Comparison of Roundabout Accident Prediction Models: Challenges of Data Collection, Analysis and Interpretation

Jiří AMBROS[a,b]*, Petr SLABÝ[c]

[a] *Centrum dopravního výzkumu, v.v.i. (Transport Research Centre), Division of Road Safety and Traffic Engineering, Líšeňská 33a, 636 00 Brno, Czech Republic, jiri.ambros@cdv.cz*
[b] *Czech Technical University in Prague, Faculty of Transportation Sciences, Department of Transporting Systems, Konviktská 20, 110 00 Prague 1, Czech Republic, ambroji1@fd.cvut.cz*
[c] *Czech Technical University in Prague, Faculty of Civil Engineering, Department of Road Structures, Thákurova 7, 166 29 Prague 6, Czech Republic, slaby@fsv.cvut.cz*

Abstract

The paper presents methods of accident studies and introduces an analysis based on prediction models. In an example, roundabout accident prediction models are compared. To this end available Czech data sets have been used; a comparison with several models from abroad was conducted as well. In addition to methodology and example description a number of challenges are stressed, including data collection, analysis and interpretation. Various ways to overcoming these challenges are mentioned, including pros and cons of specific alternatives.

Keywords: accident prediction model, roundabout, methodology, data

1. INTRODUCTION

Road safety and traffic engineering experts have traditionally been interested in accidents on intersections. Various levels of accident analysis are available. Three of them will be described in the following paragraphs.

1.1. National traffic police accident statistics

Data from annual Czech traffic police reviews [12] may be used for this “macro” analysis. Among others they report national numbers divided between 3-arm intersections, 4-arm intersections, intersections with 5 and more arms and roundabouts. The following graph (Fig. 1) shows the number of all reported accidents from last 5 available reviews (2007 – 2011).

* Corresponding author.

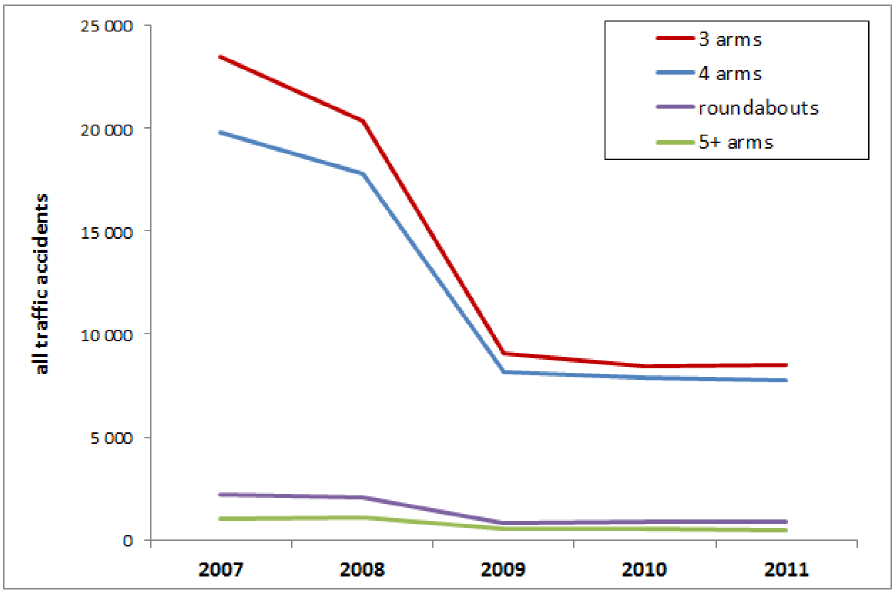

Fig. 1. Number of all accidents on specific intersection types [12]

The graph shows a significant change in 2009 – it is due to an increased change of accident reporting threshold which caused higher underreporting (for more information see e.g. [5]). Since this change influenced mainly the number of property damage only accidents, its impact may be reduced by considering just injury accidents. Fig. 2 shows that underreporting is lower in the case of these accidents and their numbers are relatively stable.

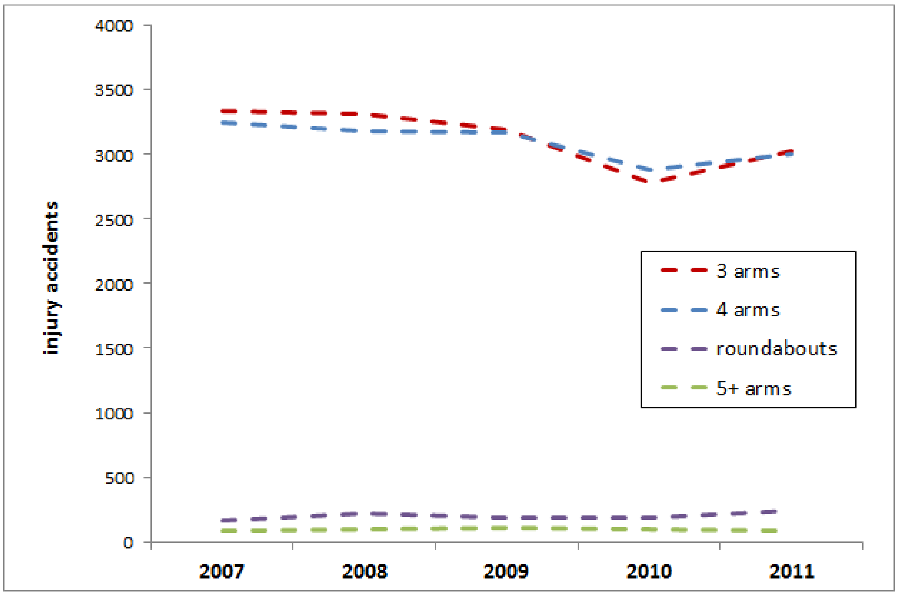

Fig. 2. Number of injury accidents on specific intersection types [12]

Nevertheless, disadvantage of this statistics is that the specific numbers do not allow comparisons of intersection types: this is because the numbers represent total numbers which reflect uneven distribution of intersection types on the Czech road network. Since the police statistics do not comprise the number of intersections, the values cannot be equalized and compared. What is more, traffic volumes on specific intersections are not reported, as well as signalization etc.

1.2. Accident analysis of selected intersections

The second level of analysis is based on a data set, representing selected intersections of specific types. Such analysis was conducted for example in a research project BESIDIDO, which will be mentioned in chapter 3. Graph in Fig. 3 presents some of its results: it compares the average annual accident frequency for each intersection type.

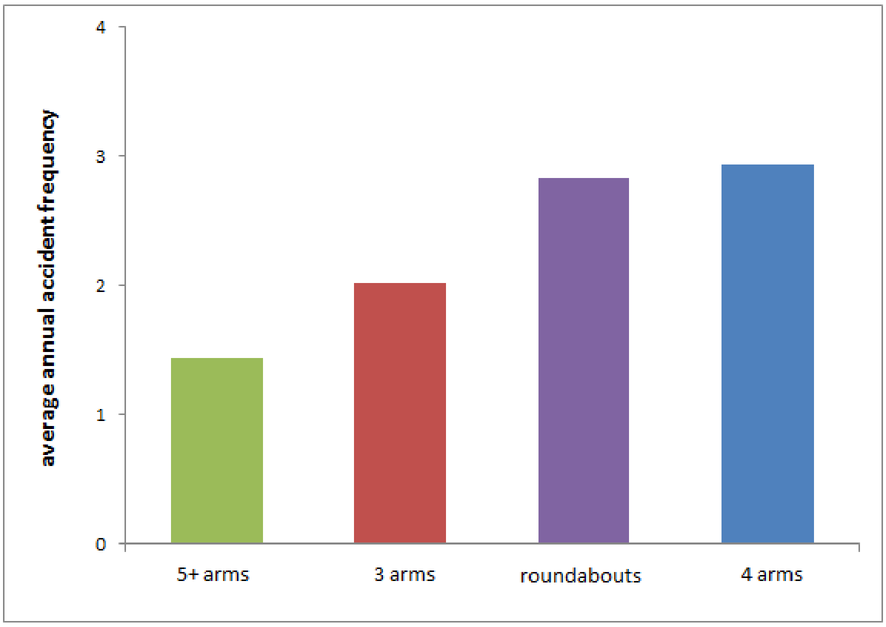

Fig. 3. Average annual accident frequency on specific intersection types [18]

Compared to the first method of analysis, the results are already equalized and may thus be used to compare the safety of intersection types. However, the explanatory power is limited by the size of a used sample.

1.3. Analysis with accident prediction models

The third level analysis employs accident prediction models, created from larger data sets. While previous methods use a retrospective analysis, prediction models enable to forecast the potential future state, i.e. prospective analysis. With a prediction model it is possible to not only compare each represented type, but also to study the impact of specific explanatory variables on accidents. The example is shown in Fig. 4: it compares accident prediction models of unsignalized 3- and 4-arm intersections in Prague [25]. The graph shows the predicted (expected) annual accidents versus traffic volume (AADT) in terms of total daily number of incoming vehicles.

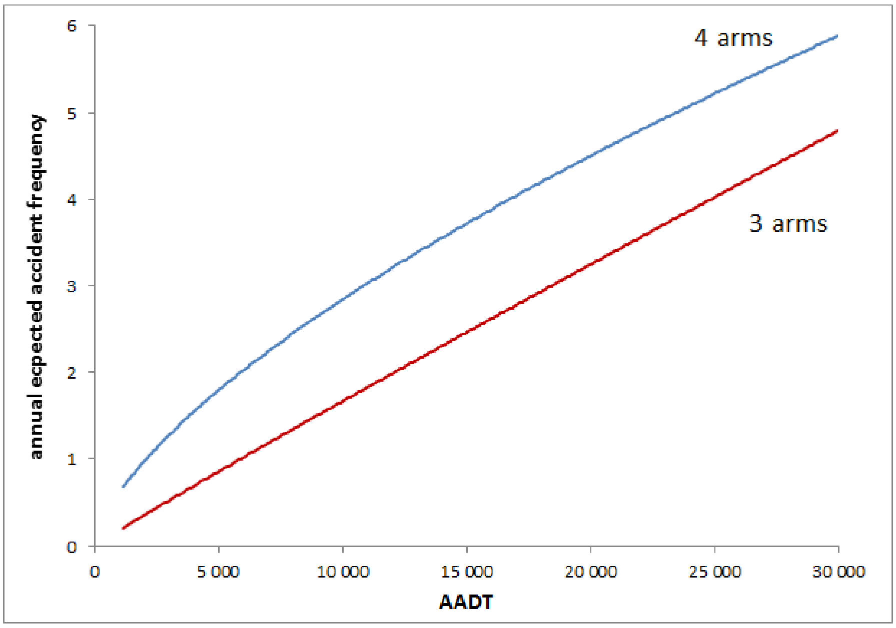

Fig. 4. Comparison of annual expected accident frequency on unsignalized 3- and 4-arm intersections in Prague [25]

Given the three mentioned accident analysis methods, the third way offers apparently the most details. This is why this method was selected for the presented paper. It aims to compare roundabout accident prediction models. Data from several research projects were used and the results were compared with some models developed in the world. These comparisons also bring a number of challenges, which will be summarized in the end, together with alternatives of possible solutions.

2. ACCIDENT PREDICTION METHODOLOGY

Prediction methodology consists of several steps which are briefly described next, for the case of intersections. For more information see for example recent Czech guidelines for the identification of hazardous road locations [8].

2.1. Selection of variables and data collection

Data availability is the main limitation: area-wide databases (secondary data) or own data collection (primary data) may be used. Data should contain the most significant variables which impact and explain accident frequency (according to previous research), are available and are not excessively correlated with each other. These explanatory variables, describing the intersection characteristics, are the independent variables (predictors). Dependent variable is accident frequency in a specific time period.

2.2. Selection of model form

The recommended function form is $E(\lambda) = \alpha \cdot AADT^{\beta} \cdot e^{\sum \gamma_i x_i}$, where $E(\lambda)$ is the expected accident frequency is a function of AADT and further explanatory variables x_i. The objective of modelling is to estimate the values of regression parameters α, β, γ.

2.3. Accident modelling

Modelling comprises selection of statistically significant explanatory variables and the estimation of values of their regression parameters. Since the accident frequency is a discrete variable with non-normal probability distribution, ordinary linear regression may not be used. Models are thus developed using generalized linear modelling. This procedure is available in a number of statistical software packages; it also allows to choose between various probability distributions of residuals and develop multivariate models.

2.4. Model quality assessment

Model quality may be assessed by several ways, e.g. checking the distribution of residuals (they should fluctuate randomly around zero), comparing information criteria (the smaller, the better), etc.

The frequency of reported accidents, which is subject to various biases, such as the mentioned underreporting, should be substituted by the expected accident frequency estimated with prediction models (see e.g. [3, 5]).

These procedures have been known world-wide for a number of decades; however this is not the case in the Czech Republic. One may ask: Does it mean that Czech studies up to now are not usable? In search for an answer to this question, the following study was produced.

3. COMPARISON OF CZECH MODELS

There have been several Czech research projects dealing with roundabout safety and accidents. The authors of the paper had three following data sets available:

- data on conversions from 4-arm intersections to roundabouts from a project of Czech Technical University in Prague, Faculty of Civil Engineering (GAČR [18, 19])
- data on other roundabout conversions from a project of Transport Research Centre (Centrum dopravního výzkumu, v.v.i.) and Czech Technical University in Prague, Faculty of Transportation Sciences (BESIDIDO [15])
- data on roundabout safety from a project of Transport Research Centre (VEOBEZ [20, 21,22])

Models have been developed according to the mentioned steps. However considering the heterogeneity of data sets, only one explanatory variable was used (AADT). Therefore, the models have a general form $E(\lambda) = \alpha \cdot AADT^{\beta}$.

Available roundabout data sets had the following sizes:

- 88 roundabout conversions (GAČR)
- 58 roundabout conversions (BESIDIDO)
- VEOBEZ data set is being enlarged [11], currently 200 roundabouts are described in detail

In order to have comparable models, only 4-arm roundabouts with 1 lane, as a typical Czech roundabout design type, were selected. It limited the data sets to 76 (GAČR), 35 (BESIDIDO) and 120 (VEOBEZ) cases.

In the course of the analysis it became apparent that BESIDIDO data set is not sufficient for the estimation of statistically significant regression parameters; thus only two data sets were used for subsequent analyses. These are the data set GAČR, labelled as “CTU” (Czech Technical University) and the data set VEOBEZ, labelled as “CDV” (Centrum dopravního výzkumu, v.v.i.). The annual frequency of accidents of all severities was chosen as a dependent variable.

The first result is in Fig. 5: CTU is in blue, CDV in red; points are accident frequency data points, curves show expected accident frequency, i.e. prediction models.

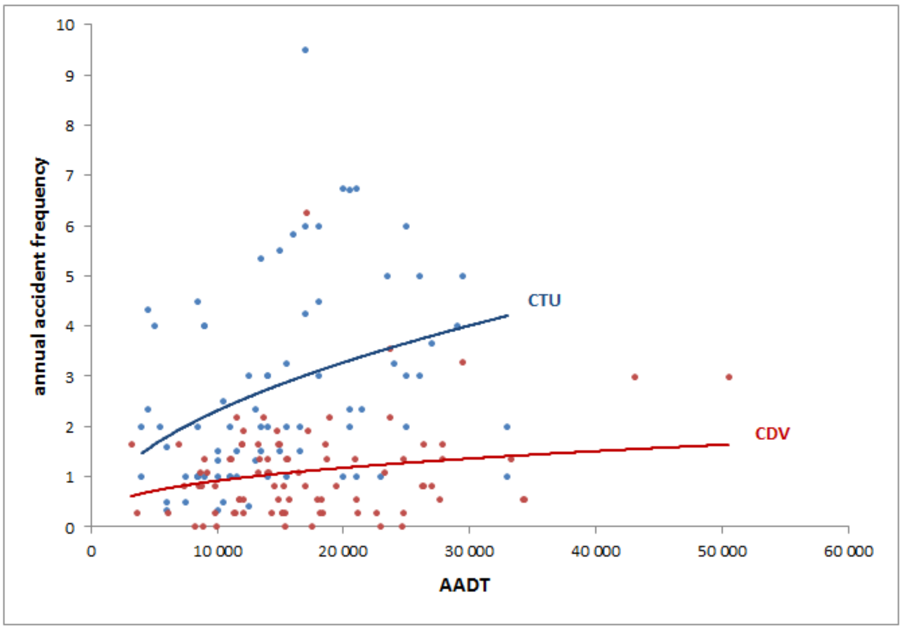

Fig. 5. Comparison of annual accident frequency on roundabouts in CTU and CDV data sets

The models are significantly different. The reason may concern time periods related to the collected data: CTU data are from 1995 – 2003, CDV data are from 2009 – 2012. Within this range several changes of underreporting took place.

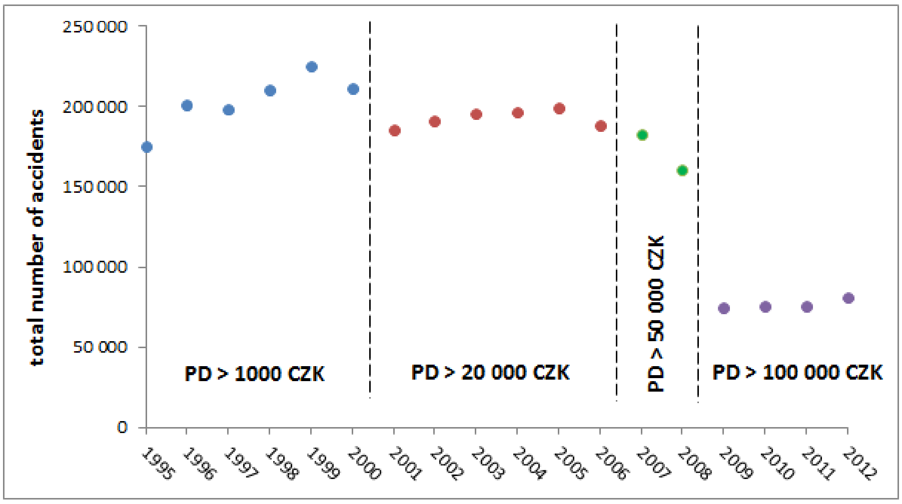

Fig. 6. Number of all reported accidents in four periods with different accident reporting thresholds based on limit of property damage (PD) in Czech crowns (CZK) [12]

Fig. 6 shows the total number of reported accidents in the Czech Republic divided into four time periods with different accident reporting thresholds (different limits of property damage). Since CTU data are from the first period and CDV data from the fourth, a relation between these periods was determined. It was calculated as a ratio of average values of numbers in these periods – the value was 2.6. Subsequently, CDV numbers were adjusted by multiplying by 2.6 in order to be comparable with CTU data. Fig. 7 shows the prediction models after this adjustment (only the central area with AADT below 30 000 vehicles per day is displayed).

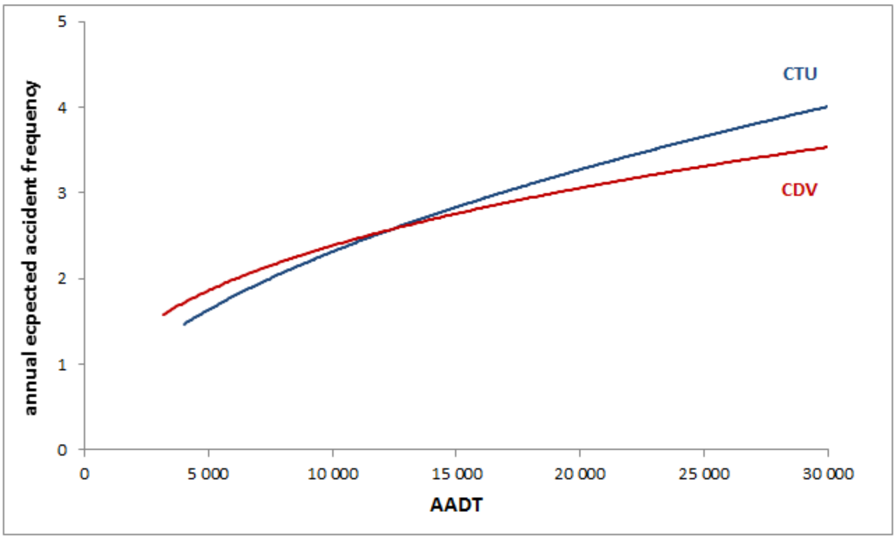

Fig. 7. Adjusted roundabout accident prediction models for CTU and CDV data sets

After the adjustment, both curves are relatively similar. One may speculate about the small differences. One of the reasons may be the fact that CTU data are from the roundabout conversions, i.e. the intersections with generally lower safety, which may be caused not only by the geometry. In contrast, CDV data were selected on an area-wide basis and are thus not influenced by this potential bias.

Nevertheless, there may be a host of influences towards the differences of both data sets safety performance. In order to illustrate them more widely a comparison with several international models was conducted and described in the following chapter.

4. COMPARISON WITH INTERNATIONAL MODELS

The above shown Czech prediction models were compared with several other models which are used abroad and were retrieved from the literature [6, 9, 13, 14, 23, 24]. They include some European examples (Belgium, France, Italy, Sweden, United Kingdom) as well as United States, Canada and New Zealand – see Fig. 8. The range of AADT values are limited between 10 000 and 30 000 vehicles per day.

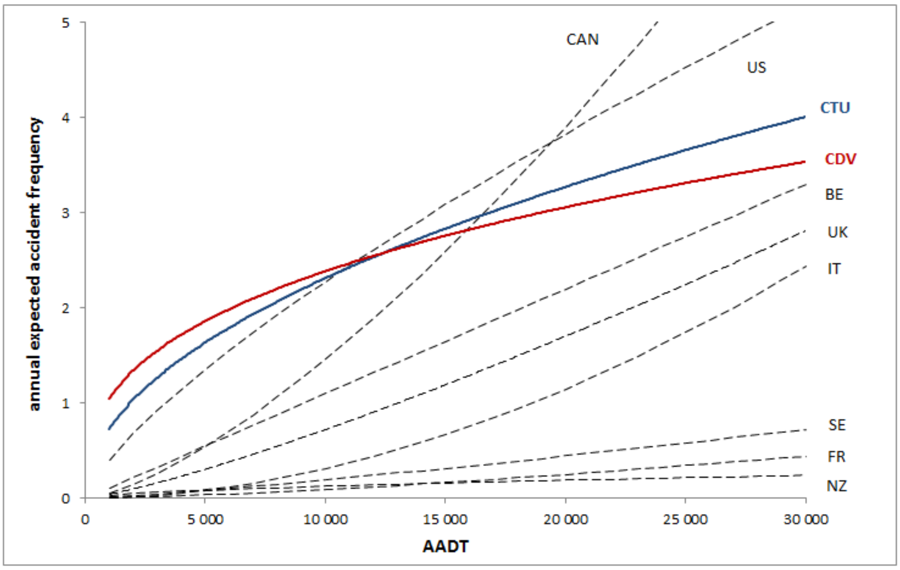

Fig. 8. Comparison of Czech roundabout accident prediction models with several international models (BE – Belgium, CAN – Canada, FR – France, IT – Italy, NZ – New Zealand, UK – United Kingdom, US – United States, SE – Sweden)

Considering the shape of curves, several conclusions may be made:

- Traditionally safe countries (Sweden, New Zealand) have the lowest expected accident frequencies.
- North American countries (United States and Canada) have similar shapes on the other side of the range.
- Most European countries (Belgium, United Kingdom, Italy) have values between those two thresholds.

However the reasons to the differences may be numerous; some of them will be listed (for more see [24]):

4.1. Various accident reporting practices

Most countries report just injury accidents and data in graph reflect this fact. They should have therefore lower values compared to the Czech models, which utilized also property damage only accidents. However there are differences with accident reporting among specific countries as well: for example in Sweden and New Zealand approximately 40% of injury accidents are reported, while in United States it is 70% and even 100% in Italy.

4.2. Definition of intersection crashes

There is no uniform criterion used for assigning an accident to an intersection. For example Belgian practice is to consider all accidents within an area of 100 m (the same holds for CDV data). However, in Canada 20 m limit is used, 30 m in Sweden and 50 m in New Zealand.

4.3. Design and traffic differences

For example roundabouts in France have a long tradition; what is more, they were built there primarily for safety reasons. On the contrary, the United States and the United Kingdom use roundabouts mainly because of capacity. These underlying concepts dictate the roundabout design, e.g. the diameter. There are also international differences in the age of roundabouts and the data sets do not cover the same time periods or rural/urban areas. Also speed characteristics and climate conditions may be significantly different.

5. SUMMARY AND DISCUSSION

The objective of the study was to compare roundabout accident prediction models. Data from Czech research projects have been used, as well as some international models. A number of challenges which result from the text will be summarized.

5.1. Availability of accident data

The availability of Czech accident data has been known for a long time; for a common user area-wide facilities are very limited. There is a public database JDVM, however, it serves mainly for basic statistical purposes and the details are not sufficient for safety analyses [1].

Increasing degree of underreporting is another challenge. The paper showed a potential way to its adjustment, however it is apparently only an approximation. It is possible to consider injury accidents only, however, it cuts the sample size needed for modelling. This fact is particularly severe with roundabouts: property damage only accidents comprised almost 80% of all accidents in presented data sets. The sample size may be extended by larger number of cases or prolongation of a time period: however, it is significantly time and resource demanding, apart from mentioning the instability of traffic characteristics.

A radical alternative is to substitute accidents with surrogate data, which do not depend on the mentioned databases. These may be the safety performance indicators, such as traffic conflicts. These data may even give a more complex picture of traffic and safety at the observed intersections, free from underreporting. Czech guidelines for the standardized traffic conflict observation and assessment is currently under development [4]; nevertheless, these data are very specific and are thus usually seen as a complement to accident data, not its substitute.

5.2. Availability of traffic engineering data

Also, Czech traffic engineering data (i.e. traffic data and road data) have been known to be an issue [1]; nevertheless, their condition is improving, e.g. with publicly available detailed results of National Traffic Census

2010. It would be also suitable to publish data from automatic traffic counters or other traffic surveys, e.g. on local roads. There are still parts of the road network which are not subjected to the national census. This is why various research activities had to cut their sample sizes, since intersections without known traffic volume data are useless. For example in a Transport Research Centre’s study [11], more than 40% roundabouts were situated on such local roads and therefore could not be included into a prediction model. What is more, traffic and road data (and accident data as well) exist separately, which impedes data searching and syntheses.

As stated before [1], a short term solution is to update and extend the outputs from currently used systems. A long term solution could be a change in data delivery systems; however it is in a collision with legal conditions.

Thus the alternative is to collect own primary data, as has been done for example in BESIDIDO project. It is however obvious that it is a demanding task with small coverage; it was also the reason why BESIDIDO data did not suffice in the presented study.

5.3. Availability of research data

Research data are outputs of accumulated data of previous research projects. The acquisition of data sets used in this study was possible only due to the cooperation between Transport Research Centre and Czech Technical University in Prague. On the other hand, data from a similar project conducted at Technical University in Ostrava were not acquired (107 roundabouts, according to [10]).

The solution would be a set-up of public database of research data. It could improve sample sizes and thus the quality of studies. It would also enable independent assessments and replications with published data.

5.4. Different methodologies

As mentioned in the introduction, accident data are non-normally distributed and linear regression and correlation techniques are thus not applicable. Even accident rate, which erroneously treats accident frequency as being in a linear relationship with traffic volume, should not be used (see more in [2]). These facts have been known abroad for a long time and they were also applied in the Czech Republic [16], however, not always with success [7]; nevertheless, they are still not a part of Czech road safety/traffic engineering knowledge.

To alleviate these methodological constraints, generalized linear modelling techniques are suggested. These procedures are routinely applied abroad and so were in this study. The most significant explanatory variable is AADT, however there are many others, as was apparent from presented Czech and international data sets. Multivariate modelling is thus a necessary further step. For example, using detailed data on 200 roundabouts [11], the following relation for annual expected accident frequency was found:

$$E(\lambda) = 0{,}0002 \cdot AADT^{0{,}576} \cdot D^{1{,}245} \cdot \Delta^{-0{,}377}$$

where D is roundabout diameter and Δ is deflection angle. The exponent signs confirm logical facts (the larger the diameter, the more accidents and the opposite for a deflection angle) and also enables to quantify their effects. Such models may then be used in studies related to safe roundabout design parameters.

6. CONCLUSIONS

The paper presented methods of accident studies and introduced analyses based on accident prediction models. The example focused on a comparison of roundabout prediction models; in addition specific challenges related to data collection, analysis and interpretation have been mentioned. Nevertheless these challenges have solutions; several of them were described in the previous text and they are summarized in the Table 1.

Table 1. Summary of challenges, their descriptions and solutions

Challenge	Description	Solution
Data collection	Insufficient quality and/or quantity of accident, traffic and road data	Data controls, updates, combination of databases Extension of samples in space and/or time Alternatively own survey, safety performance indicators/surrogate measures collection
Data analysis	Non-linear relationships, discrete data, non-normal probability distribution	Generalized linear modelling
Data interpretation	Multivariate relationships	More explanatory variables in prediction models

Proposed solutions have been known abroad, in the Czech Republic this is still not the case. Reasons may include a lack of knowledge, but also elaborateness, demanding thorough data preparation. As early as 10 years ago, it was stated that this approach is maximalist and nonstandard [17]; nevertheless, it is inevitable, as confirmed by international experience.

Acknowledgements

The study was conducted with the support of the Czech Ministry of Educations, Youth and Sports' institutional support of long-term development of Transport Research Centre.

References

[1] Ambros, J. (2011). Dopravně inženýrská data v českém prostředí: analýza dostupnosti, rozsahu a použitelnosti. *Silniční obzor*, *72*, 204-209.
[2] Ambros, J. (2012). Jak měřit bezpečnost? - 1. část. *Silniční obzor*, *73*, 78-82.
[3] Ambros, J. (2012). Jak měřit bezpečnost? - 2. část. *Silniční obzor*, *73*, 103-105.
[4] Ambros, J., Kafoňková, J., Kocourek, J., Kočárková, D., Kozel, P., Rusek, M., Turek, M., & Turek, R. (2012). Vývoj metodiky sledování a vyhodnocování dopravních konfliktů. *Silniční obzor*, *73*, 309-312.
[5] Ambros, J. (2013). *Hodnocení bezpečnosti dopravy* [online]. Brno: Centrum dopravního výzkumu, v.v.i. Retrieved May 22, 2013 from http://issuu.com/michdor/docs/m8_text_cdv?mode=window
[6] Daniels, S., Brijs, T., Nuyts, E., & Wets, G. (2010). Explaining variation in safety performance of roundabouts. *Accident Analysis and Prevention*, *42*, 393-402.
[7] Hála, M., & Slabý, P. (2006). Statistická analýza nehodovosti na pražských světelně řízených křižovatkách - 2. část. *Silniční obzor*, *67*, 74-77.
[8] *Identifikace kritických míst na pozemních komunikacích v extravilánu: metodika provádění*. (2012). Brno: Centrum dopravního výzkumu, v.v.i.
[9] Kennedy, J. (2007). *International comparison of roundabout design guidelines*. Published project report PPR206 [online]. Crowthorne: Transport Research Laboratory. Retrieved May 22, 2013 from http://www.trl.co.uk/online_store/reports_publications/trl_reports/cat_traffic_engineering/report_international_comparison_of_roundabout_design_guidelines.htm
[10] Mahdalová, I., Malina, T., Křivda, V., Cihlářová, D., Seidler, T., & Škvain, V. (2011). *Projekt výzkumu a vývoje č. CG911-008-910 Vliv geometrie stavebních prvků na bezpečnost a plynulost provozu na okružních křižovatkách a možnost predikce vzniku dopravních nehod: Závěrečná zpráva za rok 2010* [online]. Ostrava: VŠB – Technical University of Ostrava. Retrieved February 10, 2011 from www.mdcr-vyzkum-infobanka.cz/DownloadFile/32762.aspx
[11] Novák, J., & Ambros, J. (2012). Rozšíření predikčního modelu nehodovosti na okružních křižovatkách. In *20. Silniční konference* (pp. 127-132). Kongresové centrum Parkhotel Plzeň, 23. – 24. 10. 2012.
[12] *Přehled o nehodovosti na pozemních komunikacích v České republice* 1995 – 2012. Prague: Ředitelství služby dopravní policie Policejního prezídia ČR.
[13] Rodegerdts, L., Blogg, M., Wemple, E., Myers, E., Kyte, M., Dixon, M., List, G., Flannery, A., Troutbeck, R., Brilon, W., Wu, N., Persaud, B., & Lyon, C. (2007). *Roundabouts in the United States*. NCHRP report 572 [online]. Washington, D.C.: Transportation Research Board. Retrieved May 22, 2013 from http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_572.pdf
[14] Sacchi, E., Bassani, M., & Persaud, B. (2011). Comparison of safety performance models for urban roundabouts in Italy and other countries. In *90th TRB Annual Meeting*, paper 11-2538.
[15] Simonová, E., & Hrubý, Z. (2004). *BESIDIDO – Moderní úpravy komunikací ve městech a obcích pro zklidňování, vyšší bezpečnost a estetickou úroveň*. Silniční obzor, *65*, 132-133.

[16] Slabý, P., & Hála, M. (1999). Statistická analýza dopravních nehod - 3. část. *Silniční obzor*, *60*, 186-190.
[17] Slabý, P. (2003). Výzkum dopravní bezpečnosti na ČVUT Praha. *Silniční obzor*, *64*, 113-116.
[18] Slabý, P. (2005). *Problematika objektivnosti hodnocení bezpečnosti silničního provozu*. Final report of project GAČR 103/02/0065. Prague: Czech Technical University in Prague, Faculty of Civil Engineering.
[19] Slabý, P. (2007). Bezpečnost na úrovňových křižovatkách. In *Automobil v meste* (pp. 49-53). Bratislava, 27. – 28. 6. 2007.
[20] Šenk, P., & Ambros, J. (2011). Odhad počtu nehod na nově budovaných okružních křižovatkách. In *19. Silniční konference* (pp. 153-156). Kongresové centrum Zlín, 11. - 12. 10. 2011.
[21] Šenk, P., & Ambros, J. (2011). Estimation of Accident Frequency at Newly-built Roundabouts in the Czech Republic. *Transactions on Transport Sciences*, *4*, 199-206.
[22] Šenk, P., Ambros, J., & Daňková, A. (2011). Predikce ekonomických škod z dopravních nehod na okružních křižovatkách. *Silniční obzor*, *72*, 351-355.
[23] Turner, S.A., Roozenburg, A.P., & Smith, A.W. (2009). *Roundabout crash prediction models, NZ Transport Agency Research Report 386* [online]. Retrieved May 22, 2013 from http://www.nzta.govt.nz/resources/research/reports/386/docs/386.pdf
[24] Turner, S., Persaud, B., Lyon, C., Bassani, M., & Sacchi, E. (2011). International crash experience comparisons using prediction models. *Road & Transport Research*, *20*, 16-27.
[25] Valentová, V. (2013). Differences between the expected numbers of accidents for various types of intersection arrangements in Prague. In *Young Researchers Seminar 2013* [online]. Lyon. Soon to be available at http://www.ectri.org/YRS13/indexyrs13.html