–16). The different specifications did not substantially change the results, with the exception of the life satisfaction measure based on the following statement: "*If I could live my life over, I would*

Figure 7:
YoGL at age 20 by sex calculated with alternative indicators in the individual dimensions



Notes: Bars show YoGL calculated with the preferred indicators (green bar) and by replacing one indicator at a time (grey bars).

change almost nothing". A larger share of individuals reported having lower life satisfaction based on this measure than on the other life satisfaction measures, which resulted in lower YoGL for both sexes (Sensitivity 16).

4 Conclusion

As the economy of Thailand continues to grow, it is often assumed that human development in the country has been increasing as well. But because this perspective masks the heterogeneity of the population, the varying levels of poverty experienced by people with different characteristics are not granted sufficient attention. We addressed this issue by utilising the YoGL as a well-being indicator. Gathering data from Chiang Rai and Kalasin, which are located in regions of the country with a high prevalence of poverty, we demonstrated the extent to which social and economic indicators contributed to years of good life among a sample of men and women.

The application of YoGL as a well-being indicator was demonstrated in this study based on survey data collected from two province in Thailand. This "bottom-up" approach to measuring human well-being is flexible, and can easily be applied to different development contexts and population subgroups. While the sample used in this study was sufficient for the purposes of this project, its generalisability is limited. The collection of larger samples would allow for more in-depth comparisons to be made across population subgroups. In this paper, we demonstrated how YoGL can be used to compare the well-being of men and women. This approach can also be used to investigate heterogeneities based on level of education, residence status and occupational status, among many other characteristics. The techniques used to construct YoGL allow for such comparisons to be made between population subgroups and over time. Moreover, the individual constituents of YoGL can be examined to determine the sources of these heterogeneities.

References

Apinonkul, B., K. Soonthorndhada, P. Vapattanawong, W. Aekplakorn and C. Jagger 2015. Gender differences in health expectancies across the disablement process among older Thais. *PLoS ONE* 10(3): e0121310. https://doi.org/10.1371/journal.pone.0121310

Bonikowska, A., J. F. Helliwell, F. Hou and G. Schellenberg 2014. An assessment of life satisfaction responses on recent Statistics Canada Surveys. *Social Indicators Research* 118(2): 617–643. https://doi.org/10.1007/s11205-013-0437-1

Chakravarty, S. R. 2017. *Analyzing multidimensional well-being: A quantitative approach*. John Wiley & Sons. https://doi.org/10.1002/9781119257424

Dasgupta, P. 2004. *Human well-being and the natural environment*. Oxford University Press. https://doi.org/10.1093/0199247889.001.0001

Desai, M., A. Sen and J. Boltvinik 1992. *Social progress index: A proposal.* Bogota: United Nations Development Programme.

Lok-Dessallien, R. 2000. *Review of poverty concepts and indicators*. New York: United Nations Development Programme.

Diener, E., S. J. Heintzelman, K. Kushlev, L. Tay, D. Wirtz, L. D. Lutes and S. Oishi 2017. Findings all psychologists should know from the new science on subjective well-being. *Canadian Psychology/Psychologie Canadienne* 58(2): 87–104. https://doi.apa.org/doi/10.1037/cap0000063

Diener, E., E. M. Suh, R. E. Lucas and H. L. Smith 1999. Subjective well-being: Three decades of progress. *Psychological Bulletin* 125(2): 276–302. https://psycnet.apa.org/doi/10.1037/0033-2909.125.2.276

Easterlin, R. A. 1974. Does economic growth improve the human lot? Some empirical evidence. In: *Nations and households in economic growth: Essays in honor of Moses Abramovitz*, eds P. A. David and M. W. Reder, 89–125. Academic Press. https://doi.org/10.1016/B978-0-12-205050-3.50008-7

Ferreira, F. H. G., S. Chen, A. Dabalen, Y. Dikhanov, N. Hamadeh, D. Jolliffe, A. Narayan, E. B. Prydz, A. Revenga, P. Sangraula, U. Serajuddin and N. Yoshida 2016. A global count of the extreme poor in 2012: Data issues, methodology and initial results. *The Journal of Economic Inequality* 14(2): 141–172. https://doi.org/10.1007/s10888-016-9326-6

Frey, B. S. and A. Stutzer 2002. What can economists learn from happiness research? *Journal of Economic Literature* 40(2): 402–435.

Ghislandi, S., W. C. Sanderson and S. Scherbov 2019. A simple measure of human development: The Human Life Indicator. *Population and Development Review* 45(1): 219–233. https://doi.org/10.1111/padr.12205

Gil Sander, F. and A. Burgard 2011. *Thailand economic monitor: April 2011*. Washington, D.C.

Jagger, C., C. Gillies, E. Cambois, H. Van Oyen, W. Nusselder and J. M. Robine 2010. The global activity limitation index measured function and disability similarly across

European countries. *Journal of Clinical Epidemiology* 63(8): 892–899. https://doi.org/10.1016/j.jclinepi.2009.11.002

Jolliffe, D. and E. B. Prydz 2016. Estimating international poverty lines from comparable national thresholds. Policy Research Working Paper No. 7606. Washington, D.C.: World Bank. http://hdl.handle.net/10986/24148

Kahneman, D. 1999. Objective happiness. In: *Well-being: The foundations of hedonic psychology*, eds D. Kahneman, E. Diener and N. Schwarz, 3–25. New York: Russell Sage Foundation.

Karcharnubarn, R., P. Rees and M. Gould 2013. Healthy life expectancy changes in Thailand, 2002–2007. *Health and Place* 24:1–10. https://doi.org/10.1016/j.healthplace.2013.08.002

Larsen, R. J., E. D. Diener and R. A. Emmons 1985. An evaluation of subjective well-being measures. *Social Indicators Research* 17: 1–17. https://doi.org/10.1007/BF00354108

Lorenz, J., C. Brauer and D. Lorenz 2017. Rank-optimal weighting or "How to be best in the OECD better life index?" *Social Indicators Research* 134(1): 75–92. https://doi.org/10.1007/s11205-016-1416-0

Lutz, W., A. A. Lijadi, E. Strießnig, A. Dimitrova and M. Caldeira Brant de Souza Lima 2018. *Years of Good Life (YoGL): A new indicator for assessing sustainable progress*. IIASA Working Paper WP-18-007. Laxenburg, Austria: International Institute of Applied Systems Analysis. http://pure.iiasa.ac.at/id/eprint/15402/1/WP-18-007.pdf

Lutz, W., E. Striessnig, A. Dimitrova, S. Ghislandi, A. Lijadi, C. Reiter, S. Spitzer and D. Yildiz 2021. Years of good life is a well-being indicator designed to serve research on sustainability. *Proceedings of the National Academy of Sciences* 118(12): e1907351118. https://doi.org/10.1073/pnas.1907351118

Mascarenhas, A., P. Coelho, E. Subtil and T. B. Ramos 2010. The role of common local indicators in regional sustainability assessment. *Ecological Indicators* 10(3): 646–656.

McGregor, J. A. 2008. Wellbeing, development and social change in Thailand. *Thammasat Economic Journal* 26(2): 1–27.

Mclean, D. 2014. National and international indices of well-being: A critical analysis. *Journal of the Indiana Academy of the Social Sciences* 17(1): 5.

OPHI 2009. Measuring multidimensional poverty: Insights from around the world. OPHI Briefing 30. Oxford: Oxford Poverty and Human Development Initiative (OPHI), University of Oxford.

Pavot, W. and E. Diener 1993. Review of the satisfaction with life scale. *Psychological Assessment* 5(2): 164–172. https://psycnet.apa.org/doi/10.1037/1040-3590.5.2.164

Pavot, W. and E. Diener 2008. The satisfaction with life scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology* 3(2): 137–152. https://doi.org/10.1080/17439760701756946

Pierre, G., M. Laura Sanchez Puerta, A. Valerio and T. Rajadel 2013. STEP skills measurement surveys: Innovative tools for assessing skills. Social Protection and Labor Discussion Paper 1421 Washington, D.C.: World Bank Group.

Pothisiri, W., O. Prasitsiriphon and W. Aekplakorn 2020. Extent of aging across education and income subgroups in Thailand: Application of a characteristic-based age approach. *PLoS ONE* 15(12): e0243081. https://doi.org/10.1371/journal.pone.0243081

Salomon, J. A., H. Wang, M. K. Freeman, T. Vos, A. D. Flaxman, A. D. Lopez and C. J. L. Murray 2012. Healthy life expectancy for 187 countries, 1990–2010: A systematic analysis for the Global Burden Disease Study 2010. *The Lancet* 380(9859): 2144–2162. https://doi.org/10.1016/S0140-6736(12)61690-0

Spitzer, S. and D. Weber 2019. Reporting biases in self-assessed physical and cognitive health status of older Europeans. *PLoS ONE* 14(10): e0223526. https://doi.org/10.1371/journal.pone.0223526

Stiglitz, J., A. Sen and J. P. Fitoussi 2010. *Mismeasuring our lives: Why GDP doesn't add up*. The New Press.

Strezov, V., A. Evans and T. J. Evans 2017. Assessment of the economic, social and environmental dimensions of the indicators for sustainable development. *Sustainable Development* 25(3): 242–253. https://doi.org/10.1002/sd.1649

Sullivan, D. F. 1971. A single index of mortality and morbidity. *HSMHA Health Reports* 86(4): 347–354.

United Nations 2019a. 2019 Revision of World Population Prospects. Retrieved from https://population.un.org/wpp/

United Nations 2019b. *World Urbanization Prospects: The 2018 Revision*. New York: United Nations.

Üstün, T. B., N. Kostanjsek, S. Chatterji and J. Rehm 2010. *Measuring health and disability: Manual for WHO Disability Assessment Schedule (WHODAS 2.0)*. World Health Organization. https://apps.who.int/iris/handle/10665/43974

Veneri, P. and A. J. E. Edzes 2017. Well-being in cities and regions: Measurement, analysis and policy practices. *Region* 4(2): E1–E5. https://doi.org/10.18335/region.v4i2.188

Vicerra, P. M. M. and W. Pothisiri 2020. Trajectories of cognitive ageing among Thai later-life adults: The role of education using the characteristics approach. *Journal of Population and Social Studies* 28(4): 276–286. https://doi.org/10.25133/JPSSv28n4.019

Weber, D. 2016. Differences in physical aging measured by walking speed: Evidence from the English Longitudinal Study of Ageing. *BMC Geriatrics* 16(1): 31. https://doi.org/10.1186/s12877-016-0201-x

Weeranakin, P. and B. Promphakping 2018. Local meanings of wellbeing and the construction of wellbeing indicators. *Social Indicators Research* 138(2): 689–703. https://doi.org/10.1007/s11205-017-1692-3

World Bank 2019. World Bank Open Data. https://data.worldbank.org/

Yang, J., S. Wang and R. Dewina 2020. *Taking the pulse of poverty and inequality in Thailand*. Washington, D.C.: World Bank Group.

Appendix A.1: Questionnaire

Part A – Demographic characteristics

Item	Questions	Label	Skip to
A1	Time of interview......................................hours......................................minutes		
A2	Sex (Interviewer's observation)	Male.......................................1 Female.......................................2	
A3	Where were you born?	Province.................................. Country.................................. Was born here.....................77	
A4	How long have you lived in this community?	..................................years Don't know.....................999	
A5	When is your birthday?	Month... Don't know.....................99 Year .... Don't know.....................9999	
A6	How old are you?	Age...............................years..............	
A7	At present, can you read? (At least one language including Braille)	Cannot read at all.........................0 Can read with difficulty.........................1 Can read without any problems.........................2	
A8	At present, can you write? (At least one language including Braille)	Cannot write at all.........................0 Can write with difficulty.........................1 Can write without any problems.........................2	
A9	Highest educational qualification	Never been to school.........................000 Level................................	
A10	Marital status	Single (never married)................................1 Married and living together with spouse ....2 Married but spouse not living in the household ....3 Widow/Widower.........................4 Divorced.........................5 Separated.........................6	Skip to A13

Continued

Part A – Continued

Item	Questions	Label	Skip to
A11	How many living children do you currently have?	None ....0 Number of Children............................... NA (single).........................7777	
A12	Any sons or daughters who have died	No children have died.........................0 Number of children who have died .... NA (single).........................7777	
A13	What is your main occupation?	Currently not working......................... 0 Studying .... 1 Please specify occupation	Skip to B1 Skip to B1
A14	Working status	Employer.........................1 Self-employed.........................2 Family business.........................3 Public sector employee.........................4 State enterprise employee.........................5 Private sector employee.........................6 Temporary employee.........................7 Member of cooperative.........................8 Not able to specify working status...............9 Other (specify).........................10	

Part B – Health

Item	Questions	Label	Skip to
B1	Chair rise (Please Test)	Extreme or cannot do.........................1 Able to do with assistance.........................2 None of difficulty.........................3 Refuse to do the test.........................7	

Item	In the past 30 days, how much difficulty did you have in: Please choose the appropriate answers below:	None	Able to do with assistance	Extreme or cannot do	NA
B2	Standing for long periods such as 30 minutes	3	2	1	7
B3	Walking a long distance such as a kilometer [or equivalent]	3	2	1	7
B4	Washing your whole body	3	2	1	7
B5	Getting dressed	3	2	1	7
B6	Taking care of your household responsibilities	3	2	1	7
B7	Joining in community activities (for example, festivities, religious, or other activities) in the same way as anyone else can	3	2	1	7
B8	Concentrating on doing something for 10 minutes	3	2	1	7
B9	Dealing with people you do not know	3	2	1	7
B10	Maintaining a friendship	3	2	1	7
B11	Your day-to-day work/school	3	2	1	7
B12	Learning a new task, for example, learning how to get to a new place	3	2	1	7
B13	How much have you been emotionally affected by your health problems	3	2	1	7

Item	Questions	Label	Skip to
B14	For at least the past six months, to what extent have you been limited because of a health problem in activities people usually do?	Severely limited.........................1 Limited but not severely.........................2 Not limited at all.........................3	
B15	Could you tell us until which age do you think you will live?	Will live until.........................years Don’t know.........................9 Up to god.........................1 No one know when he/she will die.....................2	

Part C – Literacy

Item	Questions	Label	Skip to
C1	Now I would like you to read this sentence to me.	Cannot read at all.........................0	Skip to D1
		Able to read part of the sentence.........................1	
		Able to read almost the whole sentence 2	
		Able to read the whole sentence.........................3	
		Refuse to read.........................8	Skip to D1
		Blind/visually impaired.........................9	Skip to D1

Texts for C1

"ชาวนาเป็นอาชีพที่ต้องทำงานหนัก"	Rice farming is a hard work.
"นักเรียนสอบได้คะแนนยอดเยี่ยม"	The student passed the exam with an excellent score.
"เสร็จแล้วไซร้เดินหน้ายุ่งเข้ามุ้งนอน"	After he finished work, he went to bed.

Item	Question	Label	Skip to
C4	Circle the word that matches the picture: 	car................................1 hand...............................2 moon............................3 hair 4	

Item	Read the sentences below. Circle YES if the sentence makes sense. Circle NO if the sentence does not make sense.	No	Yes
C5	1. ลูกบอลมีทรงสี่เหลี่ยม A ball has a square shape.	1	2
	2. ลูกชายของน้องสาว คือ หลานสาวของเรา The son of my sister is my niece.	1	2
	3. สีเขียว คือ สัญญาณไฟจราจรให้รถไปได้ When the traffic light turns green, cars can go.	1	2
	4. นกเป็นสัตว์ที่บินได้ A bird can fly.	1	2
	5. ข้าวสารหนัก 50 กิโลกรัม หนักกว่าข้าวสาร 70 กิโลกรัม A 50 kilograms bag of rice is heavier than a 70 kilograms bag of rice.	1	2

Item	Question	Label
C6	Can you tell me what day of the week it is?	Day of week given correctly...............................1 Day of week given incorrectly/doesn't know day...............................2

Part D – Economic conditions

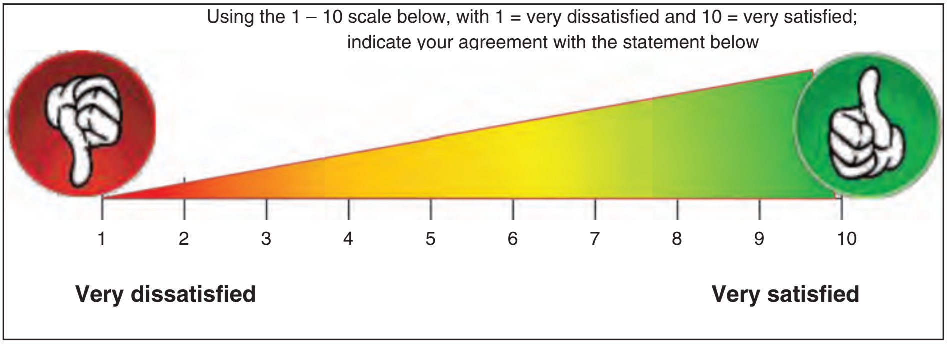
Item	Questions	Label	Skip to
D1	In the past four weeks, did you or any household member have to eat fewer meals in a day because there was not enough food?	Never ....1 Yes 2	Skip to D3
D2	How often did this happen in the past four weeks?	Rarely (once or twice in the past for weeks)....1 Sometimes (three to 10 times in the past for weeks)....2 Often (More than 10 times in the past for weeks) 3	
D3	In the past four weeks, did you or any household member have to eat a limited variety of foods due to a lack of resources?	Never...............................1 Yes 2	Skip to D5
D4	How often did this happen in the past four weeks?	Rarely (once or twice in the past for weeks)....1 Sometimes (three to 10 times in the past for weeks)....2 Often (More than 10 times in the past for weeks)...............................3	
D5	In the past 12 months, how much did you earn in all of your jobs? (including non-monetary income)	...............................Bahts per day Working .... days per month ....Bahts per month ....Bahts per year	
D6	Do you think that your total income is enough for a living?	More than enough...............................1 Enough ....2 Sometimes enough...............................3 Not enough 4	
D7	Do you have a welfare card (low-income card)?	No ....1 Yes 2	
D8	How is the current financial situation of your household?	Very bad...............................1 Bad ....2 Neither good nor bad ....3 Good ....4 Very good 5	
D9	Compared to the past five years (2013), did the financial situation of your household get...?	Much worse ....1 Worse ....2 Styed the same ....3 Better ....4 Much better 5	

Part F – Satisfaction with life

Using the 1–5 scale below, indicate your agreement with each item by placing the appropriate number on the line preceding that item. Please be open and honest in your responding.

Item	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Don't know
F1	In most ways my life is close to my ideal.					
	5	4	3	2	1	9
F2	The conditions of my life are excellent.					
	5	4	3	2	1	9
F3	I am satisfied with my life.					
	5	4	3	2	1	9
F4	So far I have gotten the important things I want in life.					
	5	4	3	2	1	9
F5	If I could live my life over, I would change almost nothing.					
	5	4	3	2	1	9

F6 Taking all things together, how satisfied are you with your life as a whole these days?



F7 What is the most important thing in your life?

First most important thing in my life is..............................
Second most important thing in my life ....
Third most important thing in my life ....

F8 What do you think are the three most serious problems in Thailand? Please rank from one to three.

...............1. Global warming
...............2. Politics
...............3. Foreign migrant labours
...............4. Drugs
...............5. Economics
...............6. Southern border provinces
...............7. Social inequality
...............8. Environment
...............9. Corruptions
...............10. Land ownership
...............11. Other (specify) ....

Part G – Dwelling characteristics and living conditions

Item	Questions	Label	Skip to
G1	Wall or building material	Brick, block, stone or cement...............1 Wood ....2 Half brick, half wood...............3 Makeshift, salvaged or improvised materials from the local area; e.g., plant-based material...............4 Reused materials; e.g., cardboard, scrap. 5 Other (specify)......../........6 Unknown...............9	
G2	What is the general condition of the area immediately around the house with respect to garbage disposal?	Lots of uncollected garbage...............1 Some uncollected garbage...............2 Very little garbage...............3 No garbage visible...............4 Not applicable...............9	
G3	Garbage disposal area	Presence of pests; e.g., cockroaches,rats.1 Bad odour...............2 Clear separation of garbage...............3 Not applicable...............9	
G4	What kind of toilet facility do members of your household usually use?	Flush toilet...............1 Squat toilet...............2 Flush toilet and squat toilet...............3 Pit latrine/Bucket toilet...............4 No toilet facility...............5	
G5	Does your household have a refrigerator?	No...............1 Yes...............2	
	Time interview finished...............hours...............minutes...............		

Appendix A.2: Data and methods

Table A.1:
Indicators and cut-off points used in the main results and in sensitivity analyses

	Health	Cognition	Out of poverty	Life satisfaction
Main YoGL	Can stand up from a chair without assistance	Can read a full sentence	Does not skip meals	Reports life satisfaction score of five or higher
Sensitivity 1	Can stand up from a chair with assistance	Can read a full sentence	Does not skip meals	Reports life satisfaction score of five or higher
Sensitivity 2	Can stand up from a chair without assistance	Can read parts of a sentence	Does not skip meals	Reports life satisfaction score of five or higher
Sensitivity 3	Can stand up from a chair without assistance	Can read a full sentence	Does not skip meals	Reports life satisfaction score of four or higher
Sensitivity 4	Can stand up from a chair without assistance	Can read a full sentence	Does not skip meals	Reports life satisfaction score of six or higher
Sensitivity 5	Positive scores on all five-item ADLs	Can read a full sentence	Does not skip meals	Reports life satisfaction score of five or higher
Sensitivity 6	Positive scores on all 12-item ADLs	Can read a full sentence	Does not skip meals	Reports life satisfaction score of five or higher
Sensitivity 7	GALI: does not report severe limitation	Can read a full sentence	Does not skip meals	Reports life satisfaction score of five or higher
Sensitivity 8	Can stand up from a chair without assistance	Reports being able to read without difficulty	Does not skip meals	Reports life satisfaction score of five or higher
Sensitivity 9	Can stand up from a chair without assistance	Reports being able to write without difficulty	Does not skip meals	Reports life satisfaction score of five or higher
Sensitivity 10	Can stand up from a chair without assistance	Able to answer correctly all reading with understanding questions	Does not skip meals	Reports life satisfaction score of five or higher

Continued

Table A.1: Continued

	Health	Cognition	Out of poverty	Life satisfaction
Sensitivity 11	Can stand up from a chair without assistance	Can read a full sentence	Does not limit variety of foods due to a lack of resources	Reports life satisfaction score of five or higher
Sensitivity 12	Can stand up from a chair without assistance	Can read a full sentence	Does not own a welfare card	Reports life satisfaction score of five or higher
Sensitivity 13	Can stand up from a chair without assistance	Can read a full sentence	Flush toilet available in the house	Reports life satisfaction score of five or higher
Sensitivity 14	Can stand up from a chair without assistance	Can read a full sentence	Does not skip meals	Life close to ideal (reports score of three or higher)
Sensitivity 15	Can stand up from a chair without assistance	Can read a full sentence	Does not skip meals	Satisfied with life (reports score of three or higher)
Sensitivity 16	Can stand up from a chair without assistance	Can read a full sentence	Does not skip meals	Would change almost nothing in life (reports score of three or higher)

Note: Indicators that are different from the main specification are marked in grey.

Table A.2:
Share of respondents by age and sex who meet all YoGL conditions, and who have positive scores in individual dimensions

Sex	Age group	YoGL	Healthy	Cognitively able	Satisfied	Out of poverty
Female	20	0.76	1	0.76	1	0.9
Female	25	0.73	1	0.87	0.93	0.83
Female	30	0.72	1	0.84	0.94	0.91
Female	35	0.53	1	0.74	0.95	0.77
Female	40	0.56	1	0.67	1	0.85
Female	45	0.42	1	0.52	0.96	0.81
Female	50	0.36	0.98	0.43	0.92	0.81
Female	55	0.31	0.97	0.39	0.94	0.81
Female	60	0.29	0.97	0.31	0.98	0.91
Female	65	0.14	0.95	0.17	0.9	0.8
Female	70	0.18	0.82	0.21	0.94	0.82
Female	75	0.24	0.83	0.29	0.95	0.93
Female	85	0.24	0.83	0.29	0.95	0.93
Female	80	0.24	0.83	0.29	0.95	0.93
Male	20	0.6	1	0.7	1	0.9
Male	25	0.91	0.91	0.91	1	1
Male	30	0.67	1	0.67	1	0.83
Male	35	0.79	0.93	0.93	1	0.79
Male	40	0.67	1	0.79	1	0.83
Male	45	0.31	1	0.47	0.87	0.73
Male	50	0.5	0.91	0.71	0.94	0.79
Male	55	0.5	0.98	0.57	0.93	0.91
Male	60	0.33	1	0.43	0.92	0.82
Male	65	0.42	1	0.45	1	0.81
Male	70	0.52	1	0.57	0.96	0.87
Male	75	0.19	0.9	0.24	1	0.71
Male	85	0.19	0.9	0.24	1	0.71
Male	80	0.19	0.9	0.24	1	0.71