Assessing Bicycle Safety in Multiple Networks with Different Data Models

Martin LOIDL and Bernhard ZAGEL
Department of Geoinformatics, University Salzburg / Austria · martin.loidl@sbg.ac.at

This contribution was double-blind reviewed as full paper.

Abstract

Safety is a key issue in the context of bicycle promotion strategies. Because such initiatives rely on a sound data basis, the road network's quality is usually assessed with a focus on bicycle safety. In this paper the indicator-based assessment model is presented as an efficient alternative to commonly used assessment models. The applicability of this GIS approach is demonstrated in a case study, where the model is applied on two different, but adjacent network data sets with diverse data models.

1 Safety Issues and Bicycle Usage

Transportation systems, especially in urban environments, tend to collapse due to increasing traffic loads and limited capacities. Apart from this, the negative side effects of a high share of motorized traffic in the modal split are constantly growing: urban areas suffer from negative ecological, economic and social impacts, such as pollutant emission (KEUKEN et al. 2005), externalities of e.g. traffic congestions (LIEB & SOMMER 2007, TIMILSINA & DULAL 2011), or social-environmental injustice (LAKES et al. 2013, LAUSSMANN et al. 2013). The limited capacities, in conjunction with the negative impacts, have brought the bicycle into the focus of planners, authorities and researchers as an urban transportation alternative (MESCHIK 2012). Countless studies have been praising the bicycle as a sustainable, cost-efficient mode of transport (SÆLENSMINDE 2004, NEWMAN & MATAN 2012, RUTTER et al. 2013), and large investments have been made in order to build or improve bicycle infrastructure. These investments have definitely led to a rising number of cyclists (see e.g. PUCHER et al. 2010). But still, there are some parameters which keep people from using their bicycle as an everyday mode of transportation. Besides the weather, travel comfort, or simply car-centred mobility habits, safety concerns are a valid objection to bicycle usage (WEGMAN et al. 2012).

This fact has been widely anticipated in research and transportation politics. THOMAS & DEROBERTIS (2013) state that only if the real and perceived safety risks are tackled, the benefits of an increasing bicycle share can be realized. This assumption is backed by numerous studies which have proven that cycling is a healthy mode of transport (see e.g. HOLM et al. 2012), but, in relation to the overall travel distance, a rather dangerous one (JUHRA et al. 2012).

Vogler, R., Car, A., Strobl, J. & Griesebner, G. (Eds.) (2014): GI_Forum 2014. Geospatial Innovation for Society.
© Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach. ISBN 978-3-87907-545-4.
© ÖAW Verlag, Wien. eISBN 978-3-7001-7652-7, doi:10.1553/giscience2014s144.

This limitation inevitably leads to the situation that on the one hand the demand for sustainable transportation systems exists, and on the other hand the bicycle would be an ideal alternative, but safety issues hamper an expanded bicycle usage to a certain degree. Hence, in the wake of the UN decade of action for road safety (www.un.org/en/roadsafety), the World Health Organization calls for safer bicycle infrastructures as an incremental part of the promotion of sustainable transportation systems: *"Policies to encourage walking and cycling need additional criteria to ensure the safety of these road users. [...] Promoting city cycling to reduce congestion cannot be encouraged if cyclists repeatedly find that their lanes cut across oncoming traffic."* (WHO 2013, 30).

In order to follow this recommendation, strategies for bicycle promotion must address the need for a high-performance and safe infrastructure, as well as user-specific information offers, with special regard to road safety. For both aspects, a solid data basis is of crucial importance. Here, geographic information systems (GIS) can significantly contribute to consolidating existing data, as well as to analyse them, and, finally, to add extra value through advanced modelling approaches. In the following, the focus is on road network quality assessment with regard to bicycle safety. The result of such an assessment can be used for status-quo analysis, for efficient planning processes and for information purposes, such as routing applications.

The rest of this paper is structured as follows: in chapter 2 the shortcomings of three commonly used methods for quality assessment are briefly discussed. Building on that, a GIS-based model approach is presented. In a case study, the applicability of this approach even for a heterogeneous data basis is demonstrated. The paper closes with a short conclusion and an outlook.

2 Assessing Bicycle Safety Risks

Road networks are commonly maintained by public authorities, who are also responsible for ensuring and improving road safety. GI systems facilitate a structured administration of all relevant road-related data. Generally the road network is stored as a geometrical dataset (graph consisting of lines = segments and nodes = junctions). Related information can then be attached as attributes to any network element. For the determination of the network's quality in terms of road safety, three main approaches are currently applied: expert assessment, accident analysis or aggregated user feedback.

2.1 Expert Assessment

Based on a catalogue of criteria, or on individual judgments, experts assess the quality of single road segments. Such approaches can result in index values, ordinal rankings, or simply verbal descriptions for every considered road segment.

From a modelling and analysis perspective, expert assessments are problematic for at least three reasons:

1. The result heavily depends on the personal judgment of the expert. This is especially the case when no – or only a rudimentary – catalogue is used. In practice, grades or verbal descriptions are most often used. Such an approach can lead to fuzzy classification results which are only of limited use for further analysis.

2. An assessment standard independent from time, location and person can hardly be established. This leads to heterogeneous, non-comparable results for different areas or time intervals.
3. A re-assessment is necessary after every physical modification of the road space. Thus, depending on the update intervals, the expert assessment approach requires by far the most effort.

Expert knowledge is essential in any assessment approach at different stages (from parameter selection to the analysis of results). But for a global approach the results from expert assessments are too vague, especially if the required effort is taken into account.

2.2 Accident Analysis

The deduction of quality measures from accident locations is a quite common approach (e.g. KIECHLE et al. 2013), although it is based on a conceptual fallacy. From the absolute number of accidents[1] (for an extensive discussion of the problematic data availability see ITF (2012)), one can neither conclude that certain locations are unsafe, nor that segments with fewer accidents are safer. As the number of accidents is related to the bicycle load, no conclusions about the quality of the road segment can be drawn. The same holds true if temporal effects, such as construction work, account for an increase of accidents.

Even if these false conclusions are not considered, one cannot deduce that road segments without or with fewer accidents are safer than others, as the following example illustrates:

> *On a cycle way with a high load of bicycles a certain amount of accidents per year are reported. A primary road with very much motorized traffic and without any bicycle infrastructure runs parallel to this cycle way. This road is avoided by most bicyclists (= low bicycle load). There are no accidents reported where bicyclists are involved (= low number of accidents). A quality assessment which is exclusively based on reported accident locations, would falsely rank the primary road higher than the cycle way.*

The analysis of accident locations is a very important task. It allows for better understanding of risk factors and it can help to detect and mitigate accident hotspots. But it cannot be used for a global road network quality assessment.

2.3 User Feedback

Using collective direct or indirect (e.g. Twitter messages) user feedback for quality assessment is a promising approach. Nevertheless there are at least three drawbacks which need to be considered in this context:

1. Feedback is not equally distributed over space. For central areas the number of feedback messages might be sufficient, but for a global application the sample size is too small.

[1] Because a statistical population (= number of bicyclists on the road) can hardly be estimated for any larger area, a risk exposure cannot be soundly calculated. What can be found in literature are rough estimations or statistical populations for only a few streets.

2. Voluntary user feedback is biased in several ways (see BROWN et al. (2014), BOYD & CRAWFORD (2012) for an extensive discussion). The sample is not representative, and thus can hardly be used for general assessment purposes.
3. In order to transfer verbal, qualitative feedback into reliable conclusions, sophisticated (semantic) algorithms are required.

User feedback is very helpful for the validation and calibration of assessment models. Furthermore it can help to improve infrastructure very efficiently. But a global quality assessment that relies entirely on user feedback is not reliable, and thus applicable for the reasons mentioned above.

3 Indicator-based Assessment Model

GI systems offer unique possibilities for the assessment of spatial entities, in this case road networks. Within a GIS, road-related data can be not only stored, but also modelled and analysed. Based upon this ability, an indicator-based assessment model with regard to bicycle safety was initially developed by LOIDL & ZAGEL (2010). This approach has been constantly refined and optimized in order to make it applicable on a meta-level, with different underlying data sources and models.

Fig. 1:
Conceptual difference and interaction between conventional assessment approaches and the indicator-based assessment model

The basic idea of the indicator-based assessment model is to identify those indicators which contribute to the potential risk, such as presence and design of bicycle infrastructure or motorized traffic load. These parameters are then weighted and compiled in a GIS-model. Assessment methods such as those described in the previous section are used for validation and calibration (see figure 1).

3.1 Model Development

The initial point of the model development is to identify factors ("indicators") which contribute to potential safety threats or accidents. This is typically done by literature review, expert, and user interviews, and/or systematic accident analysis (see figure 2).

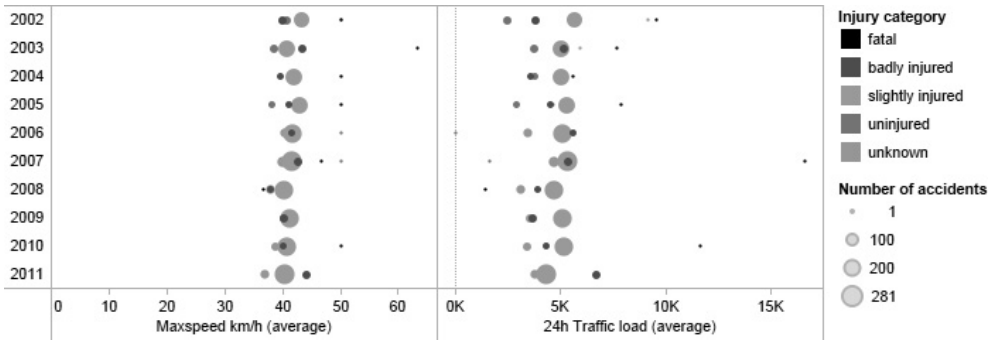

Fig. 2: Results of a visual bicycle accident analysis (n = 3.096) regarding maximum speed and traffic load

In a further step, the digital equivalents to these real-world physical indicators are determined. In this context the data model and attributive structure of the underlying data source are key factors, as will be demonstrated in the case study.

Based on empirical studies (e.g. HARKEY et al. 1998), expert knowledge, and user experiences, weights are assigned to each indicator. These weights express the relative influence of each indicator on the overall quality (in this context the safety risk). Depending on the environment, data availability, or user's preferences, the weights can be adjusted.

In the final assessment model the selected indicators with their respective weights are compiled. The model's design is linear and results in a dimensionless index value, where low values express a low safety risk and vice versa. The linear design allows for adding or removing indicators without any effect on the model's technical performance. Generally the more indicators that are used, the higher the explanatory power (given that the indicators are not redundant).

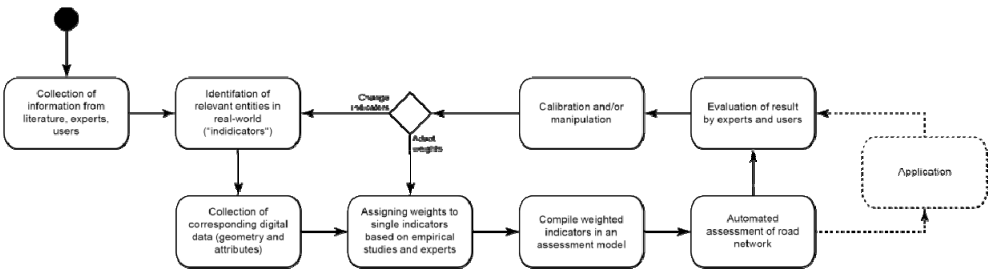

Fig. 3: Iterative workflow of model development (section 3.1), application, evaluation and calibration (section 3.2)

3.2 Model Application, Evaluation and Calibration

The indicator-based assessment model can be applied in any GIS for the calculation of the index value for every road segment. The computed result is then evaluated by experts and

users. If necessary, the model can be iteratively adapted either on the level of the indicators or the weights.

The main advantages of this modelling approach for quality assessment are:

- Transparency: the composition of the resulting index value can be traced back to the very smallest building block of the model.
- Comparability: due to the global application of the model the results within the area of interest can be easily compared.
- Adaptability: through the linear model design and the weights, the model can be adapted to any environment, data availability or user preferences. Thus it is easily transferable (in a geographical and thematic sense) and geographically scalable.
- Reproducibility: Once the model is compiled it can be integrated in automatic assessment workflows. This allows for short update intervals and employment in simulation routines.

Generally the indicator-based assessment model is designed for an application on a homogenous data basis. This means, that the data model and the attributive structure is consistent. In the following case study a workflow for the application of the assessment model for two networks with different data models is presented.

4 Case Study

Potential safety threats are different in urban and rural environments, respectively. As the road space in urban areas is much more complex, more indicators need to be considered than in rural areas. In return, the impact of single indicators, such as the maximum speed for motorized vehicles, is higher on country roads. For a seamless quality assessment across different environments, the indicator-based assessment model has to be adapted and integrated into a multi-stage workflow. The following case study is taken from an applied research project, where the results of the assessment were used as impedance in a cross-border online routing application for bicyclists (www.radlkarte.eu).

4.1 Study Area and Data

The study area is located in the Austrian-German border region around the city of Salzburg (see figure 4). The indicator-based assessment model has been applied to the whole road network. Apart from the different environments (urban/rural), network datasets from two data sources with different data models had to be incorporated.

1. For the city of Salzburg an authoritative dataset from the city administration was used. The data model of this data set is rather basic. For every road only the centre-line is represented as an edge, no matter whether there are several, physically separated lanes (e.g. main road with separated cycle way) or not. The cross-section of the road is represented on the attribute level. The network consists of 11,458 edges and has a total length of 1,119.7 km.
2. For the five other municipalities an OpenStreetMap (OSM) extract from August 2013 was used. In OSM the cross-section of the road is geometrically represented, this

means, that every physically separated road or lane is stored as an individual edge. The network in the final version[2] consists of 9,601 edges and has a total length of 941.3 km.

Apart from the geometrical representations, the attributes are structured differently in the two networks. While the authoritative graph is well structured, homogenous and complete, the OSM graph exhibits all characteristics of crowd sourced data (GRASER et al. 2013, SPIELMAN 2014), such as incompleteness or attributive inconsistencies.

In order to apply the indicator-based assessment model on a heterogeneous data basis, a meta-level for indicator definition and modelling was introduced. Here it was possible to reflect the different environments, and to harmonize the geometrical and attributive diversities.

Fig. 4: Location of the study area (left) with different data sources. The data model of the official data (dark green/grey) and OSM (light green/grey) are fundamentally different (right): e.g. separated lanes are geometrically represented in the OSM data, whereas in the official data set they are attributes of the centre-line edge.

4.2 Indicator-based Assessment Model Application

The challenge of modelling across different data models is twofold: first, the topology in a harmonized data set needs to be established, and, secondly, the indicators and their corresponding digital representations have to be identified, weighted, and compiled. Therefore, the following workflow (see figure 5) was established:

1. Define transfer points where the two networks are merged. In order to minimize the number of such points the highway bypass around the city of Salzburg with its underpasses and the river Saalach with its bridges was used for the definition.

[2] In order to make the graph routable, some geometrical post-processing (edges were separated and nodes introduced at every junction) was necessary.

2. Apply line matching and integrate transition edges where necessary. For the merge of the two networks a line matching algorithm was applied where the number of edges was equal. Where this was not the case, transitional edges (see figure 6) were introduced in order to connect the edges and ensure a correct topology. These edges have no impedance and are thus irrelevant for the cost optimization in the routing.
3. Identify relevant indicators for urban and rural environments. For the urban environment, the authoritative data set was used. Rural areas were covered by the OSM data set.
4. Find corresponding attributes which code for the indicators in the respective data sources. Since the data model and the attributive structure are very different for the two data sets, one and the same indicator is represented in various ways.
5. Adapt the indicator-based assessment model for the rural and urban area respectively (compare to the workflow in figure 3). The risk scenarios for bicyclists very much depend on the environment. Whereas parking cars in towns are e.g. a potential safety threat, they are irrelevant in the countryside. In return, the maximum speed differs in rural areas, and is generally higher compared to urban environments where the speed limits are more or less constant etc. Additionally, the data availability in the two data sets influenced the composition of the model.
6. Run the model and compute the index value. In order to make the index values for the two sub-regions comparable, a z-transformation is included.

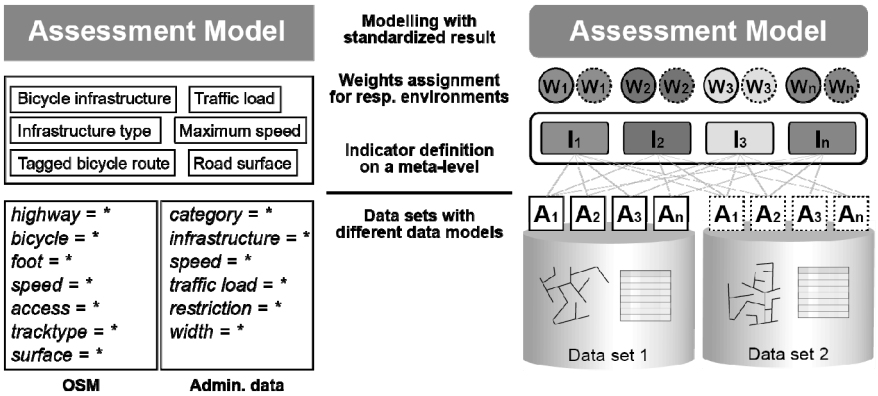

Fig. 5: Workflow for applying the indicator-based assessment model on diverse data sets: example from case study (left), generic concept (right)

The described workflow can be implemented in any GIS application. For the case study presented here, the whole processing was scripted in Python and run in an ESRI ArcGIS 10.1 environment.

4.3 Results

The result of the application of the indicator-based assessment model on a meta-level is a seamless network graph with a harmonized index value. This value indicates the potential safety risk for bicyclists for every road segment (see figure 6). Compared to common assessment methods, this approach has been proven to be much more efficient and sustainable (one setup allows for an arbitrary number of repetitions). With the application on a meta-level, the model can even be simultaneously applied on different data models with a heterogeneous attributive structure.

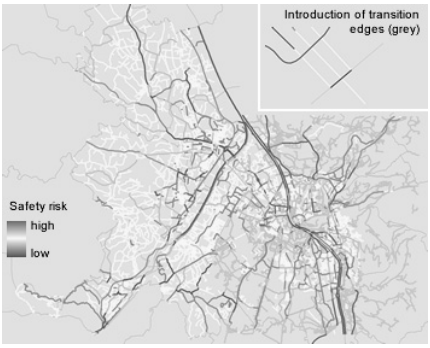

Fig. 6:
Result of a global road network assessment

The resulting index value can be used in various ways. Maps, such as the one depicted in figure 6, can serve as an objective basis for status-quo analysis. As the workflow of the assessment model is automated within a GIS, it can be used for simulation purposes: the effect of infrastructural interventions on the whole network can be tested in advance, and thus help to allocate budgets more efficiently. A third application example would be the use of the index value as impedance in a bicycle routing, which prefers potentially safe roads instead of the fastest or shortest route; in such a context, the integration of real-time information is possible through the modelling approach.

Although the indicator-based assessment model has been proven to be an applicable approach, generating valid results has two major downsides – which are in turn subjects for further research – nevertheless should be mentioned here. Firstly, a global model can hardly generate exact results for every part of the area under consideration. Models by definition exhibit a certain amount of fuzziness (caused, e.g., by data biases or wrong model parameters). The further improvement of the results is a major research goal. Secondly, the quality of the assessment results fundamentally depends on the data availability and quality. This is especially relevant in the context of crowd sourced spatial data. How to deal with biased and/or incomplete data in a modelling context is a highly relevant research question.

Acknowledgements

Part of this research was funded by the national “Klima und Energiefond” (KLIEN) and the EU INTERREG programme within the project “Grenzüberschreitende Radlkarte” (available at www.radlkarte.eu). Thanks to Robin Wendel who did most of the scripting and to Susanna Loidl for proof reading.

References

BOYD, D. & CRAWFORD, K. (2012), Critical Questions for Big Data. Information. Communication & Society, 15 (5), 662-679.

BROWN, G., KELLY, M. & WHITALL, D. (2014), Which 'public'? Sampling effects in public participation GIS (PPGIS) and volunteered geographic information (VGI) systems for public lands management. Journal of Environmental Planning and Management, 57 (2), 190-214.

GRASER, A., STRAUB, M. & DRAGASCHNIG, M. (2013), Ein systematischer Vergleich der Straßennetzwerke von GIP und OpenStreetMap im Großraum Wien. In: STROBL, J., BLASCHKE, T., GRIESEBNER, G. & ZAGEL, B. (Eds.), Angewandte Geoinformatik 2013. Berlin/Offenbach, Wichmann, 424-433.

HARKEY, D. L., REINFURT, D. W., KNUIMAN, M., STEWART, J. R. & SORTON, A. (1998), Development of the bicycle compatibility index: a level of service concept. Washington D.C., Federal Highway Administration, 116 p.

HOLM, A. L., GLÜMER, C. & DIDERICHSEN, F. (2012), Health Impact Assessment of increased cycling to place of work or education in Copenhagen. BMJ Open, 2 (4), 10.1136/bmjopen-2012-001135.

ITF (2012), Cycling Safety: Key Messages. International Transport Forum – Working Group on Cycling Safety. TØRSLØV, N., OECD.

JUHRA, C., WIESKÖTTER, B., CHU, K., TROST, L., WEISS, U., MESSERSCHMIDT, M., MALCZYK, A., HECKWOLF, M. & RASCHKE, M. (2012), Bicycle accidents – Do we only see the tip of the iceberg?: A prospective multi-centre study in a large German city combining medical and police data. Injury, 43 (12), 2026-2034.

KEUKEN, M., SANDERSON, E., VAN AALST, R., BORKEN, J. & SCHEIDER, J. (2005), Contribution of traffic to levels of ambient air pollution in Europe. In: KRZYZANOWSKI, M., KUNA-DIBBERT, B. & SCHNEIDER, J. (Eds.), Health effects of transport-related air pollution. Copenhagen, World Healt Organization, 53-84.

KIECHLE, G., MARKMILLER, T. & OBERMAYER, M. (2013), SecureRouting – sicheres Routing unter Verwendung von Unfallhäufungs- und Wetterdaten. In: STROBL, J., BLASCHKE, T., GRIESEBNER, G. & ZAGEL, B. (Eds.), Angewandte Geoinformatik 2013. Berlin/Offenbach, Wichmann, 444-449.

LAKES, T., BRÜCKNER, M. & KRÄMER, A. (2013), Development of an environmental justice index to determine socio-economic disparities of noise pollution and green space in residential areas in Berlin. Journal of Environmental Planning and Management, 1-19.

LAUSSMANN, D., HAFTENBERGER, M., LAMPERT, T. & SCHEIDT-NAVE, C. (2013), Social inequities regarding annoyance to noise and road traffic intensity. Results of the German Health Interview and Examination Survey for Adults. Bundesgesundheitsblatt, 56 (5/6), 822-831.

LIEB, C. & SOMMER, H. (2007), Externe Kosten im Strassenverkehr – Grundlagen für die Durchführung einer Kosten-Nutzen-Analyse. Bern, Schweizerischer Verband der Strassen- und Verkehrsfachleute, 115 p.

LOIDL, M. & ZAGEL, B. (2010), Wie sicher ist sicher? – Innovatives Kostenmodell zur Ermittlung des Gefährdungspotenzials auf Radwegen. In: STROBL, J., BLASCHKE, T. & GRIESEBNER, G. (Eds.), Angewandte Geoinformatik 2010. Berlin/Offenbach, Wichmann, 394-403.

MESCHIK, M. (2012), Reshaping City Traffic Towards Sustainability Why Transport Policy should Favor the Bicycle Instead of Car Traffic. Procedia – Social and Behavioral Sciences, 48, 495-504.

NEWMAN, P. & MATAN, A. (2012), Human mobility and human health. Current Opinion in Environmental Sustainability, 4 (4), 420-426.

PUCHER, J., DILL, J. & HANDY, S. (2010), Infrastructure, programs, and policies to increase bicycling: An international review. Preventive Medicine, 50, S106-S125.

RUTTER, H., CAVILL, N., RACIOPPI, F., DINSDALE, H., OJA, P. & KAHLMEIER, S. (2013), Economic Impact of Reduced Mortality Due to Increased Cycling. American journal of preventive medicine, 44 (1), 89-92.

SÆLENSMINDE, K. (2004), Cost-benefit analyses of walking and cycling track networks taking into account insecurity, health effects and external costs of motorized traffic. Transportation Research Part A: Policy and Practice, 38 (8), 593-606.

SPIELMAN, S. E. (2014), Spatial collective intelligence? Credibility, accuracy, and volunteered geographic information. Cartography and Geographic Information Science, 1-10.

THOMAS, B. & DEROBERTIS, M. (2013), The safety of urban cycle tracks: A review of the literature. Accident Analysis & Prevention, 52, 219-227.

TIMILSINA, G. R. & DULAL, H. B. (2011), Urban Road Transportation Externalities: Costs and Choice of Policy Instruments. The World Bank Research Observer, 26 (1), 162-191.

WEGMAN, F., ZHANG, F. & DIJKSTRA, A. (2012), How to make more cycling good for road safety? Accident Analysis & Prevention, 44 (1), 19-29

WHO (2013), Global Status Report on Road Safety 2013 – Supporting a Decade of Action. Geneva, World Health Organization, 318 p.