Road safety data: collection and analysis for target setting and monitoring performances and progress

Analyses using the European Road Safety Observatory

Emmanuelle Dupont  IBSR, Belgium, e-mail: Emmanuelle.DUPONT@bivv.be
Heike Martensen  IBSR, Belgium, e-mail: Heike.MARTENSEN@bivv.be
Henk Stipdonk  SWOV, the Netherlands, e-mail: Henk.Stipdonk@swov.nl
Frits Bijleveld  SWOV, the Netherlands, e-mail: Frits.Bijleveld@swov.nl
Jacques Commandeur  SWOV, the Netherlands, e-mail: jacques.commandeur@swov.nl
Eleonora Papadimitriou  NTUA, Greece, e-mail: nopapadi@central.ntua.gr
George Yannis  NTUA, Greece, e-mail: geyannis@central.ntua.gr
Constantinos Antoniou  NTUA, Greece, e-mail: antoniou@central.ntua.gr
Ruth Bergel  INRETS, France, e-mail: Ruth.Bergel@inrets.fr
Mohammed Debbarh  INRETS, France, e-mail: mohammed.debbarh@inrets.fr
Christian Brandstaetter  KfV, Austria, e-mail: christian.brandstaetter@kfv.at

Abstract

Road safety analyses aim to describe and explain road safety outcomes (road accidents and casualties), either in time or in space, as well as to forecast future developments on the basis of existing experience. The availability of reliable data is one of the most important conditions for useful road safety analyses. Within the development of the European Road Safety Observatory (ERSO), a wealth of road safety related data at European level was gathered, harmonized and made available, including road accidents data (through the CARE database), risk exposure data, road safety performance indicators and in-depth road accident investigation data. The objective of this paper is to present a comprehensive overview of analyses carried out by using data and knowledge from the European Road Safety Observatory. These analyses were carried out within the framework of the EC co-funded integrated research project SafetyNet and can be broadly distinguished into time series analyses, geographical analyses, and accident analyses. It is shown that different types of dependencies are present among observations (e.g. serial or hierarchical dependencies), calling for particular statistical treatment by means of dedicated techniques, such as multilevel analysis and time series analysis. Moreover, additional methodological questions are tackled, such as the use of exposure data; the interest of disaggregate analyses; and the potential of combining ERSO data with other data sources. A brief overview of some interesting findings from related case studies in different countries or at European level is finally provided, including combined analyses of road accidents, fatalities and exposure, structural and explanatory analyses of road safety trends, disaggregate time series analyses and forecasts, spatial modeling, joint analysis of CARE and road safety attitude data, statistical analysis of injury underreporting rates and modeling of in-depth fatal accident investigation data.

Key-words: road safety; time series analysis; geographical analysis; accident analysis; European Road Safety Observatory.
1. Introduction

Road safety analyses aim to describe and explain variation in road safety outcomes (road accidents and casualties), either in time or in space, as well as to forecast future developments on the basis of existing experience. Such analyses can be broadly classified into three groups, according to the type of question addressed and the type of data required (Martensen et al. 2008):

- Time series analyses, in which the questions addressed concern the description, explanation and forecasting of overall or particular (e.g. motorcyclists') road safety trends. In this case, the data required concern series of measurements over time.

- Geographical analyses, in which the questions addressed concern road safety differences between regions, on the basis of attributes of these regions as well as of external factors (e.g. road safety measures). The data to be used in order to answer such questions may show a geographical hierarchical structure, given that regions are nested within countries, and neighbouring regions may be more similar to each other than regions that are situated at opposite ends of the country.

- Accident analyses, in which accident mechanisms are analyzed, in terms of the contribution of road, vehicle, driver and accident characteristics to the accident outcomes. In order to answer these questions, one has to rely on relatively detailed accident data. These data typically have a hierarchical structure, because more than one road user may have been in the same vehicle and more than one vehicle may have been involved in the same accident.

These different types of dependencies among observations (e.g. serial or hierarchical dependencies), can be efficiently handled by means of dedicated techniques, such as multilevel analysis and time series analysis (Dupont & Martensen, 2007). Moreover, a number of questions related to data availability and quality may need to be addressed as well. For instance, the necessary data may not be available at the required level of detail, or may not be available through a single data source, or may not be comparable. These questions may be encountered in road accident and fatality data, but may also become even more pronounced when it comes to other necessary data for reliable road safety analyses.

Within this framework, the development of the European Road Safety Observatory (ERSO), has contributed a wealth of road safety related data at European level. These data were gathered, harmonized and made available, and include road accidents data (through the CARE database), risk exposure data, road safety performance indicators and in-depth road accident investigation data.

The objective of this paper is to present a comprehensive overview of the analyses carried out using data and knowledge from the ERSO, within the framework of the EC co-funded research integrated project SafetyNet. These analyses concern all three types of road safety questions at national or European level, demonstrating how the related dependencies were handled by means of appropriate advanced techniques, and exploiting improved data availability and quality. Moreover, their results reveal interesting road safety patterns and provide useful insight on several road safety questions.

2. Methods and data

2.1. Dependencies in road safety data

Most statistical techniques are based on building a simplified model to describe the data. Factors that are assumed to have explanatory power are introduced in this model and statistical tests are then used to evaluate whether this factor actually allows the model to describe the data better. These tests in general rely on the assumption of the independence of the observations involved. However, in several cases the independence assumption may not hold, e.g., in grouped data (some cases are more similar to each other than to others), or repeated measurements data.

Specifically in road safety analyses, dependencies among observations may be time-related, space-related or accident-related. In general, two advanced statistical techniques may be used to handle these dependencies, namely time series and multilevel analysis. A detailed presentation of the assumptions and properties of the various techniques is beyond the scope of this paper and the reader is referred to Dupont & Martensen (2007).

A summary of the type and the potential of the techniques is provided below and presented in Table 1.
Table 1: Techniques for handling dependencies in road safety data

<table>
<thead>
<tr>
<th>Potential problem for statistical tests</th>
<th>Time series analysis</th>
<th>Hierarchical (accident)</th>
<th>Hierarchical (geographical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>Well applicable</td>
<td>Unknown, probably small</td>
<td>Medium</td>
</tr>
<tr>
<td>Methods for solution</td>
<td>Time series analysis</td>
<td>Multi-level modelling</td>
<td>Multi-level modelling</td>
</tr>
<tr>
<td>Applicability</td>
<td></td>
<td>Difficult to apply</td>
<td>Possible to apply</td>
</tr>
</tbody>
</table>

Time series analysis may be performed by means of different techniques, ranging from extensions of generalized linear models and non-linear models (in which the autocorrelation in model residuals needs to be handled), to dedicated time series analysis methods such as Autoregressive Moving Average or ARMA-type models or state-space model (in which the serial correlation among observations is explicitly accounted for), while accounting for other fixed, explanatory or random effects. In Deliverable D.7.4 of the SafetyNet project, these techniques have been extensively tested with real data, confirming their overall applicability, as well as the fact that ignoring the serial correlation among observations may result in non-negligible bias.

On the other hand, multilevel analysis concerns techniques intended to capture correlations due to some hierarchical grouping of the data. Most standard statistical techniques (linear or generalized linear models, multivariate models etc.) may be extended to multilevel forms of the models, in which observations are grouped into levels and separate statistical tests are carried out for identifying explanatory effects within and among levels, while also explicitly accounting for the random variation within and among levels. In Deliverable D.7.4 of the SafetyNet project, these techniques have been also extensively tested with real data.

It was concluded that geographical hierarchies (e.g. accidents of counties nested into regions nested into countries and so on) can be relatively easily captured by multilevel modelling structures, correcting in several cases for bias in the estimation. On the other hand, accident hierarchies (e.g. fatalities nested into vehicles nested into accidents) appear to be more difficult to examine, mainly due to small sample size (i.e. small number of casualties in the same vehicle and small number of vehicles in the same accident), but at the same time seem to be less critical for the analysis.

2.2. The European Road Safety Observatory

A number of questions related to data for road safety analysis at European level have been identified, namely unavailability, incompleteness (including under-reporting issues), poor comparability, inappropriate level of disaggregation and limited accessibility (Yannis, 2000).

The development of the European Road Safety Observatory within the SafetyNet project has resulted in the collection, harmonization and publication of an important amount of data for 27 European countries. In particular (Thomas et al. 2006):

- Road accident data of the new Member States were collected, harmonized and introduced in the CARE database.
- The level of road accident injury under-reporting was estimated through pilot studies in seven European countries linking Police data with hospital data, and a common definition of injury severity was proposed.
- A thorough investigation of national risk exposure data (vehicle-kilometres of travel, vehicle fleet, driving licenses, road length etc.) in terms of availability and comparability was carried out and a first set of transformation rules was proposed. Some of this data is currently available at Eurostat.
- A set of key road safety performance indicators was defined with respect to alcohol and drugs, Alcohol and drug-use, speeds, protection systems, daytime running lights, vehicle passive safety, roads and trauma management and extensive data collection was carried out for the calculation of the indicators where possible.
- A database of approximately 1300 fatal accidents from seven European countries was assembled to describe the key characteristics of these accidents, with some interpretation of causation. A common methodology based on in-depth road accident investigation techniques was used on that purpose.
- A second database provides an in-depth description of the causation of around 1000 fatal and non-fatal accidents and identification of key risk factors, focusing on infrastructure safety and the needs of eSafety technologies.
To conclude, the ERSO includes a wealth of comparable information on both macroscopic and in-depth road safety data. In the following sections, it is shown how this data was exploited for addressing key road safety questions. In a few cases, ERSO data were combined with other data sources, which could be eventually included in the ERSO.

3. Analyses

3.1. Time series analyses

3.1.1. Structural changes in road safety trends

An analysis of the development of fatality rates over time was carried out for different European countries in relation to motorization rates, aiming to identify differences in the slopes and breakpoints of the macroscopic fatality trends between countries. Data for the period 1960-2005 were used on that purpose. The results are summarized in Figure 1.

It can be concluded that different countries reach specific motorization rates in different years (temporal landmarks) (Stipdonk et al. 2008). However, some of the examined countries exhibit their major breakpoint in fatality risk within a narrow range of motorization rate values (320-370 vehicles per 1000), suggesting that these breaks take place under similar social and economical conditions. The range is different for certain subgroups of the examined countries, which show a more complex pattern. These findings can serve as a starting point to obtain a further understanding of why, and when, these important breakpoints of fatality trends are observed.

![Selected countries – Personal risk vs. motorization](image1)

![Selected countries – Personal risk vs. time](image2)

Figure 1: Fatalities per million inhabitants vs. motorization (top panel) and time (bottom panel) in European countries
3.1.2. Simultaneous modeling of the three levels of road risk

A simultaneous analysis of the time series of exposure (e.g. vehicle-kilometres traveled); accident risk (accidents per vehicle-kilometres); and accident severity (fatalities per accident), often referred to as the three levels of road risk, can be very advantageous. Important inter-relations between road safety outcomes and exposure exist and can be optimally accounted for by using such a multivariate time series analysis. Examples for France and the Netherlands can be found in Commandeur et al. (2007).

3.1.3. The interest of disaggregation

When only global figures (e.g. total accidents, total mobility) are used to explain traffic safety changes over time, the resulting models often behave poorly. This is because, in any country, the total safety figures depend on the related figures of different vehicle groups (e.g. two-wheelers, passenger cars, heavy vehicles etc.), road users (males / females, young or elderly drivers etc.) and different road networks (e.g. urban / interurban). Consequently, further understanding of traffic safety development asks for disaggregation of mobility and accident data into different models for different components of the road traffic system. Several such analyses were carried out within SafetyNet.

An interesting example concerns three-level models of risk in France that were developed separately for main rural roads and motorways (Commandeur et al. 2007). The results show these two types of roads differ from each other in the development of vehicle-kilometres travelled, injury accident risk and accident severity (see Figure 2). Most importantly, the developments of accident risk and accident severity are dissociated for these two road types. While motorways have a lower accident risk than main rural roads, the severity of those accidents that do happen on motorway is higher as compared to main rural roads. As far as evolution over time is concerned, accident risk is generally decreasing, but not so much for motorways, whereas accident severity is generally decreasing, but not so much on rural roads.

![Figure 2](image)

Figure 2 : Estimated quarterly vehicle-kilometres, injury accidents and fatalities on motorways (left panel) and main rural roads (right panel) in France

Another example concerns the disaggregation into single- and two-vehicle accidents of passenger cars in the Netherlands (Commandeur et al. 2007). Car-car accidents were further stratified by type, namely frontal, rear and side impact car-car accidents. It was found that the risk of being involved in a single car KSI (killed or seriously injured) accident was approximately equal to the risk of being involved in a car-car KSI accident. The related risks for the three different car-car accident types were also approximately equal. These results suggest that disaggregation based on accident types does not add to the understanding of the overall trend in the Netherlands.

A similar analysis was carried out on road safety time series and mobility time series, which were stratified by age and gender, and simultaneously analysed (Stipdonk et al. 2008). Demographic data were used to obtain mobility data per capita. To the difference of those based on accident-type disaggregation, the results of this second analysis showed that in the Netherlands, the number of fatalities per traffic mode and age group can be very different (see Figure 3). At the same time, it was investigated how well these differences could be predicted using changes in population data only.
The results showed that mobility per capita, stratified by age and traffic mode, changes relatively little over time. As an example, while the number of senior citizens has increased (and will be increasing further), the distance each member of this group travelled with various transport modes changed very little. As a consequence, even when mobility data by age group are not established regularly, a relatively good estimation of changes in road safety over time may be achieved by using demographic data instead of mobility data.

**Figure 3:** Estimated annual billion-km traveled by car drivers against their age (from 0 to 100 years old) and time (front to back), including forecasts to year 2020 in the Netherlands.

### 3.1.4. Explaining the risk

Another way to obtain a better understanding of the changes in aggregate road safety indicators is to use explicit (i.e. explanatory) variables for modelling these changes. The influence of a common external factor, the weather conditions, on the aggregate road safety level has been studied and illustrated using datasets from the Athens region in Greece, from France and from the Netherlands, over a long period (Stipdonk et al., 2008).

Precipitation, temperature and frost (negative temperature) were measured at a daily level at different meteorological stations and averaged over the stations, then over the month. It was demonstrated that the changes in weather variables were significantly correlated with the changes in the numbers of injury accidents and fatalities. However, the strength and even the sign of these correlations depend on the type of road. Further research was devoted to the inclusion of mobility figures in the models used, as these figures may explain the differences related to the type of road.

Apart from its explanatory property, the weather factor is useful for providing a road safety monitoring corrected for transitory factors, and could be included in the analysis of the short term trends at the national level in Europe.

Overall, these analyses reveals the different types of insights that can be obtained on the basis of different levels of analysis, starting from the examination of macroscopic trends, to the combined analysis of different risk levels (exposure, accident risk, and accident severity), to the disaggregation of these risks into different groups of the road traffic system, and eventually by incorporating in the analysis other explanatory risk factors.

### 3.2. Geographical analyses

Another group of analyses concerns those devoted to geographically organised data. These analyses have different objectives, but they nevertheless have one thing in common: The cases that belong to the same unit (road-site, country etc.) are likely to be more similar to each other than those belonging to different units, and this calls for a multilevel modelling approach. Multilevel analyses range from ‘pure’ spatial analyses, investigating the spatial relationships of road safety outcomes (e.g., the spatial analyses of Greek accident data), to analyses of specific road safety questions in the light of geographical dependencies among road safety outcomes (e.g., the investigation of spatial effects in drink-driving and enforcement).
3.2.1. Spatial analysis

Using Greek data, it was shown how the spatial structure of a country can be integrated in an analysis of accident data. In other words, it was demonstrated how the systematic “neighbourhood structure” in the accident/fatality data can be disentangled from those differences that occur purely at random. For this purpose, the accident and fatality numbers per county (NUTS-3 level) were extracted from the CARE database. This analysis showed that differences between counties with respect to the number of accidents or fatalities per inhabitant are, for some part, determined by their location: Neighbouring counties tend to be more similar than counties located far away from each other. These data can be used to create a road-safety map for the whole country, as in Figure 4, where it is shown that spatial dependencies are stronger for accidents than for fatalities per county.

*Figure 4: Spatial relationships between fatalities (left panel) and accidents (right panel) per population in Greece*

This example illustrates the principle of spatial modelling, which can also be applied to other countries or scaled up to larger regions (e.g. countries within Europe). In particular, it can be used to identify borders in road-safety. Such borders can be political or natural borders and spatial models can be used to test which regions share unobserved components that affect their accident and fatality occurrences. Questions like this are important when determining how broad a road-safety measure has to be applied in order to be effective or which areas can be candidate for isolated measures.

3.2.2. Spatial effects of drink-driving and enforcement

The spatial relationships between road safety outcomes need to be accounted for when examining the effects of various measures or performance indicators on road safety outcomes, especially when the data is collected at geographical level.

One example concerns the results of a Belgian roadside survey on drink driving (Vanlaar, 2005), where drivers were stopped at test-sites that were selected randomly with respect to location and time and it was established whether the drivers had been drinking-driving. Taking into account the geographical dependency of the data (i.e. drivers tested were nested into different sites) it was found that the time of testing was the most important predictor: Drink-driving on weekend nights exceeds by far that at all other time points. At the individual level, gender and age were the most notable predictors with men between 40 and 54 having the highest risk of drink driving. It was also shown that these variables (weekend night, male, 40-54) had the same effect on both the probability to have drunk slightly more than the legal limit and the probability to have drunk much more than the legal limit.

Another example concerns the effects of speeding and alcohol controls on the accident and fatality number for each Greek county (Yannis et al. 2007; Yannis et al. 2008). It turned out that both enforcement measures were highly correlated (i.e. counties that executed many alcohol controls also issued many speeding controls), and that they are together associated with lower fatality numbers. Moreover, it was shown that the enforcement measures were the most effective in those regions that had the highest accident rate in the first place. This may imply that at least part of the estimated effect of the enforcement measures is caused by a regression to the mean. It was also demonstrated that enforcement had a stronger overall effect on the number of fatalities than on the number of accidents as such, suggesting that the accidents became less severe.
3.2.3. Road-safety attitudes, self-reported behaviours, and accident data

In order to investigate whether national road safety attitudes and behaviours are related to road safety levels, the accident data from the CARE data base were linked to the SARTRE database, which contains attitude data concerning road safety from 13 European countries. Attitudes and behaviour were expressed as three principal components: (1) Aggressiveness and Speeding, (2) Other Unsafe Behaviour (no seat belt, drink driving etc.), and (3) Perceived Control Likelihood. A disaggregation per country, age and gender was carried out.

The analysis yielded the conclusion that a positive attitude towards speeding and aggressiveness was more frequent in age and gender groups that also have a higher number of accidents and fatalities. It is though important to note that this statistical relation does not prove that the attitudes shown by young drivers actually caused the accidents. However, it shows that a positive attitude towards speeding and aggressiveness is most typical of the problematic groups, and might therefore be the most promising attitude to be addressed in campaigns (Dupont & Martensen, 2008a).

3.3. Accident analyses

3.3.1. Risk and protection factors in fatal accidents

The third part of the analyses concern the analyses of specific accidents types, and aim at answering questions such as:

- What differentiates single from multiple vehicle fatal accidents?
- What is the probability of being killed, given that one is involved in a fatal accident? What factors affect this probability and thus the consequences of a severe accidents for the persons involved?
- How reliable are the injury severity scores assigned by the police to road accident casualties? Are there any factors systematically affecting the misreporting of injury severity?

In-depth fatal accident investigation data of ERSO for seven European countries were used on that purpose. Although the dependencies related to the accident (road users nested into vehicles etc.) were examined in all analyses, they were not found to be significant in this dataset.

The risk and protection factors in fatal accidents were examined in Dupont & Martensen (2008b). Comparing survivors and fatalities indicating risk and protection factors for road-users involved in fatal accidents, was a particularly complex issue, due to two methodological issues:

- The accident size effect, resulting from the fact that the lower the number of road users involved in a fatal accident, the higher the probability of fatality for each road user (given that there was a fatality in every accident contained in the data)
- The accident opponent effect, according to which, the more vulnerable the road user, and the heavier the accident opponent, the higher the fatality risk for that particular road user (see Figure 5).

Both effects result from the fact that, in the examined dataset, only the worst cases are present (fatal accidents), but not the more positive ones (non-fatal accidents) to compare them with.

These differences in baseline risk were taken in account in successive steps, each of which provided particular improvements in handling these methodological problems posed by the limitation of the data to fatal accidents. First the complete dataset, including the observations made on all road-user types (i.e.: car-drivers as well as pedestrians or heavy good vehicles) was analysed. The remaining analyses focused on car-occupants specifically and took the type of the opponent into account.
Some variables emerged as “risk factors” in a consistent way for all analyses, such as road-users who could be considered as “senior” (i.e.: more than 65 years old), or drivers who did not react properly to the occurrence of the accident by braking. In the specific case of car occupants, seatbelt appeared to be an important protection factor. The risk for the road-user to decease in the accident also appears to increase with the age of the vehicle. Finally, accidents that took place on a road junction also left the road-users with increased chances to survive. The risk for car occupants also depended on the area of main damage: Generally speaking, front damages were less dangerous than side damages. This was, however, true mainly for crashes between two cars and to a lesser extent for single vehicle accidents and collisions between cars and heavy- or light good vehicles. This latter result once more underlines the importance of including the transport mode of the opponent in the analysis.

3.3.2. Injury severity mis-reporting

A separate analysis was devoted to detecting inaccurate reporting of injury severity by comparing the original police reports to those of the Accident Investigation team (e.g. slight injuries recorded as serious, fatalities recorded as serious injuries, and so on). Generally speaking, the analysis performed offered encouraging conclusions with respect to the quality of the reporting of injury severity, at least in the member states that took part in the data-collection. Indeed, only a few sources of misreporting had been identified on the basis of this analysis, all of them among the first pilot wave of the data, suggesting a general pattern according to which, the more complex the road accident conditions were (e.g. more accident participants, higher traffic flow, night time), the higher was the probability of misreporting injury severity. No notable inaccuracy could be identified in the final data base, and no systematic country variation in the few inaccuracies identified.

Overall, the in-depth accident investigation data were proved to offer more detailed information than what is typically available in national or international databases, and also more reliable.
4. Conclusion

This paper provides a concise overview of analyses carried out within the SafetyNet project, using data of the European Road Safety Observatory. First, it was described how data structures like hierarchical accident data, geographical data, and time series data can pose particular problems in terms of statistical analysis. As a solution, the use of time-series analysis and multilevel analysis has been suggested. Moreover, a number of additional methodological problems that are specific to the various types of data used were also discussed, and solutions were offered to handle them.

For time series data, results at national and European level reveal that the development of the number of fatalities needs to be considered simultaneously with the development of mobility - and preferably so, separately for different subgroups (disaggregation). Accordingly, geographical structures need to be accounted for, either when analysing regional, national or European data, and spatial effects may play an important role when examining performance indicators, measures and interventions etc. Moreover, for the analysis of fatal accident data, it was emphasized that each analysis and interpretation of these data needs to take into account that this is a very specific selection of fatal accidents only.

Along with this summary of the methodological issues that have been dealt with in the analyses, a number of additional conclusions have been achieved using these data. A good insight was provided of what can be accomplished using the data that are - or will be - available through the ERSO as a result of the SafetyNet project. As shown on the basis of the various road safety research issues that have been addressed, there are many possibilities for addressing important safety questions by applying appropriate analysis techniques on the ERSO data.

Certainly, some limitations have also been encountered in using these data. Safety Net has initiated a process in terms of data collection, and this process is certainly not concluded. In the SafetyNet analyses, the data made available so far were optimally exploited, however more numerous data, and of improved quality may provide even more potential for analysis (especially as regards risk exposure data, in-depth data, behavioural data and measures and interventions data). The ERSO is currently most useful data source with comparable information at European level and its further development will eventually enable systematic monitoring of road safety trends and identification of best practises, in order to achieve maximum casualty reduction.

Acknowledgments

Several other researchers have contributed to the various analyses presented in this paper, namely Sylvain Lassarre (INRETS), Vojtech Ekserl (INRETS), Mohammed Cherfi (INRETS) and Dimitris Katsochis (NTUA). The authors would like to thank the SafetyNet partners who provided valuable comments and suggestions on the analyses summarized in this paper, within the framework of the SafetyNet internal review process. The authors would also like to thank all the SafetyNet partners and the Members of the CARE Experts group of DG-TREN who provided useful additional data for the analyses. Special thanks are addressed to the SafetyNet WP5 research teams for making available the Fatal Accidents Investigation database.
References


