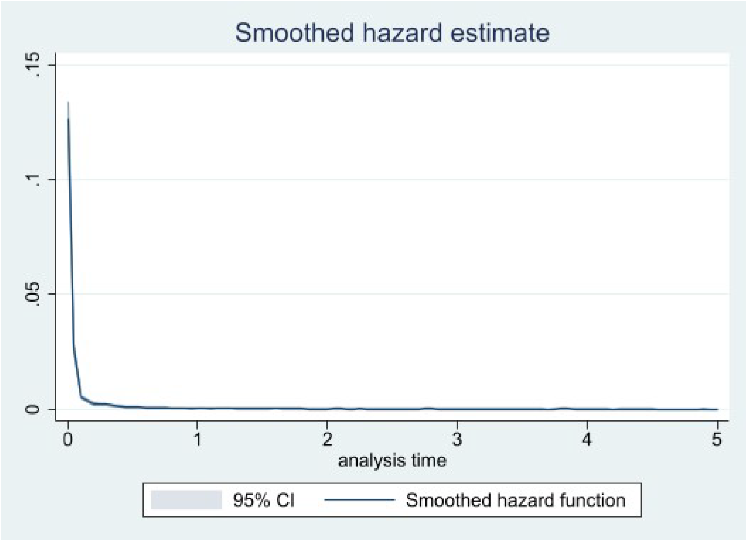
Supplementary Material

Supplement to: Idohou, E., Bocquier, P. And Guillot, M. (2025) Excess under-five mortality of children born to immigrants: longitudinal evidence from France. Vienna Yearbook of Population Research, 23. https://doi.org/10.1553/p-fdgp-bgc3

S1. Weibull model

Figure S.1 U5M, hazard function



Source: Author's calculations based on the permanent demographic sample (EDP), 1990–2020

S2. Propensity score method

The propensity score method is used to balance or make similar the characteristics of two groups in order to compare them on a given dimension. It involves synthesising all these characteristics or conditioning variables into a single variable known as the propensity score. Introduced by Rosenbaum and Rubin (1983), the propensity score defines the probability of receiving treatment based on the observed variables between two groups, with one designated as the treatment group and the other designated as the control group. Through this method, a hypothetical group of study participants is created. In our study, the group of interest (i.e., children of immigrant mothers) is similar to the reference group (i.e., children of native mothers) in terms of some observed characteristics other than migration origin (Ma, 2008).

Choice of covariates

The initial stage of propensity score estimation presents the first challenge. There is no universally accepted rule for selecting appropriate variables. Some authors suggest incorporating into the estimation procedure all variables correlated with both the treatment and the variable of interest (Quantin, 2018).

© The Author(s) 2025
Open Access This article is published under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/) that allows the sharing, use and adaptation in any medium, provided that the user gives appropriate credit, provides a link to the license, and indicates if changes were made.

Conversely, others stress the delicate balance between statistical and theoretical considerations. While the inclusion of a large number of variables may improve the precision of propensity score estimation, it also leads to an increase in the variance of the estimators. On the other hand, omitting crucial variables could introduce bias into the estimates (Lecocq et al., 2016).

The results of the Oaxaca-Blinder (O-B) decomposition, along with the existing theory on immigrant health, guided our selection of variables. Factors such as infant sex, plural birth, place of birth in France, father's SPC, marital status and the age of the father and the mother at the child's birth were considered. In the decomposition analysis (Table S.1), under-five mortality differences between children of immigrant mothers (CIM) and children of native mothers (CNM) are substantially accounted for by these variables, with the exception of the infant's sex and plural birth. However, we have incorporated all of these variables into the construction of the propensity score. It is important to note that these two groups are not directly comparable on average, and the propensity score method will therefore be used to address and correct for these differences.

Table S.1 O-B Two-fold decomposition of U5M (per 10,000 births) gap between CIM and CNM

	Coef	low95	high95	
Immigrant-native gap	19.77	14.98	24.56	
Endowment effect	2.31	0.51	4.12	
Unexplained factors	17.46	12.44	22.47	
	Coef	low95	high95	Variables
	0.00292	-0.0711	0.0769	Sex (Female)
	0.202	0.0258	0.379	Plural birth
	1.229	0.659	1.8	Place of birth (Overseas France)
	0.0719	0.0443	0.0897	Father's SPC+
	-0.719	-2.519	1.082	Parental marital status (Married parent)
	-2.216	-3.314	-1.118	Father's age
	3.739	1.763	5.715	Mother's age

Source: Author's calculations based on the permanent demographic sample (EDP), 1990–2020

Choice of method

Once the variables have been selected, they are incorporated into a model, typically a logistic regression model, with a few exceptions, such as the use of generalised boosted models, a non-parametric multivariate regression technique (McCaffrey et al., 2004). Following estimation, the obtained propensity score facilitates the formation of two groups, namely a treatment group and a control group, employing various methods. The most commonly employed methods include matching, stratification and weighting. However, choosing the most appropriate method can be challenging, as each approach possesses its own advantages and limitations. As highlighted by (Lecocq et al., 2016), no single method supersedes the others, and the selection of the appropriate method depends on the specific characteristics of the dataset at hand (Lecocq et al., 2016; Quantin, 2018).

We have chosen the method of weighting for several reasons. Undoubtedly, the full (with replacement) matching method could have been selected as the optimal choice. Indeed, as Austin (Austin, 2013) and Austin and Stuart (Austin & Stuart, 2015, 2017) have shown using Monte Carlos simulation, this method is the most optimal compared with the others, in particular weighting and partial (without replacement) matching methods. However, considering the limitations of partial matching when applied to our dataset, the substantial sample size and the memory constraints associated with full matching, weighting emerges as a viable alternative. Furthermore, as demonstrated by Austin and Stuart (2015), the weighting method can perform as well as the full matching method. The weighting method is similar to the re-weighting procedures used in survey sampling to take account of observations with varying probabilities of inclusion in the sample (McCaffrey et al., 2003). It involves the creation of a new variable incorporated into a regression model. By assigning higher scores to children of immigrant mothers (CIM), inverse weighting reduces the weight assigned to children of native mothers (CNM) and to CIM with similar observable characteristics, while increasing the weight assigned to CIM. This approach offers the advantage of preserving the entire sample for analysis, and it can produce almost unbiased estimates, particularly for large samples under specific conditions (Austin, 2016; McCaffrey et al., 2003). However, it may encounter challenges stemming from excessively high weights, particularly when certain units exhibit very low propensity scores (Austin & Stuart, 2017). To overcome this issue, truncation can be used. By setting a threshold, this technique addresses the problem of assigning extremely high weights to certain observations. In our analysis, for example, we observed a weight exceeding 30. Consequently, we applied truncation at the 99th percentile, resulting in a maximum weight of approximately 13.

In the weighting method, each unit is assigned a weight, noted w_i,which is inversely proportional to its probability of belonging to one of the two groups. More specifically, for the CNM group, the weight is determined by the inverse of the propensity score, $\mathrm{e}(\mathrm{x_i})$, while for CIM, it is determined by $1-\mathrm{e}(\mathrm{x_i})$.

Thus,

$$\boldsymbol{w_i} = \frac{\boldsymbol{z_i}}{\mathbf{e}(\boldsymbol{x_i})} + \frac{\mathbf{1}-\boldsymbol{z_i}}{\mathbf{1}-\mathbf{e}(\boldsymbol{x_i})}$$

Balancing properties

This section focuses on presenting indicators commonly employed to assess the balancing property of the propensity score. The quality of the propensity score is usually assessed and reported using statistical measures and graphs. The main indicators are mean or standardised proportion differences and variance ratios.

The standardised difference in means is the difference in the means of a (continuous) variable in the CIM and CNM, divided by the sum of the variance of the two groups.

 https://doi.org/10.1553/p-fdgp-bgc3

Thus,

$$\boldsymbol{Diff_1} = \frac{\overline{X}_{Na} - \overline{X}_{Im}}{\sqrt{\frac{s_{Na}^2 + s_{Im}^2}{2}}}$$

where $\overline{X}_{Na}$ and $\overline{X}_{Im}$ are the empirical means of the covariates X in the CNM and CIM groups, respectively, and s_{Na}^2 and s_{Im}^2 the corresponding empirical variances.

When dealing with categorical variables, this refers to the difference in the proportion of a specific category between the two groups under investigation, as demonstrated by the following equation:

$$\boldsymbol{Diff_2} = \frac{\hat{p}_{Na} - \hat{p}_{Im}}{\sqrt{\frac{\hat{p}_{Na}(1-\hat{p}_{Na}) + \hat{p}_{Im}(1-\hat{p}_{Im})}{2}}}$$

where $\hat{p}_{Na}$ et $\hat{p}_{Im}$ are the proportions for a given category.

The standardised difference approach has the advantage of not being affected by sample size, which makes it particularly relevant for matched data (Quantin, 2018). It plays a significant role in comparing differences in observable characteristics before and after adjustment, thereby ensuring the accuracy of the propensity score. To evaluate the balancing property, a visualisation of the distribution of each covariate is employed (not presented here). Covariates are considered unbalanced when the standardised difference or proportion difference is minimal. However, there is no universally recognised threshold for determining the significance of these differences. The choice of a threshold for comparison is ultimately left to the discretion of the analyst (Lecocq et al., 2016).

As previously mentioned, the balancing property of the propensity score extends beyond the mean or proportion alone. It encompasses the entire distribution of covariates. Therefore, it is suggested that additional indicators, such as variance ratios (Table S.2), are considered. These ratios are employed to assess the distributional shape of covariates beyond their central tendencies. The variance decreases as the average difference in the variable between the two groups decreases. Variance ratios close to one mean that the variances are (almost) equal in the two groups, and therefore indicate a balance in these groups (Rosenbaum & Rubin, 1983). It is worth noting that significant differences in shape may indicate a suboptimal balance even if mean differences and variance ratios are below certain thresholds (Greifer, 2022).

Table S.2 summarises some key indicators before and after adjustment (weighting). These indicators include the standardised difference in means or proportions (Diff) and the variance ratios (V.ratio). The terms "Un" and "Adj" represent abbreviated forms of "Unadjusted" and "Adjusted", respectively, denoting the corresponding values. It is clear from this table that the application of weighting has led to an enhancement in the balance across all variables. This improvement is reflected by bringing all variables below the threshold of 0.1 for standardised mean differences or proportions, although some minor imbalances remain. In particular, a more substantial balance is achieved for the father’s age and the parental marital status, which exhibits a much higher degree of balance.

Table S.2 Some balancing indicators

Socio-economic characteristics	Type	Before adjustment		After adjustment (weighting)	
		Diff	V. Ratio	Diff	V. Ratio
Infant sex	Binary	-0.00001	NA	-0.001	NA
Plural birth	Binary	-0.00522	NA	-0.00466	NA
Place of birth of the infant in France	Binary	-0.09486	NA	-0.03910	NA
Mother's age	Continuous	-0.17657	0.83819	-0.03851	0.81105
Father's age	Continuous	0.463925	0.60387	0.00473	0.93160
Parental marital status	Binary	-0.43487	NA	-0.00514	NA
Father's SPC	Binary	0.05330	NA	0.03619	NA
Propensity score	Continuous	0.88730	0.43416	0.01119	0.94219

Note: Variance ratio is only computed for continuous variables.
Source: Author's calculations based on the permanent demographic sample (EDP), 1990–2020

References

Austin, P. C. (2013). The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in Medicine*, 32(16), 2837-2849. https://doi.org/10.1002/sim.5705

Austin, P. C. (2016). Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. *Statistics in Medicine*, 35(30), 5642-5655. https://doi.org/10.1002/sim.7084

Austin, P. C., and Stuart, E. A. (2015). Optimal full matching for survival outcomes: A method that merits more widespread use. *Statistics in Medicine*, 34(30), 3949-3967. https://doi.org/10.1002/sim.6602

Austin, P. C., and Stuart, E. A. (2017). The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes. *Statistical Methods in Medical Research*, 26(4), 1654-1670. https://doi.org/10.1177/0962280215584401

Greifer, N. (2022). *Covariate Balance Tables and Plots: A Guide to the cobalt Package* (Version R package version 4.4.0). https://cran.r-project.org/web/packages/cobalt/vignettes/cobalt.html

Lecocq, A., Ammi, M., and Bellarbre, É. (2016). Le score de propension : Un guide méthodologique pour les recherches expérimentales et quasi expérimentales en éducation. *Mesure et évaluation en éducation*, 37(2), 69-100. https://doi.org/10.7202/1035914ar

Ma, S. (2008). Paternal Race/Ethnicity and Birth Outcomes. *American Journal of Public Health*, 98(12), 2285-2292. https://doi.org/10.2105/AJPH.2007.117127

McCaffrey, D. F., Ridgeway, G., and Morral, A. R. (2004). Propensity Score Estimation With Boosted Regression for Evaluating Causal Effects in Observational Studies. *Psychological Methods*, 9(4), 403-425. https://doi.org/10.1037/1082-989X.9.4.403

Quantin, S. (2018). *Estimation avec le score de propension sous R*. (Série des documents de travail « Méthodologie Statistique » 2018/01). Insee. https://www.insee.fr/fr/statistiques/3546202