Risk exposure data availability, collection methodologies and use in the EU

George Yannis
Assistant Professor
National Technical University of Athens
Athens, Greece
geyannis@central.ntua.gr

Eleonora Papadimitriou
Research Associate
National Technical University of Athens
Athens, Greece
nopapadi@central.ntua.gr

Philippe Lejeune
Head of Unit
Centre d'Etudes Technique de l'Equipement du Sud Ouest, France
philippe.lejeune@equipement.gouv.fr

Frits Bijleveld
Researcher
SWOV Institute for Road Safety Research
Leidschendam, Netherlands
frits.bijleveld@swov.nl

Joao Lourenço Cardoso
Head of Unit
Laboratório Nacional de Engenharia Civil
Lisbon, Portugal
joao.cardoso@lnec.pt

Abstract

The objective of this paper is the analysis of the current state-of-the-art in the EU as regards risk exposure data for road safety analyses, in terms of data gathering, availability and use. In this framework, the definitions and properties of the various exposure measures used in road safety analysis are examined (vehicle- and person-kilometres, vehicle fleet, driving licenses, road length, population, number of trips, and time in traffic), in light of their relationship to the needs and uses of risk exposure. Moreover, the methods of collecting exposure data at national level (travel surveys, traffic counts, and national registers) are presented and assessed, in order to determine the advantages and limitations of each method, which in turn determine the quality and usability of the available exposure data. Finally, the potential of international accident risk comparisons is discussed, through an investigation of the exposure data availability and quality in the International Data Files (EUROSTAT, ECMT, IRTAD, UN/ECE, and IRF). The results of the analysis allow for an overall picture of the current risk exposure data availability, collection methods and exploitation potential to be drawn, highlighting common practice and necessary steps towards an improved risk exposure data framework in the EU.

1. Introduction

In road safety analysis, exposure data is often used in order to obtain risk estimates, those being defined as the rate of the number of accidents (or casualties) divided by the amount of exposure of a population over a time period. Risk figures may be used for different purposes; however their main use concerns the comparison of safety performance among different units, populations or countries.

In theory, continuous exposure measurements of different road user categories in different modes and different road environments would be required and could provide detailed exposure estimates to the degree of disaggregation of the respective accident data. In practice, such

measurements are not possible. Consequently, road safety analyses need to compromise to some (approximate) estimates of exposure, which may be more or less accurate and representative of the true exposure of the examined population (Golias and Yannis, 2001).

Today, there is an important potential for road accident investigation at the international level, as a national framework for the collection, processing and analysis of accident data is operational in all EU Member States. The development of the CARE European system for the collection and analysis of accident data at EU level, including comparable disaggregate data, is a major step forward in this direction and now provides useful results. However, the absence of a system similar to CARE for exposure data collection and exploitation considerably limits the possibilities of reliable and useful analyses of accident data, including the analysis of risk rates and especially the international comparisons.

The objective of this research is the analysis of the state-of-the-art in risk and exposure data availability, collection methodologies and use in the European Union (NTUA, 2005). This work was carried out within the scope of the SafetyNet project of the 6th Framework Program for Research, Technological Development and Demonstration of the European Union. More specifically, the analysis aims to explore the concepts of exposure and risk, as well as the theoretical properties of the various exposure measures in use in road safety. Moreover, it aims to present an overall picture of the existing methods for collecting exposure data for national risk estimates. Finally, the potential of international risk comparisons is investigated, mainly through the International Data Files with exposure data.

In order to meet these objectives, the following methodology was adopted: firstly, an exhaustive bibliography review was carried out. Additionally, a set of National Reports was created by the Institutes involved in the analysis, providing representative examples of exposure data availability, collection methods and use from seven representative European countries: Denmark, France, Greece, Hungary, Norway, the Netherlands and Portugal. Furthermore, a separate survey was devoted to the investigation of the International Data Files, as far as exposure data availability and quality is concerned. The survey was carried out by means of personal interviews with the maintainers of the related databases of several international organizations.

The results of the analysis allow for an overall picture of the current risk exposure data availability, collection methods and exploitation potential to be drawn, highlighting common practice and necessary steps towards an improved risk exposure data framework in the EU.

2. Needs and uses of risk and exposure data

In road safety analysis, exposure data can be used in two manners (Hakkert and Braimaster, 2002, Hauer, 1995):

- To obtain risk data in the form of outcome per unit of exposure.
- To describe differences in the road safety situation.

An example of the first type is the number of fatalities per inhabitant. Such a figure is called a risk. In general it is defined by "the number of outcomes" divided by "the amount of exposure". An example for the second case is the number of motor vehicles per inhabitant.

In road safety analyses, different exposure measures are used, according to data availability and quality, as well as the particular objective of the analysis. These measures may vary significantly in terms of the potential level of disaggregation and the possible underlying bias in their estimates. It should therefore be noted that no general rule is available concerning the preferred measures of exposure. Vehicle- and person kilometres of travel, as well as the time spent in

traffic, are conceptually closer to the theoretical definition of exposure and can be theoretically available to a satisfactory level of detail. However, under certain conditions, other available exposure measures may be equally efficient for the purposes of a particular analysis and / or more reliable e.g. vehicle fleet, road length, drivers population etc. These alternative exposure measures may also have other, explanatory or descriptive uses.

Figure 1 summarizes the characteristics of the different exposure measures used in road safety analysis. It should be noted, however, that the features presented in the Table concerns the theoretical potential of exposure measures. In practice, the availability, quality and disaggregation level of exposure measures may be compromised by limitations and particularities of the respective collection methods. For example, sampling methods may impose a series of errors in the estimates. Additionally, the use of data sources that were not designated to provide exposure data may result to difficulties in the full exploitation of the data.

A detailed discussion on the different collection methods of exposure data and the respective effects on the usability of the estimates, are presented in the section 3 of this paper.

	Road safety Outcome		**Context**			**Variation**		**Disaggregation potential**			
Exposure measure	Accidents	Persons	Traffic	Mobility	Epidemiology	Temporal	Regional	Road User category	User characteristics	Vehicle characteristics	Road characteristics
Vehicle - kilometres	•	•	•			•	•		•	•	•
Person - kilometres		•	•	•		•	•	•	•	•	•
Road Length	•		•				•				•
Vehicle Fleet	•		•							•	
Population		•			•		•	•	•		
Driver population		•	•						•		
Number of trips	•	•	•	•		•	•	•	•	•	•
Time in traffic		•	•			•	•	•	•	•	•

Figure 1. Comparison of exposure measures

3. Methods for collecting exposure data in the EU

In this section, the main collection methods for "raw" exposure data are presented by means of examples from several European countries. The analysis focuses on methods involving an intrinsic statistical error (i.e. sampling-based methods). However, other methods concerning more objective measurements (i.e. databases and census) are also described and assessed. A general conclusion that can be drawn is that there is no unique or standard method for obtaining the same exposure measure. Accordingly, different exposure measures can be derived out of data collected by means of one method. Moreover, it is obvious that there is much less disaggregation potential in exposure data, even those collected by the more detailed methods, compared to accident data.

Vehicle- and passenger-kilometres of travel, as well as time spent in traffic can be collected through (national) travel surveys, allowing obtaining information by both person, vehicle and road network characteristics. The main advantage of the travel surveys (compared to non-survey collection methods) is that these surveys have persons as a unit, making it possible to compare groups of persons. However, these surveys are carried out by personal interviews on a sample of the entire population (although in some cases an age threshold is in place) and therefore the data

obtained are, optimally, only an acceptable approximation of the actual risk exposure. Additionally, a number of possible biases (sampling, non response or measurement errors) may occur and should be treated accordingly where possible (Cochran, 1963). For example, experiences with travel surveys indicate that short travels (e.g. by foot or by bicycle) are often not reported, while motorized trips are often overestimated.

The international comparability among the produced exposure data is often limited, mainly because of several incompatibilities among the national definitions (road network, vehicle categories etc.) and/or characteristics (different use of various transport modes in different countries e.g. mopeds and motorcycles). Moreover, travel surveys often have main purposes other than to provide exposure data (FHWA, 1994). Consequently, the different definitions between travel surveys and accident databases often create problems when travel surveys are used for exposure purposes.

On the other hand, traffic counts systems, which are also widely used for vehicle-kilometres estimates, are not suitable to distribute exposure according to person characteristics (age/gender groups). The seasonal (e.g. weekly, daily, hourly) variation of exposure can be estimated by means of traffic counts, as the measurements are usually continuous over time. Traffic counts may give good estimates of average annual daily traffic (AADT), but there are practical problems involved in calculating vehicle-kilometres from AADT.

Additionally, this method is also sample-based, and the measurement points may or may not be representative of the national / regional traffic, as in most cases the systems are operational on the principal National and interurban road network (local or urban roads often not included). Problems are also encountered in the classification by vehicle type; in some traffic counts systems the level of detail is insufficient, and in other cases a bias in the detection of particular vehicle categories (e.g. two-wheelers) is observed.

The two methods discussed above present different advantages and limitations, however they are the only methods that can produce vehicle-kilometre estimates. However, because of the difficulties in the implementation and operation of such systems, in most countries the vehicle fleet and driving licenses national registers are also used to calculate exposure. A problem when using such registers to estimate risk is that these are certainly very crude estimates of exposure, giving quite unreliable risk estimates.

Quite often the registers are optimally used to calculate risk in combination with sample studies (travel or mobility surveys) of average driving distances, resulting to vehicle-kilometre estimates. However, in most cases the number of registered vehicles and the number of licensed drivers are directly used as exposure measures. It should also be noted that data from such databases are known sometimes to lead to some (often uncalculated) overestimations, due to insufficient updating of the registers; scrapped vehicles are not always removed from the vehicle fleet files and deceased drivers are not always removed from the driving licenses' files, due to the fact that, in both cases the registers were not created to provide exposure data (NTUA, 2005). More accurate estimates of the actual number of (active) vehicles may be obtained through vehicle inspection databases (not available in most countries) or vehicle taxation databases (not accessible in most countries).

As far as availability of road network length data is concerned, in most countries the available information concerns the National road network (motorways, national roads etc.), whereas more detailed information e.g. roadway geometry, is less available. Regional/local road network length estimates may also be partly available at regional/local authorities. The growing use of advanced methods (e.g. digital mapping, GIS etc.) is expected to improve data availability and quality in the coming years.

In Figure 2, the main characteristics (exposure measures, variables and values collected) of the various collection methods are presented in several European countries.

Method	National Travel Surveys							Traffic counts					Vehicle register				Driver licenses					Road length			
Country	**Distance travelled**	**Time spent in travel**	-by gender	- by age	- by experience	- by mode*	- by road type	**AADT**	**Traffic volume**	- hourly variation	- seasonal variation	- vehicle classification	**New entries**	**Scrapped vehicles**	- by vehicle type	- by vehicle age	**New entries**	**Deceased drivers**	-by gender	- by age	- by license type	- National roads	- Regional roads	- Local roads	- Intersections
Norway	•		•	•		•		•	•	•		•	•		•		•		•	•	•	•	•		•
Greece**	•		•	•	•	•	•	•	•	•	•	•	•	•	•••	•	•	•	•	•	•	•	•		•
Portugal		•				•••	•	•	•	•	•	•••	•		•		•	•	•	•	•	•	•		
Netherlands	•	•	•	•		•••		•					•		•••		•		•	•	•	•	•	•	•
France	•	•	•	•	•	•••	•	•	•	•	•		•		•••	•	•		•	•	•	•	•	•	
Hungary								•	•	•	•	•••	•		•••		•		•	•	•	•	•		
Denmark			•	•		•••			•	•	•	•	•		•••		•				•				

** more bullets indicate a more detailled classification*
*** the travel survey in not official; traffic counts system was operational up to 1993*

Figure 2. Exposure data collection methods and variables in selected EU countries

From the analysis, the following conclusions can be drawn:

- The features and specifications of each method may vary significantly among countries
- Accordingly, the availability, disaggregation and comparability of exposure measures (in terms of definitions, variables and values) are quite diverse.
- The disaggregation level theoretically possible for an exposure measure is seldom achieved in practice
- Data from different sources (collection methods) are often used to produce a national exposure estimate, i.e. different data sources may function complementarily for the calculation of a single exposure measure
- In general, it is not always clear how the exposure estimates are obtained from the "raw" data collected by means of the various methods.

Consequently, national exposure and risk estimates, when available, are seldom comparable at EU level, especially as far as vehicle- and passenger-kilometres are concerned.

4. International data files with exposure data

These national risk exposure estimates are collected, exploited and published through a number of International Data Files (IDF) in the field of transport and road safety. The main IDF involved in road accident and exposure data in the EU are the following:

- Eurostat (European Commission, Luxembourg)
- ECMT (European Conference of Ministers of Transport, Paris)
- UNECE (United Nations - Economic Commissions for Europe, Geneva)
- IRTAD (International Road Traffic Accident Database, Koeln)
- IRF (International Road Federation, Geneva)

These data files are useful and accessible data sources, as a result of several decades of important data collection efforts. However, they have different objectives; they collect different data in different forms and structure, and are maintained by organizations with different scopes and policies. In particular, although the main data sources are national authorities, in some cases (IRTAD, IRF) other sources are also used (e.g research results, other studies at national, regional or local level, private sources etc.), complicating data comparability among IDFs. Moreover, the availability of exposure data among the data files varies significantly, in terms of both countries and years availability, and variables and values availability.

It is interesting to notice that the exposure data available in the IDFs are in a much more aggregate form that the exposure data collected at national level, as reported by the countries. In Figure 3 an overview of the IDFs examined in the framework of the present analysis is presented, focusing on the availability of exposure data and the related disaggregations. Moreover, a detailed comparison of exposure data published by the IDFs was carried out, in terms of availability and quality, and several interesting results and conclusions were obtained:

- It appears that the more disaggregate national exposure data are not exploited within the context of IDFs.
- Significant differences are observed among the IDFs in the published figures for each exposure measure; these differences are more important for the more "sophisticated" exposure measures (i.e. vehicle and passenger kilometres).
- These differences may be partly due to the different national sources and definitions used
- However, another reason may concern insufficient data quality control within the IDFs.

Despite the limitations discussed above, the considerable effort made during the last decades for gathering and exploitation of road safety related data is clearly reflected to these IDF (ECMT, 2004, IRF, 2005, IRTAD, 1998, UNECE, 2005). The fact that there are various IDF for exposure data at European level is positive for the road accident statistics users, because they can choose from a variety of information. The objectives and scopes of these data files, as well as the quantity and quality of available data contained inside the IDF, differ among the various data providers, making them to function complementarily in most of the cases. Consequently, particular caution is required from the data users, in order to optimally use the available information in reliable road safety analyses.

	Data File description									**Data availability**					
Data file	Number of countries	Examined time series	Transport statistics	Accident statistics	Other statistics	Collection method	Disaggregate / Aggregate	Data Access	Data quality control	Vehicle-kilometers by mode	Passenger-kilometers by mode	Number of vehicles by type	Number of drivers	Population by gender/age	Road length by road type
Eurostat	25	1960-	•	•	•	Quest.	Agg.	free	limited	•	•	•		•	•
ECMT	50	1960-	•	•			Agg.	free	limited	•	•	•		•	
UNECE	55	1960-	•	•			Agg.	free	limited	•	•	•		•	•
IRTAD	29	1970-	•	•	•	Quest.	Agg.	members	limited	•	•	•		•	•
IRF	84	1995-	•	•	•	Quest.	Agg.	members		•	•	•			•

Quest: questionnaire

Agg.: aggregate

Figure 3. Overview of exposure data in the International Data Files

5. Discussion and recommendations

Summarizing, the availability and quality of risk exposure estimates in the EU Member States varies significantly, and is related both to the exposure measures used and the characteristics of the respective collection methods. In particular, significant efforts are made at national level to improve data availability, disaggregation and reliability; however the lack of a common European framework for the collection and exploitation of exposure data limits significantly the comparability of the detailed national data. On the other hand, the International Data Files containing road safety related data, including exposure data, provide useful aggregate information in a systematic way and are currently the only sources allowing international comparisons; however more effort is required to further improve the availability and quality of these data.

It can be deduced that a series of problems, namely poor data availability, insufficient reliability, inappropriate disaggregation (in relation to accident data) and limited accessibility are the main limitations to the full exploitation of exposure data at European level. It is also obvious that the most useful exposure data are the least available.

Further work and research should therefore focus on data compatibility and availability, namely through a common framework including common data requirements and definitions and a pan-european data collection system. In particular, this framework should focus on the collection of disaggregate time series of exposure data by road user, mode and network characteristics, and should be organized to provide data in a consistent and systematic way.

Within this framework, it can be deduced that vehicle- and person-kilometres of travel are the most appropriate exposure measures, especially in the context of road traffic safety analysis, as they are closer to a theoretical concept of exposure and can be estimated at a satisfactory level of disaggregation (i.e. combined by user, vehicle and road characteristics). Consequently, a common European framework should mainly focus on these exposure measures.

Accordingly, it is obvious that different collection methods may be used for vehicle- and person-kilometres estimates, namely travel surveys and traffic counts, each one presenting different features and difficulties. In particular, travel surveys, being more flexible in their design, may provide a higher level of disaggregation, having both persons and vehicles as units. On the other hand, traffic counts systems are the only method, which practically can provide continuous exposure measurements over time. Consequently, a common exposure data collection framework should include both travel survey and traffic counts elements. The specific elements of the calculation process of exposure measures would be an important and complex task.

Certainly, the establishment and operation of such a system would be a complex and time-consuming task, which would also involve a significant effort and cost. However, given the importance of an improved risk exposure data availability and quality, to support and monitor an efficient road use and safety policy at EU-level, it is necessary to promote its development.

In order to deal with the current exposure data needs, the gathering and harmonization of the existing information shall certainly contribute to the improvement of the potential for exploitation of the existing exposure data.

Acknowledgment

This paper is based on work carried out within the scope of the SafetyNet project of the 6th Framework Program for Research, Technological Development and Demonstration of the European Commission.

References

1. Golias, J. and Yannis, G.., (2001).'Dealing with lack of exposure data in road accident analysis'. *Proceedings of the 12th international Conference Traffic Safety on three Continent*, Federation of European Road Safety Institutes, Transportation Research Board, Moscow, September 2001.
2. NTUA - National Technical University of Athens (2005). *State of the art Report on Risk and Exposure Data*. Deliverable 2.1 of the SafetyNet IP - Building the European Road Safety Observatory, European Commission, Brussels, www.erso.eu.
3. Hakkert, A. S. and Braimaister, L. (2002). *The uses of exposure and risk in road safety studies*. SWOV report R-2002-12. SWOV, Leidschendam, the Netherlands.
4. Hauer, E. (1995). *On exposure and accident rate*. Traffic Engineering and Control, 36 (3), pp. 134-138.
5. Cochran, W.G (1963). *Sampling Techniques*. Second edition, Wiley, 1963, third edition, 1977.
6. FHWA - Federal Highway Administration (1994). *FHWA Study Tour for National Travel Surveys*. USDOT - FHWA, Washington DC.
7. ECMT (2004). *Annual Report 2004*. Joint OECD/ECMT Research Center. May 2004.
8. IRF (2005). *World Road Statistics - 2004*. IRF, Geneva.
9. IRTAD (1998). *Definitions and data availability*. OECD - RTR Road Transport Research Programme. BASt Germany, June 1998.
10. UNECE (2005). *Annual bulletin of transport statistics for Europe and North America*. UNECE, Geneva.