

COMPARATIVE ANALYSIS OF ROAD ACCIDENT DATABASES FOR CREATING A HIGH-QUALITY ACCIDENT PREDICTION MODEL

USPOREDNA ANALIZA BAZA PODATAKA CESTOVNIH PROMETNIH NESREĆA U SVRHU STVARANJA KVALITETNIJEG MODELA PREDVIĐANJA

Krunoslav Tepeš¹, Davor Brčić², Dino Šojat³

Abstract: The lack of traffic safety is one of the leading causes of death and disability in the world. Besides internal costs for traffic participants, traffic accidents produce significant external costs to society at national, regional and local level. Identification and quantification of traffic accident causes on a global scale is increasingly emphasized, and different mathematical models as tools for traffic accident prediction are considered to use. The analysis of traffic accidents is based on a set of comprehensive information obtained from road accident summaries in the past. When developing a mathematical model which analyses and predicts road accidents, it is necessary to process as many as possible data collections related to causes and consequences of accidents. This approach ultimately leads to a realistic database which has a uniform structure. From that kind of database, it is possible to read the current state properly, and to set up trends and correlations of traffic safety indicators for the entire road transport. This paper presents two sets of road accident data: in the European environment and in the Republic of Croatia. The analysis of these two data sets draws conclusions in order to create a more precise traffic accident prediction model on the national level.

Keywords: road safety, road accident, database, risk factor

Sažetak: Nedostatak sigurnosti cestovnog prometa predstavlja vodeći uzrok smrtnosti i invaliditeta u svijetu koji osim internih troškova za sudionike u nesrećama producira i značajne eksterne troškove za društvo na nacionalnom, regionalnom i lokalnom nivou. Identifikacija i kvantifikacija uzročnika nastanka cestovnih prometnih nesreća u globalnim razmjerima sve se više naglašava i iskazuje potrebu korištenja raznih matematičkih modela, kao alata za prognoziranje cestovnih prometnih nesreća. Analiza cestovnih prometnih nesreća temelji se na skupu sveobuhvatnih informacija dobivenih iz retrospektivnih zbirnih podataka o nesrećama. Pri izradi matematičkog modela za analizu i prognozu cestovnih prometnih nesreća nužno je obraditi što više skupova podataka o uzrocima i posljedicama nastanka prometnih nesreća. Tim pristupom u konačnici se dobiva realna statistička podloga uniformirane strukture iz koje je zatim moguće adekvatno iščitati aktualno stanje, trendove i korelacije kretanja sigurnosnih pokazatelja cestovnog prometnog sustava. U radu je prikazan skup podataka o cestovnim prometnim nesrećama europskog okruženja, te podataka koji se prikupljaju na nacionalnom nivou Republike Hrvatske. Analizirajući ta dva skupa podataka o cestovnim prometnim nesrećama, crpe se zaključci u cilju stvaranja kvalitetnijeg modela predviđanja prometnih nesreća na nacionalnom nivou.

Ključne riječi: sigurnost prometa, cestovna prometna nesreća, baza podataka, rizični činitelj

¹ Professional expert, Tepeš Krunoslav, M.Sc. Traffic Eng. (Ph.D. Cand.), City of Zagreb, City Office for Physical Planning, Construction of the City, Utility Services and Transport, Department of Transport, Trg Stjepana Radića 1, Zagreb, Croatia, krunoslav.tepes@zagreb.hr

² Head of the Department of Urban Transport, prof. Brčić Davor, Ph.D., University of Zagreb, Faculty of Transport and Traffic Sciences, Department of Urban Transport, Vukelićeva 4, Zagreb, Croatia, davor.brcic@fpz.hr

³ Professional associate, Šojat Dino, M.Sc. Traffic Eng., University of Zagreb, Faculty of Transport and Traffic Sciences, Department of Urban Transport, Vukelićeva 4, Zagreb, Croatia, dsojat@fpz.hr

1. INTRODUCTION

Road accidents and their consequences (deaths, severe injuries and disabilities) present a major social, economical and political issue on the global, regional and national level. Accidents are also considered as the most serious problem of the whole transportation system.

According to present trends, accidents aspire to become the third cause of death by 2030 (The global burden of disease: 2004 update).

In an effort to improve the condition of road safety, i.e. to make the whole transportation system safer, it is necessary to develop the idea of road safety management. The development of different mathematical accident models (able to predict the number and category) represents an unavoidable analytic tool whose purpose is to identify and quantify the key factors contributing to traffic accidents in order to manage road safety more effectively (Elvik, 2004).

The quality of the model depends on the incoming data – these data gathered from the documentation and statistics has to be collected systematically, and processed with the support of as many high-quality input parameters (information).

This paper presents the current development state of the official information system in charge of road accident data processing in the Republic of Croatia in comparison to current activities and methods in the European Union. The purpose of this comparison is to improve the methods of road accident data collection (to increase the quantity of information) within the existing road accident databases on the national level. The analysis of road accident database in the Republic of Croatia and foreign expertise altogether draw conclusions regarding basic problems and limitations existing in the current national database. Also, the analysis points out the need to synchronise the accident data collection process with the European Union in order to manage traffic safety more efficiently and to achieve the goals proposed by the National Program on Road Safety in Croatia for 2011-2020.

2. DATA COLLECTION – SOURCES AND METHODOLOGY

From a statistical point of view, road accidents and their consequences are considered random events and they are relatively rare if total distance travelled by vehicles is considered. Even though it is almost impossible to predict spatial and temporal coordinates of a certain accident, a set of these accidents can have almost perfectly predictable patterns if precise mathematical and statistical relationships are defined (Fridstrom and Ingebrigsten, 1991).

For this reason, it is essential to establish the national and local transport policies ensuring continuous monitoring of road safety indicators if there is an initiative to determine whether the current trends are in correlation with the objectives pursued, measures implemented or expectations planned. The monitoring is based on studying direct traffic safety indicators on annual basis, which are the number of road accidents and their consequences – the number of deaths and injuries.

On the global scale, both identification and quantification of road accident causes is increasingly emphasized and therefore recognized as a key component in the process of accident forecasting and prevention. This is done by using different mathematical models and algorithms, especially nowadays when there is a high presence of modern IT technology within all branches of traffic science.

The usage of the available historical sets of road accident data records is one of the most relevant ways to identify and classify risk factors contributing to road accidents with all characteristics and consequences. The reason for using these data sets is the fact that cumulative effects of specific measures for increasing traffic safety implemented can be adequately measured and evaluated. This provides traffic policy makers a very important feedback when considering whether the set of measures and strategies resulted in desired goals.

2.1. Statistical traffic accident database

In the past, traffic accident data was not adequately collected, unified and filtered to become appropriate for statistical analysis at the national level. The data often turned out to be incompatible, which was the result of independently adopted analytical principles, structures and characteristics within each source of traffic accident data registries.

The difference between each individual road accident database is evident from the amount of basic road accident data that needs to be input in the accident questionnaire. The difference is evident by looking at the number of variables, values or pages required when entering data (Tab. 1).

In the EU state members, definitions and methodologies applied to study the road accident data are not entirely unified, which makes the implementation of a high-quality comparative analysis difficult. In addition to the national level, the unified definitions and methodologies have to be put through the accident databases at the local and regional level in order to facilitate the comparison of processed data and to ensure credibility.

At this moment, unified international traffic accident databases have been emerging in the world. For this purpose, standards were adopted with an example of good practice – the establishment of the Glossary for Transport Statistics, accepted by UNECE, EUROSTAT and ECMT. The establishment of such kind of database ensures the following:

- Traffic accident data comparison at the national level,
- Ranking of countries,
- Indication of urgency to get international support,
- Information about the development and progress of each country,
- Better identification of vulnerabilities in the transport system,
- Detection of differences in the level of safety at different points of road infrastructure (Road accident investigation guidelines for road engineers, 2007).

Besides the process of unifying national road accident databases, numerous studies and researches in the field of road safety (Kim et al, 1995; James, 1991) also have to rely on the integration of data collected from different sources which get the credit for these activities. Such sources in most cases are institutions involved in transportation, police institutions and health institutions.

Table 1. The fundamental amount of information collected through road accident questionnaires (The DUMAS project, 2000)

Number	Country	Number of variables	Number of values	Number of pages
01	Germany	30	149	3
02	France	68	348	1
03	Italy	15	197	2
04	The Netherlands	38	148	2
05	Belgium	61	239	4
06	Luxembourg	26	153	2
07	United Kingdom	50	255	4
08	Ireland	29	171	1
09	Denmark	45	202	1
10	Greece	20	168	1
11	Spain	78	277	1
12	Portugal	22	82	2
13	Croatia	38	119	2

The establishment of a high-detail, high-quality database and the development of different statistical models assumes interconnections of different databases from different sources, so that the relevant information about causes and consequences of accidents can be collected from different points of view and processed, analysed and interpreted accordingly.

Although the technique mentioned above is not an innovation in the analysis of traffic accidents, it should be noted that the most significant component for the justification of the technique is the integration of data from traffic police and health institutions to establish relationships between circumstances preceded the accident and the cause of deaths. The process proposed provides extensive data validation (determination of inconsistencies and discrepancies) among individual databases and relevant information upgrades.

The connection of databases also requires the cooperation between a large number of entities, each responsible for maintaining statistical records related to traffic accidents and corresponding casualties in accordance to methods and procedures agreed upon in order to process the data efficiently (Ferrante et al, 1993).

Although significant advances in the process of accident data collection and analysis have been made, the process is still underdeveloped and rigid – the information explaining causes of road accidents is still poor, and in some reports even omitted. All the databases mentioned have the same problem, which comes down primarily to consistency and precision of what is normally considered to be “routine” traffic accident data either from the police institutions, health institutions or some other sources of information.

3. AN OVERVIEW OF INTERNATIONAL ROAD ACCIDENT DATABASES

Standardisation of information gathering protocols and definitions regarding traffic accidents significantly varies between EU countries, preventing further accident data comparisons. Furthermore, the tools used for processing the same data as well as the very structure of subjects responsible for data collection and processing (in most cases, police institutions) differ between EU countries, which further makes the comparison of processed data more difficult.

The result is a fact that a large number of countries had reviewed their statistical traffic accident models and developed a scheme on the basis of their databases in cooperation with other relevant sources of data nationwide. This created preconditions for establishing international traffic accident reports which are both synchronised and standardised according to the European Union monitoring system.

One of the measures for synchronisation of national and EU accident data systems is the implementation of an international accident data collection system. An example of good practice is a European traffic accident database called CARE (Community data bank on road traffic accidents in Europe), which is, after several upgrades, known in its present form as CADaS (Common Accident Data Set). The database is managed by Directorate-General for Mobility and Transport (DG TREN).

The establishment of the database defined a unified analytical methodology with a minimum number of standardised variables to be used in order to achieve a comparable level of traffic accident information between EU member states. In addition, it is proposed within the analysis to collect a total of 88 variables and 481 values (Tab. 2). The variables proposed are divided according to their relevancy in the analysis, and values are divided according to level of detail that can be achieved in the collection process. The next system structure represents the decomposition of variables onto 4 basic info groups – traffic accident, vehicles involved, persons and roadway where the accident occurred.

The additional reason for data integration from sources mentioned above is the limitation of data (risk factors) at individual scale and similarities in analytical methodology.

An important feature of databases (both national and international) is the form of displaying – the large amount of traffic accident data is recorded and displayed in the form of categories with several classifications and levels of importance.

Table 2. The structure of CADaS – variables and values (Recommendation for a Common Accident Data set, 2011)

Variable	Number of variables			Number of values		
	High importance (H)	Low importance (L)	Total	Detail values	Alternate values	Total
Traffic accident (A)	7	5	12	76	15	91
Roadway (R)	16	21	37	151	15	166
Vehicle (V)	7	10	17	110	7	117
Participant (P)	15	7	22	96	11	107
Total	45	43	88	433	48	481

In this particular area, traffic experts have the responsibility to apply their skills of processing and interpretation to effectively identify specific traffic accident causes for each individual category by using a variety of tools, techniques and applications able to compress variable categories (Dupont and Martensen, 2007).

4. AN OVERVIEW OF SOURCES OF ROAD ACCIDENT DATA IN THE REPUBLIC OF CROATIA

In accordance to possibilities of technology and legal obligations, the relevant national database has been systematically upgraded and modernised, but until this point the data has not been merged within a unique database.

The principal purpose of the database is to detect hazardous locations on the road network, sometimes referred as the accident blackspots, as well as the most significant risk factors of road accidents on the basis of statistical traffic accident data to make the database act preventively.

The sources of data regarding traffic accidents on the national level can be obtained from public and civil institutions responsible for data collection and processing, of which the most relevant data is the one from the Ministry of the Interior and from the Croatian Bureau of Statistics. The data can also be obtained from health institutions running internal records, insurance companies, Croatian Roads, Croatian Freeways, county road administrations, Vehicle Center of Croatia (CVH) and the National Protection and Rescue Directorate (DUZS).

The legal obligation of institutions is to provide the collected and processed traffic accident data on the national level and the level of each county, according to the Law on Official Statistics (Zakon o službenoj statistici, 103/03, 75/09), in jurisdiction of the Croatian Bureau of Statistics and the Ministry of the Interior. The contents of such information are details of each road accident ever reported: number of traffic accidents with injuries according to type, consequences, cause, conditions and participants (vehicles, drivers, pedestrians and passengers) as well as the number of deaths and severe injuries.

Besides the road accident classification mentioned above, the whole range of other relevant information and possible risk factors specific to road accidents can be shown: roadway category, road material and surface conditions, weather conditions (visibility), driver specifications (age, gender, exposure to alcohol or narcotics), vehicle condition, etc.

The data used for establishing previously mentioned bulletins on road safety in the Republic of Croatia draws the information from these two data sources:

- Records and statistical monitoring of registered vehicles,
- Records and statistical monitoring of road accidents.

The Information Subsystem of the Ministry of the Interior is an important system for tracking safety indicators in traffic which is in charge of statistical and analytical monitoring of traffic accidents obtained in the investigation process related to accidents. The same system is used for later data entries, whereas the data is gathered from the road accident questionnaires compliant to European standards (Pravilnik o načinu postupanja policijskih službenika u obavljanju poslova nadzora i upravljanja prometom na cestama, 141/11).

For the purpose of making specific traffic safety reports and safety assessments, besides the accident data itself, the Information Subsystem of the Ministry of the Interior provides access to traffic accident data for individual micro-locations, i.e. the sections of road network on which traffic accidents occur more frequently (data about municipalities, towns, roads, house numbers, etc.).

In addition, it has been two years since the traffic police introduced detection of high-risk locations (blackspots) as well as geographical positioning of every road accident with Global Positioning System and GPS devices, which altogether reduce the costs and simplify the whole accident positioning process

The number of input variables included into statistic and analytic process of the Information Subsystem as well as the system itself, has been systematically improving and modernising in response to emergence of sophisticated technology over the years. During 2009, there have been several updates of applications dealing with traffic accidents, traffic violations, driving schools and vehicle inspection stations. Also, instead of the previous PN-10 questionnaire, the new UPN questionnaire for statistical monitoring of accidents has been introduced. From a statistical viewpoint, the new questionnaire raised the quality of data to a higher level due to updates and changes made in it.

The traffic accident questionnaire mentioned above (UPN) currently consists of 38 questions. Police officers have to input the accident data, which is the basis for precise traffic accident database (Zovak et al, 2011). The revised data collection method introduced in UPN has been valid since January 2010.

Since the Information Subsystem of the Ministry of the Interior which is in charge of collection and analysis of the traffic accident data has already been developed appropriately (sets are consistent, extensive and categorised), it is currently possible to develop accident prediction models for the future.

The national traffic accident database should contain all relevant elements which can improve the analysis of road accidents. These elements are correlated with the same systems of other EU countries for the purpose of creating the conditions in which the unique system in charge of monitoring events and incidents in traffic can be implemented. This has been a request from relevant organisations in the European Union for a long time.

5. SUGGESTIONS FOR FUTURE ACTIVITIES TOWARDS THE UPGRADES OF NATIONAL ROAD ACCIDENT DATABASE

Besides the improvement of the procedures during the investigation process (traffic accident analysis, collision diagrams, GPS devices, etc) and the implementation of traffic police training, the improvement of the existing road accident database at local and national level requires the incorporation the data gathered through:

- Vehicle Center of Croatia,
- The department in charge of transport inspection (Ministry of Maritime Affairs, Transport and Infrastructure),
- The departments in charge of the infrastructure management in the City of Zagreb⁴,
- Database including traffic volume and road conditions.

This last item is regulated by the Law on Roads (*Zakon o cestama*, 84/11, 22/13, 54/13, 148/13) and with the Decision of Unclassified Roads (*Odluka o nerazvrstanim cestama*, 19/99, 19/01, 20/01), it lies under the jurisdiction of the city government.

This approach for creating a comprehensive database has not been established yet in the Republic of Croatia, but the development of such database would enhance the existing statistics and documentation. Also, this would result in systematic identification and elimination of the most critical locations on the entire road network and at each level (local, national, society).

With this kind of analysis, it is possible to point out the disadvantages of roads and areas surrounding them as well as the disadvantages of traffic and the user behaviour.

Since the comprehensive database mentioned above would have interdisciplinary features, it should be handled by the Ministry of Maritime Affairs, Transport and Infrastructure.

6. NATIONAL ROAD ACCIDENT DATABASE – BENEFITS AND ADVANTAGES FOR THE LOCAL COMMUNITY

In the scope of creating the national road accident database and statistical methods for connecting, classifying and predicting the number of accidents, it should be mentioned that such research were not carried out in the Republic of Croatia in any form.

With the complexity and the large number of parameters (which are independent variables) responsible for road accidents given, there has not been established a single scientific approach which would create an accident prediction model.

Within the Ph. D. studies, a research on the accident prediction model in the urban environment and the local community is in progress at the Faculty of Transport and Traffic Sciences.

The creation of a unified traffic accident database on the national level will create the preconditions to make the data on the local level available, by both content and legislation.

The formation of local entities in charge of traffic safety on the local level (within the local government) with political support, well-defined objectives, programs and provided funds would create preconditions for better and more substantial activities for improving safety in road traffic.

Consequently, with such activities related to road safety improvements, the local communities would also create conditions for achieving the objectives set by the National Program on Road Safety in Croatia for 2011 – 2020.

7. CONCLUSION

The analysis of road accidents represents primarily the studies of official statistical reports from several institutions (dominantly from traffic police) in order to identify the most important causes of accidents with high precision. By improving information systems and expanding existing traffic accident database, researchers are provided new possibilities in terms of data collection methodologies.

⁴ The departments in charge of the infrastructure management in the City of Zagreb are Croatian Freeways, Croatian Roads nad the City of Zagreb itself, which had established the Zagreb Roads as the branch of Zagreb Holding for such activities.

Various statistical models now used to predict traffic accidents are becoming very important for the perception of road safety because they represent the indicators of performance. If the accident prediction curve does not match the actual curve, it means that certain unknown factors influence on traffic and cause road accidents. In that case, it is necessary to identify those factors as soon as possible.

This fact has resulted in the possibility of establishing relationships between different variables in the modelling process. These variables can be social, demographic, economic, environmental and political with an aim to describe, explain or predict the level of safety in road traffic.

With complexity of the problem given in this paper, described with a great amount of critical parameters (independent variables) which are the elements of accident incidental factors, there is still no single methodological approach for the development of an accident prediction model. Accordingly, many authors in the field of transport science have been systematically worked on a variety of methods, techniques and statistical models to predict traffic accidents in the past two decades. During that period, older input data and variables have been upgraded with a variety of new data sets according to time periods of the studies themselves.

The problem that is often present in determination of input variables and elaboration of mathematical models is the lack of data for longer periods of time, which results in their negligence in further calculations, and ultimately, very poor reliability of calculation.

This paper presented the accident data collection method at the level of the Republic of Croatia and the level of the European Union.

When establishing a traffic accident prediction model, the following should be considered: limitations in data collection process, insufficient number of relevant data about the causes, number and consequences of traffic accidents, underreporting and the lack of transparency and unavailability of databases. This results in incompatibility between the sources of information, and the impossibility to create a unified traffic accident database which would serve as a platform for traffic accident prevention.

Despite the above, it is important to mention that the national accident registry under the Ministry of the Interior is currently one of the best sources of traffic accident information in the Republic of Croatia with the ability to help traffic experts in their future research in the field of road safety.

8. BIBLIOGRAPHY

- [1] Common Accident Data Set (CADA_S), Recommendation for a Common Accident Data set, Reference Guide, Version 3.11., 2011.
- [2] DENCO, Universia di Brescia, INRETS, TRL & Aristotle University of Thessaloniki (2000). The DUMAS Project. Developing Urban Management and Safety. WP-4. Accident Investigation.
- [3] Dupont, E. and Martensen, H. (Eds.) (2007). Multilevel modelling and time series analysis in traffic research – Methodology. Deliverable D7.4 of the EU FP6 project SafetyNet.
- [4] Elvik, R. (2004). Does use of formal tools for road safety management improve safety performance? Transportation Research Record: Journal of the Transportation Research Board 2318 (1), 1-6.
- [5] Ferrante A.M., Rosman D.L., Knuiman M.W. (1993). The construction of a road injury database, Accident Analysis and Prevention 25(6), 659-665.
- [6] Fridstrom, L., Ingebrigsten, S. (1991). An aggregate accident model based on pooled, regional time-series data, Accident Analysis and Prevention 23(5), pp. 363-378.
- [7] James, H.F. (1991). Under reporting of road traffic accidents, Traffic Engineering and Control 32, 573–583.
- [8] Kim, K., Nitz, L., Richardson, J., Li, L. (1995). Personal and behavioural predictors of automobile crash and injury severity, Accident Analysis and Prevention 27(4), 469–481.
- [9] Odluka o nerazvrstanim cestama (2007), Službeni glasnik Grada Zagreba, br. 19/99, 19/01, 20/01.
- [10] PIARC, Road accident investigation guidelines for road engineers, 2007.
- [11] Pravilnik o načinu postupanja policijskih službenika u obavljanju poslova nadzora i upravljanja prometom na cestama (2011), Narodne novine, br. 141/11.
- [12] The global burden of disease: 2004 update, World Health Organization, 2008.
- [13] Zakon o cestama (2014), Narodne novine, br. 84/11, 22/13, 54/13, 148/13.
- [14] Zakon o službenoj statistici (2013), Narodne novine, br. 103/03, 75/09, 59/12.
- [15] Zovak, G., Šarić, Ž., Čop, A. (2011). Conceptual model for structuring traffic accident data, Proceedings of the 22th International DAAAM Symposium – DAAAM 2011.