Deliverable 2.1.
State of the Art Report on Risk and Exposure Data

Contract No: **Integrated Project No. 506723: SafetyNet**
Acronym: **SafetyNet**
Title: **Building the European Road Safety Observatory**

Integrated Project, Thematic Priority 6.2 "Sustainable Surface Transport"

Project Co-ordinator:
Professor Pete Thomas
Vehicle Safety Research Centre
Ergonomics and Safety Research Institute
Loughborough University
Holywell Building
Holywell Way
Loughborough
LE11 3UZ

Organisation name of lead contractor for this deliverable:
NTUA - National Technical University of Athens

Due Date of Deliverable: **30/10/2005**

Submission Date: **30/10/2005**

Report Author(s):
George Yannis, Eleonora Papadimitriou, **NTUA,**
Philippe Lejeune, Vincent Treny, **CETE-SO,**
Stig Hemdorff, **DRD,**
Ruth Bergel, Mouloud Haddak, **INRETS,**
Péter Holló, **KTI,**
Joao Cardoso, **LNEC,**
Frits Bijleveld, Sjoerd Houwing, **SWOV,**
Torkel Bjørnskau, **TØI**

Project Start Date: **1st May 2004** **Duration:** **4 years**

Project co-funded by the European Commission within the Sixth Framework Programme (2002 -2006)		
Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (inc. Commission Services)	X
RE	Restricted to group specified by consortium (inc. Commission Services)	
CO	Confidential only for members of the consortium (inc. Commission Services)	

Transport
Project co-financed by the European Commission, Directorate-General Transport and Energy

Table of Contents

Table of Contents **2**
Index of Figures **4**
Index of Tables **4**
Executive Summary **5**

1. Introduction **10**
1.1. Objective of the analysis 10
1.2. Methodology and structure 10
1.3. Risk exposure data background and definitions 12
References 16
2. Risk exposure data in road safety analysis **17**
2.1 Statistical background and limitations 17
2.1.1. The Poisson Limit Theorem for accident counts 17
2.1.2. Over-dispersion 19
2.1.3. Normal approximations 19
2.1.4. The case of victims counts 19
2.1.5. The role of trials in exposure 20
2.2 Needs and uses of risk figures 21
2.2.1. The need of risk figures 21
2.2.2. General definition of risk 21
2.2.3. Definitions and usability of risk 21
2.2.4. Statistical implications 23
2.2.5. Other uses of risk exposure data 24
2.2.6. Discussion 24
2.3. Risks in use in road safety analysis 26
2.3.1. Vehicle kilometres 28
2.3.2. Person kilometres 28
2.3.3. Road length 29
2.3.4. Fuel consumption 29
2.3.5. Vehicle fleet 29
2.3.6. Population 30
2.3.7. Driver population 30
2.3.8. Number of trips 30
2.3.9. Time in traffic 30
2.3.10. Other methods for risk estimates 31
2.4. Synthesis 32
References 34
3. Review of methods for collecting RED in the EU **35**
3.1. Introduction 35
3.2. Travel surveys 37
3.2.1. The need for surveys 37
3.2.2. The sample framework 37
3.2.3. Survey errors 38
3.2.4. Exposure measures collected and disaggregation level 39

3.2.5. Limitations 40
3.3. Traffic counts systems 41
3.3.1. General description 41
3.3.2. Sampling framework and continuity over time 41
3.3.3. Exposure measures collected and disaggregation level 42
3.3.4. Limitations 43
3.4. Vehicle fleet databases 44
3.5. Driver licenses databases 47
3.6. Road registers 49
3.7. Other methods 50
3.7.1. Model estimation of vehicle kilometres based on fuel sales 50
3.7.2. Odometer readings at regular vehicle inspections 53
3.8. Synthesis 54
References 57
4. International data files with exposure data 58
4.1. General 58
4.2. Brief description of the analysed IDF 60
4.2.1. Eurostat 60
4.2.2. ECMT 60
4.2.3. UNECE 60
4.2.4. IRTAD 61
4.2.5. IRF 62
4.3. Comparative analysis of the collected RED 63
4.3.1 Road length 63
4.3.2 Vehicle kilometres 66
4.3.3 Transport measurement 68
4.3.4 Registered vehicles 70
4.3.5 Population 73
4.3.6 Driving licenses 75
4.3.7 Fuel consumption 75
4.3.8. Number of trips 75
4.3.9 Time in traffic 75
4.4. Synthesis 76
References 80
5. Conclusions and recommendations 81
5.1. Summary of the state-of-the-art 81
5.2. Recommendations 88

Annex I - Expectation and variance of a ratio 90
Annex II - Examples of practical implementation of surveys in the EU 91
Annex III - Survey errors 98
Annex IV - Examples of practical implementation of traffic counts systems in the EU 103
Annex V - IDF data gathering questionnaire 111
Annex IV - IDF responses to the questionnaire 114

Index of Figures

Figure 1.1. Structure of activities within the state-of-the-art survey 10
Figure 2.1. Example of comparison of different risks (traffic approach versus epidemiology approach) 27
Figure 2.2. Example of comparison of different risks based on IRTAD data for 2003 27
Figure 4.1. Comparison of motorway length data in EUROSTAT and IRF .. 65
Figure 4.2. Comparison of main highway length data in EUROSTAT and IRF 65
Figure 4.3. Comparison of secondary highway length data on EUROSTAT and IRF 66
Figure 4.4. Comparison of distance traveled by passenger cars in EUROSTAT, IRF, ECMT and IRTAD 67
Figure 4.5. Comparison of distance travelled on private vehicles in EUROSTAT, IRF and ECMT 69
Figure 4.6. Comparison of distance traveled on public vehicles in EUROSTAT, IRF and ECMT 70
Figure 4.7. Comparison of number of cars in EUROSTAT and IRF 72
Figure 4.8. Comparison of number of buses, lorries and tractors in EUROSTAT and IRF 72
Figure 4.9. Comparison of number of motorcycles and mopeds in EUROSTAT, IRF and UNECE 73
Figure 4.10. Comparison of the number of country inhabitants on EUROSTAT, IRF, ECMT and IRTAD 74

Index of Tables

Table 2.1. Advantages, limitations and optimal use of different exposure measures 33
Table 3.1. Overview of travel surveys in the EU Member States 40
Table 3.2. Overview of traffic counts systems in the EU. 43
Table 3.3. Overview of vehicle registration databases in the EU 46
Table 3.4. Overview of driving license databases in the EU 48
Table 3.5. Overview of road registers in EU countries 49
Table 4.1. Road classes considered in different IDFs 64
Table 4.2. Exposure data availability* on the analysed IDF 76
Table 4.3. Risk indicators in international data files 78
Table 5.1. Comparison of exposure measures..... 82
Table 5.2. Exposure data collection methods and variables in selected EU countries..... 85
Table 5.3. Overview of exposure data in the International Data Files 86

Executive Summary

The **objective** of this Report is the analysis of the state-of-the-art in risk and exposure data availability, collection methodologies and use in 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 and a bibliography database on risk and exposure data was developed. 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 the following organizations: EUROSTAT, ECMT, UNECE, IRTAD and IRF.

From the results of the analysis, it was deduced that, **comparing risk rates**, especially at international level, may be a very complex task. Both accident counts and exposure measures present some theoretical and practical limitations and are subject to estimation errors, which may compromise their usability. Especially as far as exposure is concerned, 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 accidents data. In practice, such measurements are not possible, therefore, road safety analyses need to compromise to some **approximations of the actual exposure**, which may be more or less accurate and representative. Different exposure measures may be used, according to data availability and quality, as well as the context of the analysis. It should be noted that no general rule can be adopted on the preferred measures of exposure.

However, it can be deduced that the most appropriate measures of exposure appear to be **vehicle- and passenger-kilometres of travel**, because they are closer to the theoretical concept of exposure and can be available, in theory, to a satisfactory level of detail. However, **other exposure measures** are often used, namely the vehicle fleet and the drivers' population, the road

network length, the fuel consumption, as well as the entire population, mainly because they involve less complex collection methods.

The theoretical features of the various exposure measures were analyzed in detail in the framework of the present research. In practice, however, the availability, quality and disaggregation level of exposure measures may be compromised by limitations and particularities of the respective **collection methods**. The main sources of exposure data include travel surveys, traffic counts systems, vehicle fleet register, driving licenses registers, roads registers and population registers.

Travel surveys are carried out in most European countries, in order to collect information on traffic and mobility patterns. From the data collected (namely distance traveled, time spent in traffic and number of trips), vehicle- (actually driver-) and passenger-kilometres estimates can be obtained. The main advantage of **national travel surveys** (compared to other collection methods) is that these surveys have persons as a unit, making it possible to compare groups of persons, and are usually designed to achieve a high level of data disaggregation by person, vehicle and road network characteristics. However, travel surveys are sample-based self-reporting information collection methods, consequently a number of possible biases (sampling, non response or measurement errors) may occur and should be treated accordingly where possible.

On the other hand, in most countries **traffic counts systems** are in place, providing data on traffic volumes, which are used to obtain vehicle kilometres estimates. An important advantage of using this method is that the seasonal variations of exposure can be captured, as the measurements are usually continuous over time. However, this method is does not allow distributing exposure by to person characteristics. Additionally, this method is also sample-based, in the sense that measurement points are placed on specific sections of the main road network, which may or may not be representative of the entire road network, and usually local or urban roads are not included. Problems may also be encountered in the automatic classification of vehicles.

The two methods discussed above present different advantages and limitations, however they are the only methods that can produce detailed vehicle- and person-kilometres estimates. However, because of the difficulties in the implementation and operation of such systems, in most countries the **vehicle fleet and driving licenses registers** are also used to calculate alternative exposure measures. The problem when using such registers to estimate risk is that these are certainly very crude estimates of exposure, giving quite uncertain risk estimates. It should be noted that, data from such databases are known to lead to some (but often uncalculated) overestimations, due to insufficient updating of the registers.

Accordingly, **roads registers** are often used to apply the length of roads as an exposure measure. However, in most countries the available information concerns the main road (motorways, national and rural roads etc.), whereas information on roadway geometry is less available, and regional/local road length estimates are less available.

From the analysis of examples of implementation of the above methods in the selected European countries, 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) is 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.
- According to the above, it can be deduced that the national exposure and risk estimates may not always be comparable at EU level.

However, in most countries some national exposure estimates are available, which are collected, exploited and published through the **International Data Files** (IDFs) in the field of transport and road safety. The main IDFs involved in road accident and exposure data EU are the following: Eurostat, ECMT, UNECE, IRTAD and IRF. These data files are useful and accessible aggregate 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, in some cases by different national sources, and are maintained by organizations with different scopes and policies. Consequently, the availability of exposure data among the data files varies significantly, in terms of both countries and years available, and variables and values available.

In the framework of the present analysis, 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:

- The exposure data available in the IDFs are in a much more aggregate form than the exposure data collected at national level
- Accordingly, 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 are partly due to the different national sources and definitions used
- However, another reason may concern insufficient data quality control within the IDFs.

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 RED limits significantly the comparability of the detailed national data**. On the other hand, the International Data Files containing road safety related data, including RED, 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 and limited accessibility are the main limitations to the full exploitation of risk and exposure data at European level. It is also obvious, from the analysis presented in this Report, that the most useful RED are the least available. Further work and research should focus on improving data compatibility and availability, namely through a common framework including **common data requirements, definitions and collection methods**.

In particular, from the results of the state-of-the-art survey on risk and exposure data, which was carried out in the framework of the present research, the following **recommendations** are suggested, towards a common risk exposure data framework:

- Priority should be given to the collection of **vehicle- and person-kilometres** of travel, these measures being the most appropriate exposure measures in the context of road and traffic safety analysis.
- The common framework should focus on the collection of **disaggregate time series of exposure** by road user, mode and network characteristics, and should be organized to provide data in a consistent and systematic way.
- Consequently, **both travel survey and traffic counts** methods should be exploited, allowing for flexibility, high level of disaggregation and continuity over time in the exposure estimates.
- **Additional data sources** could be exploited to benchmark or validate the exposure estimates and improving data reliability and accuracy

- The specific **calculation process of exposure** measures should be defined and standardized.

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

However, in order to deal with the current needs, the gathering and harmonization of the existing information is equally important for the improvement of the **exploitation potential of the existing exposure data**. The harmonization of the definitions of exposure measures, variables and values between countries (at the most disaggregate level), as well as within the International Data Files (aggregate level), in accordance to the existing accident data, as well as the current and future exposure data needs, would be an important first step to improve comparability of the existing disaggregate data.

1. Introduction

1.1. Objective of the analysis

This Report presents the results of a bibliographical study and review of the current methodologies and practices dealing with Risk Exposure Data (RED) in the EU. The main objective of the analysis consists in getting a good knowledge of the current RED state-of-the-art in the EU, in terms of data gathering, availability and use.

In particular, the analysis aims at analyzing the concept of risk in road safety and the relationship between risk and exposure. In this framework, the definitions and properties of the various exposure measures used in road safety analysis are presented.

Moreover, the methods of collecting exposure data at national level are presented and assessed, in order to determine the advantages and limitation 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.

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.

1.2. Methodology and structure

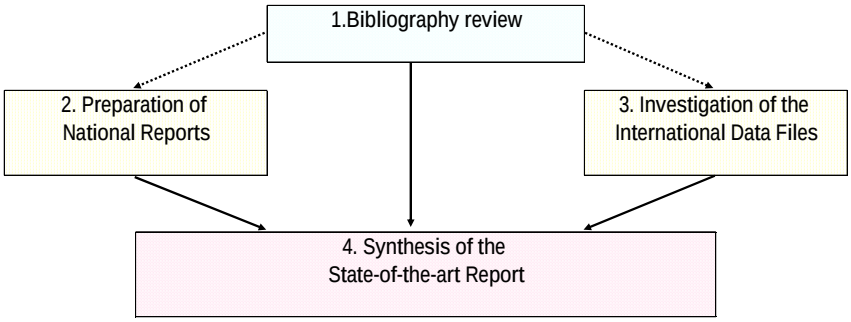


Figure 1.1. Structure of activities within the state-of-the-art survey

The methodology adopted in order to meet the objectives of the analysis is presented in Figure 1 above.

1. An exhaustive bibliography review on risk exposure data availability, collection methodologies and use in the EU was carried out. The results were summarized in a bibliographic database of documents related to risk exposure data.

2. A set of National Reports was prepared by the partners involved in the analysis, concerning risk exposure data collection methods and use in their country. The purpose was to obtain examples for exposure data typical collection and exploitation methodologies from different countries (Denmark, France, Greece, Hungary, the Netherlands, Norway and Portugal).

   Each National Report focused on the characteristics, advantages and difficulties encountered within the following exposure data collection methods:

   ● National Travel Surveys
   ● Traffic Counts systems
   ● Registered Vehicles databases
   ● Driver license databases
   ● Road registers
   ● Other sources of exposure data

   Additionally, studies on exposure data use at national or regional level were described and assessed in the report. The various examples were used in the analysis and are also available in the Annexes of the present Report.

3. A separate inquiry was devoted to the investigation of the International Data Files, as far as risk exposure data availability and quality is concerned. The participation of EUROSTAT in the technical meetings that were held in the framework of the analysis allowed for the collection of useful information. Additionally, a complete survey including visits to several International Data Files (EUROSTAT, ECMT, IRTAD, UN/ECE, IRF) was carried out, in order to collect information on all the RED related issues (collection methods, data availability and comparability).

   The detailed results of the investigation of the International Data Files are also available in the Annexes of the present Report.

4. On the basis of the results of the above activities, the synthesis of the state-of-the-art was carried out.

In particular, in the present **Chapter 1** of the Report, the objectives of the analysis are presented and the methodology adopted is described. Moreover,

a general background of risk exposure data issues is presented, in terms of definitions and needs.

Chapter 2 concerns the use of risk exposure data in road safety analyses. The theoretical concept of risk is initially presented, and the basic statistical properties of accident and exposure data are discussed. Moreover, the theoretical relationship between accidents, exposure and risk is investigated. Within this framework, the different uses of risk figures are presented. Finally, the properties of the different exposure measures that can be used are presented, and their usability is assessed according to their properties and the context of the analysis.

In **Chapter 3**, the main methods for collecting risk exposure data are presented. The analysis aims at highlighting the features, advantages and limitations of each method, and showing how these parameters may compromise the usability of data in practice, in relation to their theoretical properties. It should be noted that the analysis focuses on the estimation errors that may rise from the collection method, affecting the calculation of the national exposure estimates. For each method, implementation examples from several EU countries are presented.

In **Chapter 4**, the results of the investigation of exposure data availability and quality in the International Data Files are presented, in order to highlight the current exploitation of the available national exposure estimates. Firstly, the main characteristics of the examined International Data Files are presented. Additionally, the available national exposure figures are compared among the International Data Files, and an overall assessment of the potential of international comparisons is carried out.

Finally, in **Chapter 5**, the state-of-the-art on risk exposure data is summarized. The main conclusions drawn from the review of current exposure data related practices are presented, in terms of advantages and limitations of the different exposure measures, the related collection methods and the usability of the produced national exposure estimates. According to the above, the need for a common framework for risk exposure at EU level is discussed, and a set of general guidelines is suggested towards this goal.

1.3. Risk exposure data background and definitions

When accidents or fatalities are to be compared between countries, the scale of different counties has to be considered: all other conditions equal, large populations tend to have more fatalities than smaller populations. The number of fatalities also depends on the number of vehicles i.e. on the level of motorisation. For instance, large populations with low level of motorisation have relatively low number of fatalities (Smeed, 1968, Holló, 1997). In this

example, the population and vehicles figures are used as measures of exposure. A number of other measures may be used for this purpose as well. It is common practice to use the ratio of the number of accidents or casualties, divided by the amount of exposure for a comparison between counties. Such a ratio is called a "risk".

As far as accident data are concerned, the systematic efforts at EU level led to the creation of the CARE accident database, which includes the fifteen national accident databases with a significant number of harmonized data (common variables and values). This allows for comparisons of absolute numbers (counts) of accidents and related casualties at a satisfactory level of disaggregation (i.e. combined analysis per person, vehicle, road network and other characteristics).

On the contrary, as far as exposure data are concerned, a series of difficulties and limitations are encountered in road safety analyses. The problem that most often arises is that it is unclear whether exposure data are available and, when they are available, whether exposure data are comparable. In particular, exposure data collection efforts are carried out at the national level only, whereas no standard methodology for the collection and use of the data exists. The exploitation of this data is carried out mainly through the International Data Files, in which the national exposure estimates are available; however data comparability issues are to be dealt with.

The main purpose of the present analysis is to create the overall picture on current exposure data availability, collection methods and use, and identify the necessary improvements towards a common European framework for exposure data comparable over the Member States, to be combined with the existing road accident data (e.g. the CARE database) in order to provide usable risk estimates.

In order to reach this objective, it is necessary to acquire a good knowledge of the state-of-the-art on the collection of exposure data in the different Member States. It should be noted that the analysis focuses on a set of exposure measures commonly used in practice, as follows:
- Road Length
- Vehicle Kilometres
- Person Kilometres
- Fuel Consumption
- Population
- Driver Population
- Vehicle Fleet
- Number of Trips
- Time in Traffic

In Chapter 2, the different properties and uses of these exposure measures are discussed in detail. In the following section, the definitions adopted by

EUROSTAT in “The Glossary for Transport Statistics” (European Commission, 2003) for the above measures are presented.

Road Length

- Definition of "road”: Line of communication, (traveled way) using a stabilized base other than rails or airstrips open to public traffic, primarily for the use of road motor vehicles running on their own wheels. Included are bridges, tunnels, supporting structures, junctions, crossings and interchanges. Toll roads are also included. Excluded are dedicated cycle paths.
- Unit: km

Vehicle kilometres

- Definition: "Vehicle kilometres" of a country is defined as the total number of kilometres travelled within the borders of the country by road vehicles, where "road vehicle" is a vehicle running on wheels and intended for use on roads.
- Unit: vehicles x km

Person kilometres

- Definition: "Person kilometres" of a country is defined as the total number of kilometres travelled within the borders of the country by persons, regardless of their age.
- Unit: persons x km

Fuel consumption

- Definition: "Fuel consumption" of a country is defined as the total consumption of energy by road motor vehicles in the country in Terajoules. Energy can be in the form of gasoline, diesel, LPG, electricity, or some other energy type which is used for the propulsion of road motor vehicles.
- Unit: Terajoules

Population

- Definition: "Population" of a country is defined as the whole number of inhabitants in the country. This number does not include temporary visitors and tourists from other countries, nor illegal immigrants.
- Unit: number of persons

Driver population

- Definition: A "driver" is defined as a person in the possession of a license (possibly a novice license) to drive a road motor vehicle, not necessarily in the possession of a vehicle or having the possibility to use a vehicle.
- Unit: number of persons

Vehicle Fleet

- Definition: "Vehicle fleet" of a country is defined as the total number of road vehicles owned by the country's population, where "road vehicle" is defined as a vehicle running on wheels and intended for use on road.
- Unit: number of vehicles

Number of trips

- Definition: "Number of trips" of a country is defined as the total number of trips made by persons, regardless their age, in the country. A return trip counts as two.
- Unit: number of trips

Time in Traffic

- Definition: "Time in traffic" of a country is defined as the total time spent in travel by persons, regardless their age, in the country.
- Unit: Unit of time (hours, minutes, and seconds)

In theory, if data collection was carried out on the basis of the definitions presented above, data comparability and compatibility with the related accident data (CARE data) would be relatively straightforward. However, in practice, more or less significant deviations from these definitions may be observed, as data collection is carried out at national level, without a common framework. Moreover, it is not clear whether sufficient data quality control is carried out in the International Data Files, which are the main source of international exposure data.

These issues, as well as other related aspects of exposure data availability, collection methods and use are discussed in the present Report.

References

1. European Commission (2003). Glossary for Transport Statistics. Intersecretariat Working Group on Transport Statistics. EC - UNECE - ECMT.
2. Smeed, R.J. (1968). Variations in the pattern of accident rates in different countries and their causes. Traffic Engineering and Control10 p.364–71.

2. Risk exposure data in road safety analysis

This chapter discusses the needs and uses of risk exposure data in road safety analysis. As the basic concepts of road accident statistics play a central role in road safety analysis, this chapter first discusses this topic. After an introduction to the statistical properties of the accident process, the related consequences on the general use of accident rates are discussed. Using that as a starting point, the needs and uses of risk figures are described, focusing on an assessment of the theoretical properties and characteristics of exposure measures.

2.1 Statistical background and limitations

2.1.1. The Poisson Limit Theorem for accident counts

This section discusses statistical properties of accident related outcomes, restricted to accident and victim counts. Statistical properties are best described by the related distributions or densities. This section is devoted to a discussion on the statistical distributions of aggregated accident counts, with some reference to the distribution of victim counts. It should be noted that the term "accident distribution" always denotes the statistical distribution of the number of accidents and not, for instance, the (spatial)_distribution of accidents over an area or the (temporal) distribution over time.

A good starting point for a discussion on the basic concepts of road accident statistics is the work by the French mathematician Simeon Denis Poisson more than 150 years ago (Elvik, 2004, Feller, 1968), who investigated the properties of binomial (Bernoulli) trials. A Bernoulli trial is an experiment that has two possible outcomes: success or failure. This type of experiment seems to be a very useful building block for modelling road safety. For instance, a pedestrian crossing a road could be seen as an experiment with a (fortunately) very small probability of accident occurrence. A similar argument could be used for a vehicle passing through a road section, a vehicle driving past a road side obstacle, or two vehicles encountering each other on the road.

The original work of Poisson assumed the probability of success to be the same at each trial. Poisson could then prove that the distribution of the sum of all successful trials would tend to a Poisson statistical distribution. The above restriction has since been relaxed (Feller, 1968). More specifically, it is not necessary that all probabilities are equal, rather the sum of all probabilities

should tend to a finite λ and it's maximum (Feller, 1968) or sum of squares (Shorack, 2000) should tend to nil.

For the practice of traffic safety analysis this result has the following consequence:

"if a traffic safety outcome can be regarded as the sum of the outcomes of many independent events that had a (very) small probability of resulting in an accident, the distribution of that sum (i.e. the number of events that resulted in an accident) will tend to the Poisson distribution with parameter equal to the sum of the probabilities of resulting in an accident".

The following remarks should be taken into account:

1. this result is relevant to the distribution of the number of accidents, *not* the number of victims or similar outcome of accidents.
2. it is important to understand the role of independence in this theorem. It should be quite reasonable to assume that the *outcomes* of the events are independent.
3. when accident registration problems are to be considered, the concept of "a small probability of resulting in an accident" can be replaced by "a small probability of resulting in an accident *and being registered*".
4. a different but no less important accident registration issue is that usually only accidents exceeding a certain level of severity are considered. This would yield: "a small probability of resulting in an accident *with a certain severity* and being registered". Even if these probabilities are different for each trial, the distribution of resulting number of accidents still tends to a Poisson distribution.
5. an alternative approach to the Poisson distribution of counts, based on counting processes, requires that the registration system cannot be saturated by the accident process (e.g. limited police resource allocation to less severe accidents would have an effect on the theorem above).

The above theorem should be considered a *limit* theorem. For road safety analysis this means that the distribution of accident counts becomes undistinguishable from a Poisson distribution only in the *limiting case*. Thus in practice, the distribution of accident counts will never be exactly like a Poisson distribution. The limiting character of the theorem is due to the number of trials on which it is based. If a count is based on a high number of trials, it is likely that for all practical purposes the count follows a Poisson distribution. For instance, annual national counts of a general type of accidents are practically Poisson distributed. However, care must be taken when the actual number of trials is rather low (Lord, Washington and Ivan, 2005). This can be the case when a rare accident type is studied or road sections with limited traffic volumes are studied.

2.1.2. Over-dispersion

Following Hauer (2001), the following phenomenon is commonly encountered in road safety analysis. "After the unknown model parameters are estimated, one usually finds that the accident counts are "overdispersed", i.e. the differences between the accident counts and model predictions are larger than what would be consistent with the assumption that accident counts are Poisson distributed" (Hauer, 2001). This phenomenon also occurs in settings where one would consider the distribution to be practically identical to the Poisson distribution. The problem rises from the replications used in the generic model as described by Hauer (2001). Even if the accident distribution would be indistinguishable from the Poisson distribution, replications would never be carried out under the exact same conditions. Therefore, both the number of trials and the values of the probability of resulting in an accident will differ between replications. In other words: replications will be drawn from a *different* Poisson distribution each time. A more extensive discussion from the viewpoint of different probabilities can be found in Lord, Washington and Ivan (2005). In Hauer (2001) and the references therein, more information on how overdispersion can be estimated, can be found.

2.1.3. Normal approximations

Very often the distribution of the number of accidents is approximated by a Normal (Gaussian) distribution. The most common procedure is to assume (first approximation) a Poisson distribution first, consequent on that assume the variance of accident counts to be equal to the expected value λ and then approximate (second approximation) the Poisson distribution with a Gaussian distribution with mean parameter and variance parameter equal to λ. Sometimes "overdispersion" is added to the variance parameter. When no statistical model is available, the expected value λ is usually estimated by the observed count. In modelling situations instead, the expected value λ is mostly estimated by the model prediction of the observed count.

It should be noted that the approximation of the Poisson distribution by a Gaussian distribution deteriorates when the accident counts are smaller. There is no general rule as to what value the counts should exceed in order for the approximation to be sufficiently reliable, as that is dependent on the application and the needed accuracy.

2.1.4. The case of victims counts

Given that an accident occurs, the expected number of victims resulting from that accident is a very complex problem. Obviously, it is dependent on the number of persons involved in that accident. Therefore, the distribution of the

number of victims is very complex. When done so at all, approximations can be made based on compound distributions. It can be assumed, however, that the victim counts are overdispersed, more than accident counts. The amount of overdispersion depends on the variation of the number of victims per accident. This means that victim counts from accidents that rarely involve more than one victim, will be less "extra" overdispersed than victim counts from accidents that involve more than one victim more often, compared to the overdispersion of the accidents.

It is important to note that the underlying conditions that determine the distribution of victim counts also influence the distribution of injury accident counts, because that distribution is dependent on the distribution of accidents in general, and the probability of an injury occurring in an accident. For example, if young individuals tend to travel in cars with all their friends (full occupancy) they are, all other things equal, more likely to be involved in an injury accident than individuals driving alone in the very same car, because when something happens (e.g. exactly the same tree is hit at the same place and speed) in the first case it is more likely that somebody is injured (hence an injury accident) than in the second case (no injury, no count). It should be also noted that, in some cases, the occupancy rate can have influence on the number of accidents as well.

2.1.5. The role of trials in exposure

In the context of risk exposure data, the number of trials plays a central role. The number of trials is the number of times road users in general are exposed to a possible accident. Therefore, the number of trials should be the best theoretical measure of exposure.

Obviously if, all other things kept equal, the number of trials increases, the expected number of accidents increases. For instance this would likely be the case where instead of one country, two countries are considered, or if two years instead of one year are considered. In most practical situations, however, the relationship is more complicated. In fact, it is reasonable to assume that, in most cases, a change in the number of trials will change the probabilities of an accident given a trial.

This number of trials is probably the best exposure measure theoretically available. Unfortunately, in practice one has to resort to more practical measures of exposure, which may be more or less close to the theoretical concept of exposure and may present different practical and conceptual advantages and limitations. Some of these issues are discussed the next section.

2.2 Needs and uses of risk figures

2.2.1. The need of risk figures

If one needs to compare the road safety situation between countries (e.g. between road categories or different modes), one somehow has to measure the road safety performance in each country and compare the measurements according to some scale.

The first candidate road safety performance measure will be one of the number of (serious of fatal) accidents or the number of victims, or a combination of such measures (per unit of time). However, although the number of fatalities on the roads is an important and informative road safety performance measure, it may not adequately address all analyses needs. For instance, if the road safety problem is to be compared with other health hazards, it is common to compensate for the number of persons at risk of being killed in a road accident. On that purpose, the annual number of persons killed in road accidents in a certain year divided by the relevant population size is often used. Accordingly, a number of other road safety performance measures were and are still being introduced for specific purposes. The general basic form of a road safety performance measure, commonly called a risk or rate, as well as its various forms and uses, discussed in this section.

2.2.2. General definition of risk

As discussed in Hakkert and Braimaister (2002), there are a number of definitions of risk in use in different forms of safety science, road safety or elsewhere. The approach taken in the present discussion is practical: a risk is the expected road safety outcome, given a certain exposure. The outcome is usually the number of accidents or victims of a certain type, but fundamentally need not be. For instance it could also be monetary loss due to the socio-economic consequences of road accidents.

2.2.3. Definitions and usability of risk

The above, general definition of risk effectively defines risk as a function, mapping exposure onto safety outcomes. Hauer (1995) calls this function a "safety performance function", more precisely mapping exposure per unit of time onto the (number of) accidents per unit of time. Hauer (1995) further argues that there are objections to a linearity conjecture for this function in road safety.

For comparing risks at an international level a risk will be considered that can be regarded as a rate (per time unit). Such rates are more easily compared:

Risk = the expected number of accidents / the amount of exposure.

Thus, as a consequence, per time unit:

The expected number of accidents = Risk * the amount of exposure (1)

It should be noted that, in order to remain compatible with the safety performance function approach by Hauer (1995), the risk can be defined as the original safety performance function divided by the amount of exposure. It is important to note this and its consequence that risk cannot be regarded independent from exposure, if only because of its definition. A similar argument can be used in relation to other influences like time, region, country, or other conditions.

Furthermore, for many applications comparing road safety performance, it is actually assumed that risk differs because of differences in the conditions present during the observations.

Additionally the following comments can be made with respect to the relevancy of these findings in the current context:

- The consequences of not considering a non-linear relationship will be most important when exposure varies significantly within a unit, for instance studying hourly observations on a road section over all hours of a day. When variations are small and relatively stable, for instance when national accident statistics with population figures are considered, the relationship may be well approximated by a linear function within the (small) range of different actual exposure figures. Note that such an approximating function not necessarily crosses the origin, as is assumed in equation (1).
- Because such an approximating function not necessarily crosses the origin, as is assumed in (1), the prominent "mistake" in the use of risk as described by Hauer (1995) is still relevant in this situation: the risk, calculated as the number of accidents divided by a certain exposure A, then multiplied by another exposure B may not yield a very accurate estimate of the number of accidents given exposure B.
- One can only use a function that is known. Effectively exploiting the approach by Hauer (1995) requires that the safety performance function is known for each level of aggregation (e.g. country) considered in a comparison.
- In many cases, actual observation units consist of monthly or longer time aggregations. A typical observation would be the number of accidents and exposure for a certain month or year on a particular road type in some area of one country. If the aggregation is over a heterogeneous set of road

sections, it will be very difficult to asses the safety performance function of the aggregation even if it were available for all contributing sections.

Given the results of Hauer (1995) and the comments above, it is assumed that in many cases the benefits of using safety performance functions is expected to be limited. It is therefore suggested to use risk data in the sense of the number of accidents or victims per amount of exposure, and limit its use to initial comparative analysis. More complex, predictive analysis should be based on more elaborate (non-linear) models, when possible.

2.2.4. Statistical implications

As already mentioned in the beginning of this section, if one needs to compare the road safety level between countries, some measurements of road safety have to be compared. On that purpose, it is important to determine how accurate these measurements (approximately) are.

In particular, the following issues have to be considered

1. Observations are likely to be biased: not all accidents may be counted and / or exposure may be under or over estimated. Moreover, no estimates for these biases may be available. If biases appear to be large and one is unable to correct for them, no reliable comparison can be made.
2. the number of accidents is intrinsically variable: it is impossible, except for the case in which no accident can possibly occur, to predict the exact number of accidents. If one has to assess the potential variation in one observation, a Poisson approximation may be sufficient when the actual count is large enough. However, if two apparently equal areas need to be compared (or even the same area for a different time period), overdispersion issues have to be considered.
3. the exposure figures are likely to be estimates themselves. This means that the variance in their estimates (i.e. the variance of the measurement error in the estimates) needs to be accounted for as well. Standard textbooks offer approximations to the variance of ratios, sometimes by means of linearization (delta method) or by means of simulation (See Annex I).
4. In addition to the variance due to the fact that exposure figures are estimates (measurements), it also has to be considered that the exposure measures are approximations, proxies to the "true" exposure (e.g. one vehicle kilometre may not be the same as another one; the same number of vehicles may be used for more kilometres in a different time period).

The possible biases mentioned above have to be borne in mind and, when possible, corrected or accounted for. Sometimes knowledge of bias may prohibit further analysis. The consequences of unknown accident and exposure variations can only be assessed in the context of a statistical model, however no general model is available. It should be noted that the

presentation of related models is not within the scope of this document. As far as exposure measurement errors are concerned, these should be accounted for in risk estimates, and this may be the most significant limitation in the use of exposure estimates. As discussed in detail in the next Chapter, the different methods for obtaining exposure estimates may account for measurement inaccuracies to a more or less efficient way.

2.2.5. Other uses of risk exposure data

Besides being used as actual exposure measures, risk exposure data can also be used as explanatory information. For instance, differences in the road safety level between countries may be explained by differences in the level of motorisation (the number of vehicles per inhabitant), or the number of driver licences per age group. Moreover, the number of person kilometres compared to the number of vehicle kilometres may reveal differences in occupancy rates, which in turn may influence accident severity.

2.2.6. Discussion

The rate defined by "the number of outcomes" (e.g. accidents, casualties etc.) divided by "the amount of exposure" can be used as an estimate of risk. For instance, in the health risk context, the total number of fatalities in road accidents per inhabitant can be used for certain applications. Accordingly, in the transportation context, the total number of fatalities in road accidents per vehicle-kilometres of travel may be a more appropriate risk rate.

Due to the potentially non-linear relationship between the number of outcomes and exposure (Hauer, 1995), care has to be taken when risks are multiplied by exposure figures to obtain estimates of the number of outcomes under different circumstances. Continuing the above example, if the health risk as the total number of fatalities in road accidents per inhabitant is calculated for a country, using that same figure (multiplied by double the population) to estimate the number of fatalities when the population has doubled, is likely to result in unreliable estimates. It should be considered that, in a case where the population in a country doubled, the health risk probably changed as well.

It is suggested to restrict the use of risk figures, defined as the rate of some road safety outcome divided by some amount of exposure, to compare different risks and acknowledge that risks may be different simply because exposure and other related factors are different.

Moreover, when risks are compared, their potential biases have to be considered and, when possible, corrected. When exposure estimates are considered, their measurement error (i.e. their statistical accuracy, usually

defined by means of the variance) has to be accounted for. At least an estimate of the variance for the road safety outcome is needed. When no such estimate is available, risks cannot be compared as it may not be clear whether differences in risk are coincidental. If the measurement error in the exposure data cannot be ignored, it has to be accounted for as well.

2.3. Risks in use in road safety analysis

In road safety analysis, different risks (rates) may be used according to the objectives of the analysis, as well as the most suitable exposure data available. As already discussed, a risk can be defined as the rate

$$\text{risk} = \frac{\text{road safety outcome}}{\text{amount of exposure}}$$

The road safety outcome is usually the number of accidents or casualties, (fatal accidents, accidents with hospitalised or fatally injured victims, fatalities, persons injured). For special purposes this can be further restricted to a specific period, area, time of day, vehicle types involved, accident type etc.

The amount of exposure is mostly selected based on its theoretical importance. However, quite often the preferred exposure measure is not available, or available in an inadequate level of disaggregation. In such cases, an alternative exposure measure may be selected.

Exposure measures under review in this document can roughly be classified into two groups.

- Traffic estimates: road length, vehicle kilometres, fuel consumption, vehicle fleet.
- Persons at risk estimates: person kilometres, population, number of trips, time in traffic, driver population,

This categorisation is somewhat arbitrary and some measures can well be considered within the other category. For instance, often person kilometres are preferred over vehicle kilometres when fatalities are to be compared, because differences in vehicle occupancy rates may be captured by person kilometres (and not by vehicle kilometres). However, when the subject of a study is the occupancy rate, a comparison based on vehicle kilometres may be more reasonable.

Therefore, there is no general rule as to what is the "best" or "preferred" exposure measure. Accordingly, when population data are applied as a measure of exposure, countries with low levels of motorization achieve a more favourable position, while it is prejudicial to those with high levels of motorization (Holló, 2000). In the following Figure 2.1, the relationship between fatalities per population (epidemiology approach) and fatalities per vehicle fleet (traffic approach) is presented (Trinca et al., 1988), as an example of the different results of different exposure measures.

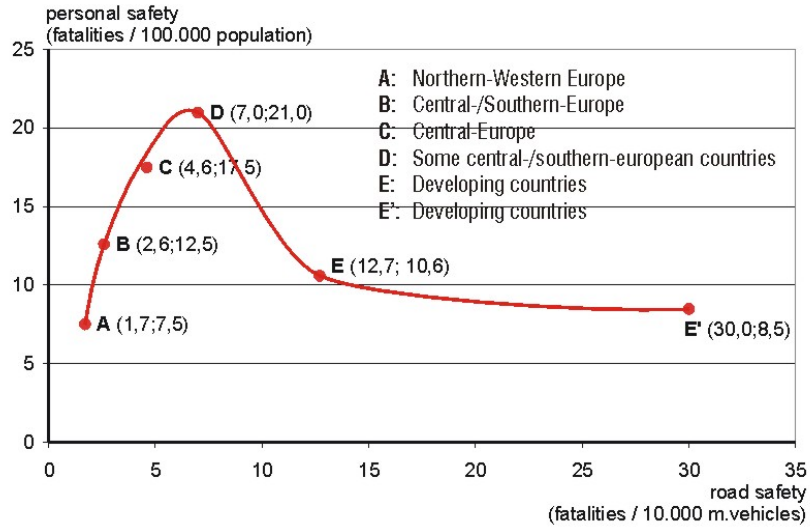


Figure 2.1. Example of comparison of different risks (traffic approach versus epidemiology approach)

The following Figure 2.2 shows one part of the above theoretical curve, based on IRTAD data of 2003 (Hollo, 2004). The model takes the development of motorization into account and consequently the individual countries, according to their motorization level, are "moving" along the theoretical curve (as a result of consistent road safety activities).

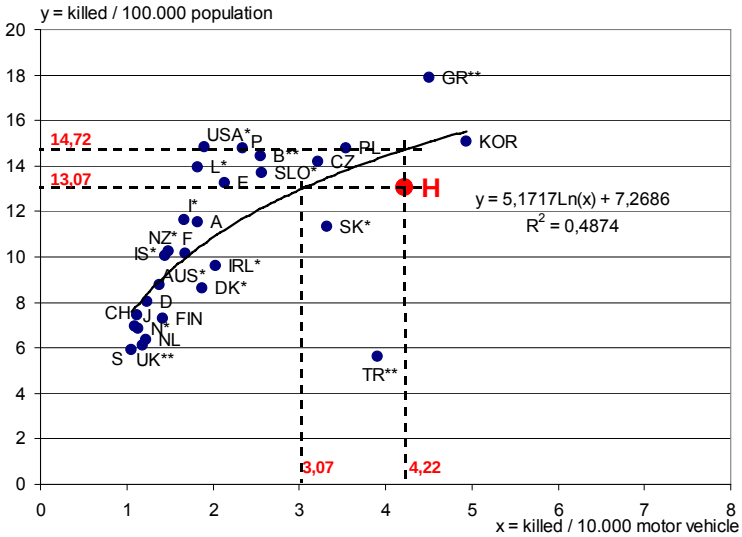


Figure 2.2. Example of comparison of different risks based on IRTAD data for 2003

Therefore, it is obvious that an optimized approach would include the application of different exposure indicators at the same time according to the needs and the context of the analysis. A detailed discussion on the advantages and limitations of the examined exposure measures is presented in the following sections.

2.3.1. Vehicle kilometres

The number of vehicle kilometres is probably the most often preferred exposure measure. One important practical advantage of the use of vehicle kilometres (over road length, fuel consumption, driver population and vehicle fleet) is that, in theory, it may be available to a significant level of disaggregation: time, vehicle type, road type, driver characteristics etc. None of the other exposure measures can usually allow for this level of detail. For that reason it is probably the best measure (alongside *person kilometres*) to capture the regional and temporal variations in the accident process. It is also useful for analyses on specific features of the road safety problem.

When used in combination with road length, it allows for an estimate of traffic density (i.e. number of vehicle kilometres divided by road length).

Vehicle kilometres are most often used in combination with accident counts. However, as it is often tightly related to person kilometres, it can also be used in conjunction with victim counts.

It should be noted however, that, in practice, the availability and the level of disaggregation of vehicle kilometres may vary significantly and is strongly dependent on the type and features of the collection method. For instance, vehicle kilometres obtained by means of traffic counts are usually available per road and vehicle characteristics, while a disaggregation by person characteristics is only possible for data obtained by means of travel surveys. A detailed discussion on the different collection methods of the data used to obtain vehicle kilometres estimates, and the respective effects on the usability of the estimates, are presented in the following Chapter 3 of this report.

2.3.2. Person kilometres

The number of person kilometres is probably the second most often preferred exposure measure. Its application is mostly geared towards casualties counts. However, due to the fact that the person- and vehicle-kilometres estimates are often obtained though the same data source (e.g. travel surveys, traffic counts etc.) person kilometres can be derived from vehicle kilometres and vice versa. Moreover, *driver kilometres,* which are sometimes used to substitute vehicle kilometres, can often be derived from person kilometres.

In general, depending on the way person kilometres are obtained (see Chapter 3 for more details), person kilometres may be available at an even higher level of disaggregation than vehicle kilometres, i.e. including the road user category (driver or passenger) or trip purpose classification. In practice, however, these additional parameters are rarely available in accident statistics.

As for vehicle kilometres estimates, the characteristics of the collection method may significantly affect the final outcome in terms of person kilometres estimates. For example, in case person kilometres are obtained through travel surveys, a substantial sample error, or other sampling bias should be considered (see Chapter 3).

2.3.3. Road length

Road length is a basic exposure measure. As opposed to person and vehicle kilometres, it can not capture temporal variations in the use of roads in an area. Moreover, due to the time needed for planning and development of road infrastructure, the measure may be sensitive to economic influences in a latent manner. On the other hand, road length may be a very useful exposure measure when used for developing countries, or for correcting for the sheer size of countries.

2.3.4. Fuel consumption

Fuel consumption could be regarded as an alternative for vehicle kilometres. It is also sometimes used for the estimation of vehicle kilometres (Fridstrøm, 1999, Cardoso, 2005). One of the drawbacks of this exposure measure, compared to the actual vehicle kilometres of travel, is that short term fluctuations in road use may not be easily captured. Obviously, fuel is consumed some time after sale, which cannot be determined precisely. Accordingly, it is also difficult to determine where fuel is consumed. Therefore, fuel sales are probably best used at an aggregated level, possibly national and annual. However, when comparing countriesadditional parameters should be taken into account, such as fuel efficiency of motor vehicles, pricing differences etc.

2.3.5. Vehicle fleet

Like fuel consumption, the number of vehicles in the vehicle fleet could be regarded as an alternative of vehicle kilometres. However, it is not recommended to be used as a replacement of vehicle kilometres, as it is possible that vehicles on average drive more kilometres over time. Nevertheless, comparing the number of accidents corrected for the number of vehicles is likely to be informative. Furthermore, vehicle information, mainly type, age and physical characteristics that are not likely to be easily available for vehicle kilometres may be available for the vehicle fleet. For private cars, the age of the owner may also be of interest. Influences of foreign vehicle fleets (e.g. from neighbouring countries) may have to be considered. See Chapter 3 for a further discussion on vehicle fleet data.

2.3.6. Population

The relation between population figures and health hazards is often studied, especially in the epidemiological or demographic context. An advantage of the use of population figures over most of the other exposure measures is that in many cases the figures are relatively accurate. Population figures may be available for several variables, most likely age and gender. Figures for specific groups of road users (e.g. schoolchildren or professional drivers) may also be available.

2.3.7. Driver population

Vehicle fleet and driver population are both related to the amount of traffic (vehicle kilometres) in a country. For many purposes, driver population may be considered as an alternative to vehicle fleet information. The measure is likely to share most of the information available for population figures, but may also include an estimate of the drivers experience, and his or her behaviour (e.g. if demerit points are available). However it should be noted that the amount of time a driver holds his license may not be an accurate estimate of his experience. Moreover, it may not be comparable between countries, due to differences in the licensing and registering frameworks.

Driver population figures may be used in a way similar to population figures when drivers casualties are considered.

2.3.8. Number of trips

The number of trips can be regarded as similar to the number of person kilometres. If trip length remains the same, results using the number of trips as exposure and the number of person kilometres should be similar. For the same reasons vehicle fleet figures may still be informative when vehicle kilometres are available, the number of trips may be informative when person kilometres are available. It is most likely that the data on the number of trips is based on the same sources as the number of person kilometres, consequently the same level of disaggregation will be available. It should be noted that the number of trips divided by the number of drivers is additional interesting information in road safety analyses.

2.3.9. Time in traffic

The same comments as those for the number of trips apply for the time spent in traffic, except that time in traffic is likely to follow person kilometres more

closely than the number of trips. The main difference is that time in traffic may to some degree account for the development (differences in) the average travel speed. Moreover, the background idea may be different: only while being involved in traffic – moving or halted – one is exposed to being involved in an accident.

However, difficulties may be encountered in the disaggregation of time spent in traffic, especially as regards comparisons. For example, comparing the time spent in traffic between motorways and urban areas, or between riding a bicycle, sitting in a bus and driving a car, may be complicated.

2.3.10. Other methods for risk estimates

In order to estimate accident risk when no exposure data is available, alternative methods are applied, namely the quasi-induced method for measuring exposure.

The quasi-induced exposure technique is based on the assumption that in every road accident, in which two vehicles are involved, there is one driver responsible for the accident and one innocent driver involved randomly from the total population of drivers. Consequently, the innocent driver can be considered as a sample of the total population of the drivers and reflects the exposure of any specific driver population defined on the basis of certain characteristics (Haight, 1973, Lyles et al. 1991). The basic requirement for the use of this method is the identification of the driver who provoked the accident. It should be noted that a series of tests should be conducted beforehand, in order to eliminate possible bias that might limit the application of the technique.

The use of the quasi-induced exposure technique may not completely overcome the need for traffic data. But the most important advantage of the quasi-induced exposure technique is the fact that it allows for detailed analysis at the level of disaggregation of the existing accident data (Golias, Yannis, 2001). However, the use of the quasi-induced exposure technique is limited by the fact that it concerns only drivers and not all road users (passengers and pedestrians) and that it requires the knowledge of the "at-fault" and the "innocent" drivers. Additionally, this method concerns mainly accidents in which at least two vehicles were involved whereas its use in single-vehicle accidents is not recommended.

Another very common approach is based on analyses that use experimental groups and control groups. It is assumed that, for the issues under study, a control group should "behave" the same way the experimental group does. This effectively means that it is assumed that exposure in the control group is proportional to the exposure in the experimental group.

2.4. Synthesis

In road safety analysis, exposure data can be used in two manners:

1. To obtain risk data in the form of outcome per unit of exposure.
2. 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.

Due to the potentially non-linear relationship between the number of outcomes (i.e. road accidents) and exposure (Hauer, 1995), care has to be taken when risk estimates are multiplied by exposure figures to obtain estimates of the number of outcomes under different circumstances. It should be underlined that risk and exposure are not independent and that, when exposure changes, other factors determining the risk are likely to change too. Therefore, it is suggested to focus (if not restrict) the use of risk figures to compare different risks and acknowledge that risks may be different because exposure and other related factors may be different.

When risks are compared, the potential biases in both accident outcome and exposure measure should be considered and, when possible, corrected for. When estimates are considered sufficiently useful, their measurement error has to be accounted for.

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. These alternative exposure measures may also have other, explanatory or descriptive uses.

Table 2.1. Advantages, limitations and optimal use of different exposure measures

Road safety Outcome	Accidents / Persons	Persons	Accidents	Accidents / Persons	Accidents	Persons	Persons	Accidents / Persons	Persons
Amount of exposure	Vehicle - kilometres	Person - kilometres	Road Length	Fuel consumption	Vehicle Fleet	Population	Driver population	Number of trips	Time in traffic
Context of analysis	Traffic	Traffic - Mobility	Traffic - Infrastructure	Traffic	Traffic	Epidemiology	Traffic	Traffic - Mobility	Traffic
Temporal variation of risk	•	•		•				•	•
Regional variation of risk	•	•	•			•		•	•
Disaggregation level of risk									
Road User category		•				•		•	•
User characteristics	•	•				•	•	•	•
Vehicle characteristics	•	•		•	•			•	•
Road network characteristics	•	•	•					•	•
Optimal use			Developping countries	Aggregate level	When average traffic is the same	Comparing health hazards	When average traffic is the same	When average trip length and travel speed are the same	
Possible bias	sampling	sampling	economic influences	pricing differences, vehicle efficiency			licensing framework	sampling	sampling

Table 2.1 above summarizes the discussion presented in the present Chapter as far as the advantages, limitations and optimal use of different exposure measures. 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 following Chapter 3 of this report.

References

1. Elvik, R. (2004). "Traffic Safety" pp 16.3-16.20 In: Kutz, Myer (ed): Handbook of Transportation Engineering. McGraw-Hill, New York.
2. Feller, W. (1968). An Introduction to Probability Theory and Its Applications, Vol I. Wiley, New York.
3. Golias, J., Yannis, G, 2001. Dealing with lack of exposure data in road accident analysis. Proceedings of the 12th international conference "Traffic Safety on three Continents", Federation of European Road Safety Institutes, Transportation Research Board, Moscow, September 2001.
4. Haight, F., 1973. Induced exposure. Accident Analysis and Prevention 5.
5. 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.
6. Hauer, E. (1995) On exposure and accident rate. Traffic Engineering + Control, 36(3), pp. 134-138.
7. Hauer, E. (2001) Overdispersion in modelling accidents on road sections and in Ampirical Bayes estimation. Accid. Anal. Prev. 33. p. 799-808.
8. Holló, P. (1997). The Role of Motorisation Level in International Comparisons of Road Safety. The Fourth International Conference on Safety and the Environment in the 21st Century, November 23-27, 1997, Tel- Aviv, Israel, Proceedings, pp. 144-154.
9. Holló, P. (2000). Relationship between motorisation and fatality rates based on IRTAD data. International Seminar on Road Traffic and Accident Data Needs for the New Century, Vienna, September 2000. Proceedings.
10. Holló, P. (2004). Road Safety in Hungary. PIARC Technical Committee 3.1. Road Safety, 2nd meeting, Budapest, 27th-29 th October 2004
11. Lord, Washington and Ivan (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accid. Anal. Prev. 37. p. 35-46.
12. Lyles, R.W., Stamatiadis, P., Lighthizer, D.R., 1991. Quasi-Induced Exposure Revisited. Accident Analysis And Prevention 23 (4).
13. Rice, J. A. (1995). Mathematical statistics and data analysis. Duxberry Press, Belmont, CA.
14. Shorack, G. R. (2000). Probability for Statisticians. Springer , New York.
15. Trinca, G,. Johnston, I., Campbell, B., Haight, F., Knight, P., Mackay, M., McLean, J. & Petrucelli, E. (1988). Reducing traffic injury – a global challenge. Royal Australasian College of Surgeons, Sydney.

3. Review of methods for collecting RED in the EU

3.1. Introduction

In the previous Chapter, the uses of risk exposure data in road safety analysis were discussed, focusing on the theoretical advantages and limitations of the different exposure measures and their optimal use according to the purpose of the analysis. It was also underlined that, some of the main theoretical features of the various exposure measures may be affected in practice by the characteristics and particularities of the data collection method.

This Chapter concerns a presentation of the various methods used to collect "raw" risk exposure data (RED) over the EU Member States. The term "raw" data is used to denote data that are used to derive exposure estimates. The approach to focus on "raw" data implies a certain distance from the initial subject of this report, as the actual relation to traffic safety is covered in less detail in the present Chapter. More specifically, emphasis is put on the core issues of the methodology. The data collection methods examined are:

- Travel Surveys
- Traffic counts
- Vehicle fleet registers
- Driving licenses registers
- Road registers
- Other methods

It should be noted that there is no standard method for the collection of each exposure measure (FHWA, 1997). In particular, different exposure measures may be derived from one collection method (for instance a travel survey may be used to collect vehicle kilometres, but may at the same time be used to obtain the number of trips, the time spent in traffic, vehicle ownership, driver license holdership etc.). Accordingly, data collected by different methods may be used to produce an exposure estimate (for instance, passenger kilometres estimates may be obtained by using vehicle kilometres derived by traffic counts and vehicle occupancy rates obtained through surveys). Therefore, one of the objectives of this Chapter is to highlight the different exposure estimates that can be derived by each method or combination of methods.

However, the main objective of this Chapter is the presentation of the characteristics (advantages and limitations) of each collection method. The review of methods includes the general description of each method, a review of practical implementation issues and a presentation of the disaggregation level (variables and values examined) available. On that purpose,

representative examples from seven EU Member States are used: Norway, Denmark (Hemdorff, 2005), France (Haddak et al. 2005), Greece (Yannis, Papadimitriou, 2005), Hungary (Cseffalvay, Holló, 2005), the Netherlands (Bos et al., 2005) and Portugal (Cardoso, 2005). As mentioned in Chapter 1, the main source of examples was a set of National Reports on data collection methods, availability and use in the respective countries.

Moreover, this chapter focuses on either random or systematic errors in the "raw" data, which may affect further use of the derived exposure data in road safety analysis. Consequently, methods without relevant errors in their data collection mechanism are not discussed in detail (e.g. population data, for which very accurate digital registrations are usually available). Additionally, some methods are well established in other fields than road safety. For instance, obtaining road (network) length information may be well covered in engineering literature, while the resulting level of accuracy may be by far sufficient for practical use in road safety analysis. Similarly, fuel sales (consumption) data may be well covered in economic literature. On the other hand, survey methodology may be well covered in statistical literature, however many aspects are worth discussing in relation to road safety analysis.

Summarizing, methods for collecting exposure data, which have an intrinsic statistical error or other bias that may affect risk calculations, are mainly discussed in this chapter. Within this framework, the examples from different countries are used to identify common practice, optimal implementation and deficiencies of each method.

3.2. Travel surveys

3.2.1. The need for surveys

In theory, in order to determine the travel patterns of a population, one would have to monitor the related activities of each member of this population at a constant rate over the examined period. Apart from the obvious difficulties in carrying out such measurements (e.g. cost and time restrictions), it has been well established in the literature that it is not necessary to obtain this level of exhaustiveness and accuracy for practical purposes. More specifically, the alternative is a survey on part of the examined population. Depending on the purpose of the survey and the resources available, the sample type and size, the frequency and duration, and the amount of information collected may differ.

3.2.2. The sample framework

Every survey is intended to obtain information on some target population. In travel surveys, this population usually consists of either actual persons or households, vehicle owners etc. In all cases, the sample chosen is intended to be representative of the entire target population. It should be noted, however, that the sampling process may be more or less successful, in terms of representativeness, due to random or systematic errors, or survey design inadequacies. For example, in case a travel survey is made part of a national population census (e.g. the Portuguese national travel survey), the resulting sample is representative for the national population, as the target population of the travel survey is also the national population.

There are essentially two ways (potential) respondents are selected for travel surveys in the EU Member States:

- Respondents are drawn from a database (e.g. telephone directory or other demographic database) and are contacted by telephone or mail.
- Respondents are randomly selected and contacted, on the roadside or by telephone, and then it is verified whether each respondent is a member of the target population (e.g. by verifying the respondents age, license holdership etc.).

Most travel surveys in the EU use telephones for contacting respondents, even if the questionnaire is on paper or in electronic form (e.g. computer assisted telephone interview). Moreover, most surveys nowadays feature some sort of a "call-back" system, in order to retry respondents that could not be contacted. E-mail and internet surveys also start to appear, although these methods are not currently used to obtain exposure data (FHWA, 1994).

3.2.3. Survey errors

As mentioned above, a survey is a method used to obtain information from the population by means of asking a (random) selection of individuals. Different sorts of errors may result from this sampling method, which are usually referred to as sampling errors, non-response errors and measurement errors. These three main kinds of errors can be detailed the following way (Cochran, 1963):

A basic property of a survey is that a selection of individuals is asked for information, instead of the entire population. This means that, if another selection of individuals is asked for the same information, this information is likely to be different to some extent from the information from the original selection of individuals. For example, in terms of vehicle kilometres, the average kilometres traveled may vary with the selection of individuals asked (interviewed). This type of error is called *sampling error*.

Moreover, not all individuals asked actually respond to a survey. This phenomenon, called non-response, cannot be ignored, although it is difficult to correct for its consequences. Non response occurs because people may not accept to be interviewed. However, there also may be other causes. For instance, people that travel a lot are less likely to be interviewed at home by means of a telephone survey, whereas they are more likely to be interviewed by a roadside survey. Consequently, for example, average kilometres traveled reported by people interviewed at home may differ from average kilometres traveled reported by people interviewed at roadside surveys. This type of error is called *non-response error*.

Another important property of this method is that individuals (as subjects) are asked for information, whereas in most other exposure data collection methods, more or less objective measurements are made on subjects. This implies that the individuals that responded the survey give the answers they judge as the correct ones. This, however, may not always be successful (e.g. under- or overestimation of time / distance traveled). This type of error is called *measurement error*.

Based on the elaboration above one can distinguish the following types of error in surveys:

- Sampling error: The error in the data caused by the fact that only a sample of the examined population is interviewed
- Non-response error: The error caused by the fact that some individuals that could or should have been interviewed are not interviewed.
- Measurement error: The error caused by the fact that some individuals interviewed give wrong or inaccurate answers.

A more detailed discussion on all three types of survey errors is presented in Annex III.

3.2.4. Exposure measures collected and disaggregation level

Usually travel surveys are designed to capture all travel by a respondent on a prescribed day, mostly the day before the contact with the respondent (e.g. Denmark, Norway Tøi), however, in some cases information on a specific day is asked in advance and in other cases additional information on other / previous days is asked (e.g. Greece NTUA).

The data reported by travel survey respondents may concern one or more of the following exposure measures:

- Distance traveled
- Time spent in traffic
- Number of trips

Additional information may include the number of vehicles in use, the number of licensed drivers, the number of active drivers, vehicle occupancy etc.

The above information is usually available by mode of travel. It should be underlined though that the main advantage of travel surveys (compared to other methods for collecting the above exposure data) is that they have persons as units, allowing for the exposure data to be distributed by person characteristics such as age, gender, driving experience, nationality etc., and for comparisons between groups of persons to be carried out.

The conceptual framework of travel surveys allows for a very high level of disaggregation. More specifically, travel survey results may allow for combined comparisons per person, vehicle and network characteristics. However, the availability of variables and values recorded varies significantly among countries. In Table 3.1 below, an overview of the features of the travel surveys of seven EU Member States is presented. It should be noted that the Table is not exhaustive, as some information was not included in the National Reports and therefore empty cells may not necessarily reflect not available data. However, main characteristics and particularities of each survey, in terms of design features, measurement and disaggregation are highlighted. Moreover, the available sample of countries (examples) is clearly representative of EU common practice.

More detailed information from the examined countries can be found in Annex II.

Table 3.1. Overview of travel surveys in the EU Member States

	Norway	Greece	Portugal*	Netherlands	France	Denmark
Survey type						
Level	official	not official	official	official	official	official
	national	national	national	national	national	national
Organization	TOI	NTUA		CBS	INSEE	TU
Duration	1 year	3 months			1 year	
Frequency	4 years	-	10 years	1 year	7-10 years	1 year
Since	1980	-	1900	1978	1966	
Age restrictions	>12	>16		>11 **		18-80
Type of travel	private	all	commuters	all	all	all
Unit	distance	distance	time	distance, time	distance, time	distance
Modes						
Car	•	•	•	•	•	•
Twowheelers	•	•	•	•	•	•
Public transport	•		•	•	•	•
Pedestrians	•		•	•	•	•
Bicyclists	•			•	•	
Variables						
Age	•	•		•	•	•
Gender	•	•		•	•	•
Experience		•			•	
Road type		•			•	

** part of the national population census*

*** until 1994*

3.2.5. Limitations

It is obvious that several limitations in the full exploitation of the data in road safety analyses may be encountered. In some countries not all modes or age groups are examined. Experiences with travel surveys indicate that particular short travels (by foot and by bicycle) are often underreported, whereas motorized trips are often overestimated, both in terms of time and distance. The number of variables recorded varies significantly among countries, and the respective differences in the definitions may complicate international comparisons even more. Other limitations include the fact that no information on the variations over time can be made available by means of travel surveys.

Perhaps the most important limitation of the exploitation of travel surveys is that it is not clear how the "raw" data collected is translated to exposure measures, such as vehicle kilometres or person kilometres. Travel surveys normally have other purposes than to give exposure data and are usually mobility-oriented, rather than traffic-oriented.

In most cases, (national) exposure estimates are made using additional data from other sources, together with the travel survey data. It is obvious that the differences among countries in travel survey design and implementation, the definitions of variables and the formulae used for the calculation of the exposure measures may render the comparison among these estimates very misleading.

3.3. Traffic counts systems

3.3.1. General description

In most European countries traffic counts systems are in place. Those can be divided into "human" and machine versions. Counting procedures based on human observations have some advantages due to the fact that humans are able to intelligently categorize vehicles and conditions, whereas machine versions only have limited sensors available for that purpose. However, human involvement is expensive and only functions properly at a limited traffic intensity (for instance at road toll stations). The machine counterparts have the advantage of not being sensitive to fatigue and to some extent being able to cope more easily with higher traffic intensities (except of saturation conditions).

3.3.2. Sampling framework and continuity over time

In the ultimate case in which all road sections are equipped with a counting system, the vehicle kilometres driven during a given unit of time can be computed as the product of the number of vehicles counted during this unit of time and the section length. Accordingly, by summing the products for all road sections, the total number of vehicle kilometres during this unit of time is obtained. In order to be exhaustive, this computation would require all road sections to be equipped with a counting system, and for all successive periods of time.

In practice, for obvious reasons, counts are made for only a limited number of road sections or sites, which are usually located on the main interurban network. The computation described above may then be achieved by summing the products for the equipped road sites; however the result may be biased. One way of dealing with this limitation would be to consider “virtual counting points” for every road section and attributing traffic by means of models, as done in the Netherlands, for instance. Another way to handle this issue would be to provide weights for each counting spot and then calculate the weighted sum, on all road sites on which counts are made, as an estimate of the total vehicle kilometres driven on the aggregate network or country. This method is used in Norway and in France.

Although the actual traffic in some cases may not be estimated accurately enough in the method described above, relative comparisons between periods or regions defined this way could be carried out at an acceptable level of accuracy.

Another relevant aspect is that the counting process may not be continuous over time. This situation is introduced in the Hungarian "system of national

road traffic census". In particular, a choice is made to "rotate" the vehicle measurement system, allowing for a more comprehensive coverage of road sections (practically the whole national road network) in the "road traffic census", at the expense of not having continuous coverage of traffic in a much smaller sample of locations/road sections.

It should be noted that this procedure of limiting the number of road counting sites to permanent counting sites introduces an additional sampling error to the data, which is similar to the sampling error obtained with travel surveys. However, this case is different in the sense that the "respondents" (i.e. counting sites in this case), are not randomly selected.

3.3.3. Exposure measures collected and disaggregation level

Traffic count systems are designed to capture all traffic on a road segment or cross-section, usually by some vehicle classification and seasonal variation included. The traffic counts (number of vehicles) may be exploited for the calculation of the following measures:

- Traffic volume and AADT (Average Annual Daily Traffic)
- Origin - Destination distributions
- Vehicle kilometres of travel

As mentioned above, this information is usually available by vehicle type (to the extent that this can be captured by sensors) and road type. In Table 3.2 below, an overview of the features of traffic counts systems in the EU is presented, based on representative examples from seven countries. It should be noted that the Table is not exhaustive; more detailed information can be found in the related documentation. However, the main characteristics of each system, in terms of measurement and disaggregation features are summarized. Moreover, the available sample of countries (examples) is clearly representative of EU common practice.

More detailed information from the examined countries can be found in Annex IV.

Table 3.2. Overview of traffic counts systems in the EU.

	Norway	Greece*	Portugal	Netherlands	Hungary	France	Denmark
Coverage	National	National	National	National	National	National	National
Number of permanent stations	230	8	61			250	
Total number of stations		902	712			1500	
Continuity	continuous	continuous	every 5 years		rotating	continuous	continuous
Estimates	AADT	AADT, O/D	AADT, traffic volume	AADT	AADT, traffic volume	AADT	traffic volume
Time scale							
hourly variation	•	•	•		•	•	
seasonal variation	•	•	•		•	•	
Variables							
vehicle type		•	•		•		
road type							•
region	•				•	•	•
direction					•		
intersection		•			•		

** until 1993*

3.3.4. Limitations

It should be underlined that the main advantage of traffic counts systems (compared to travel surveys) is the potential of continuity of measurements over time.

However, as mentioned above, traffic counts systems only count vehicles, not vehicle kilometres and there are practical problems involved in the calculation of vehicle kilometres from traffic volumes or AADT. For example, when distribution by road type is sought, a combination of traffic counts and traffic assignment modelling is required, as it is not feasible to evaluate traffic volumes in all network segments, especially in urban areas. Moreover, as mentioned above, measurement points may or may not be representative of the national / regional traffic, as local or urban roads are usually not included. Finally, problems may also be encountered in vehicle classification.

One of the most significant limitations of exposure data derived from traffic counts is that these have vehicles as a unit, not allowing for comparisons among groups of persons with different characteristics. Moreover, it would be extremely difficult to compare vehicle kilometres estimates derived from traffic counts systems and travel surveys.

3.4. Vehicle fleet databases

As mentioned in the previous Chapter 2, the number of vehicles in use or the number of registered vehicles is a useful measure, which can be used to compare or correcting of the level of motorization between countries. In particular, the number of vehicles in use may be used as an exposure measure, i.e. as an alternative of vehicle kilometres traveled under certain conditions (e.g. average distance traveled is equal at the examined level of disaggregation). In general, however, vehicle fleet data usually is used as an exposure measure when vehicle kilometres are not available. Two approaches are available for exploiting vehicle fleet information; firstly, by collecting vehicle ownership data through surveys, which are described in a previous section and are not further discussed in this section.

Secondly, by means of the various vehicle-related databases, which are maintained in most countries:

- Vehicle registration databases
- Vehicle taxation databases
- License plate registers
- Vehicle inspection data

In addition to the formal administration related databases, some private databases may also be available for research (e.g. in Portugal a database of insured vehicles is exploited for research).

Best practices depend on the legal framework per country on the following issues:

- Do all vehicles have a registration number? This is particularly of interest for mopeds, bicycles and / or light motorcycles. Is the registration number unique?
- Do all vehicles have license plate identification? This is particularly of interest for mopeds, bicycles and / or light motorcycles. Is the identification unique?
- Is taxation imposed on vehicle ownership?
- Is road use taxation imposed on vehicle ownership?
- Is insurance mandatory for each vehicle?
- Are technical inspections mandatory for each vehicle?

The exposure measure most widely used is the number of registered vehicles, as relevant databases exist in most countries and the data can be easily accessible. However, in some countries, not all vehicles may be registered and attributed a license plate number in the same database, and the exception usually concerns moped and / or light motorcycles (e.g. Norway, Greece, Hungary, Denmark). Moreover, in some countries the plate number may be related to the driver license rather than attached to the vehicle

registration. It should be also noted that the lack of a specific framework for replacing lost license plates may also affect the accuracy of the data.

Based on the legal and administrative differences among countries, it is difficult to prescribe in general how a vehicle database should be set up. However, it is obvious that such databases are useful only if their maintenance is guaranteed. Effectively, a database on license plate identification is only useful when in practice every vehicle has unique license plate identification. Another problem may be the timeliness of the registration: new vehicles should be included shortly after their introduction on the roads, but more importantly, vehicles no more in use should be removed from the registrations or indicated as no more in use. It should be noted that in most countries, vehicles officially retired are efficiently removed from the registers, at least during the recent years. On the contrary, other types of vehicles no longer in use (e.g. scrapped or abandoned vehicles) are practically not removed from the registers. Within this framework, it is likely that databases where at least on a regular bases, some form of cost is involved in (prolonging) the registration are most reliable.

It is quite likely that data from vehicle taxation databases would be more reliable and representative of the actual number of vehicles in use. However, in most countries these databases are independent (not linked to the vehicle registration databases), and in most cases not accessible due to confidentiality or other reasons. Accordingly, vehicle inspection information would be more representative of the actual number of vehicles in use; these databases are also usually maintained by the same authority that maintains the vehicle registration data. However, in most countries, vehicle inspection data are not collected at a regular basis and / or not maintained in organized databases.

The above alternatives (vehicle inspection and taxation data), when available, could be used either as main sources of the number of vehicles in use, or as a means to correct for the particular inaccuracies of the vehicle registers discussed above. It would be also possible to account for these potential biases by comparing the number of vehicles in the registers with survey data, when available; however there may be practical and statistical difficulties in carrying out such comparisons.

As far as the variables recorded in the databases are concerned, these include in most cases vehicle type, brand, weight and registration date. However, the number of vehicle types, as well as the related definitions used, may vary significantly among countries. In the following Table 3.3 an indicative overview of the characteristics and information available in the official vehicle registers is presented for seven European countries.

Table 3.3. Overview of vehicle registration databases in the EU

	Norway	Greece	Portugal	Netherlands	France	Hungary	Denmark
Official database	Public Roads Administration	Ministry of Transport	Ministry of Internal Affairs	Central Bureau of Statistics		Ministry of Internal Affairs	Police
Type of vehicles							
mopeds	•			•	•		
motorcycles	•	•	not <125 cc	•	•	•	•
cars	•	•	•	•	•	•	•
trucks	•	•	•	•	•	•	•
buses/coaches	•	•		•	•	•	•
Variables							
Brand	•	•	•		•		
engine size	•	•	•				
vehicle age	•	•			•		•
weight	•		•				•
new entries	•	•	•	•	•	•	•
retired vehicles		since 1994			•		
scrapped vehicles	•						
Publication	new entries total vehicles	new entries total vehicles		new entries total vehicles		new entries total vehicles	
Other sources			Taxation data Inspections data private data				

It is obvious that the main limitations in the exploitation of such registers, apart from the fact that the number of registered vehicles is by nature a crude estimate of exposure, concern poor maintenance and update of the registers (i.e. removing vehicles not in use). Additionally, it is obvious that no disaggregation per driver or person (other than potentially the owner, which is not necessarily a natural person) characteristics can be obtained by the vehicle fleet data. However, the fact that these registers contain objective information (i.e. are not sample-based) is an important advantage.

3.5. Driver licenses databases

As the number of vehicles in use may be used as an alternative of vehicle kilometres under certain conditions, the number of drivers can be used as an alternative of person kilometres under certain conditions (e.g. average distance traveled by drivers is equal at the examined level of disaggregation). Although in practice these conditions are often violated, the number of drivers is a more appropriate exposure measure compared to the entire population, especially in the context of road safety analysis. Moreover, it can be made available more easily compared to person kilometres, as in most countries driving license registers are maintained and exploited on that purpose.

Two approaches are available for collecting information on the number of drivers; firstly, by collecting driving license holdership or active driver data through surveys, which are described in a previous section and are not further discussed in this section. Secondly, by exploiting the national driving license registers.

The number of driving licenses, as recorded by the countries, apart from the obvious limitation of not including traffic information, may also involve other difficulties. In particular, the number of driving licenses may or may not reflect the actual number of drivers, for several reasons; first of all, data may concern the number of licenses, not the number of licensees. In this case, a license of a given category may also implicitly allow for a license in a "lower" category. For example, a moped license may not be required for a driver who has a "higher" class license (e.g. passenger car license in the Netherlands, motorcycle license in Greece). Additionally, in some countries a lorry license implies a car driver license. In general, some hierarchy may exist among driver licenses and only the "highest level" of the driver license is recorded (much like the highest level of education is recorded instead of each intervening step).

Another source of data inaccuracy may rise from insufficient maintenance of the databases, in terms of update process. In particular, while newly issued driving licenses are sufficiently recorded, withdrawn licenses are less accurately recorded in most cases. Moreover, in most countries it is not known whether and how deceased drivers are removed from the files. For instance, in Greece, drivers in the database appearing to exceed a certain age threshold are manually removed. However, such practices are hardly efficient, as additional bias may be introduced.

Finally, it is obvious that information on the number of inactive drivers (e.g. drivers who hold a license but do not practice driving, at least on a regular basis) can not be obtained by means of license registers.

Similar to vehicle databases, driver license databases can be corrected and improved when compared to survey data. In some counties (e.g the

Netherlands), driver license holdership information is also part of the travel survey. This provides the added value of the opportunity to combine license holdership information with travel information. Survey data may also provide information on licence holdership of foreigners and exclude persons emigrated or temporarily living abroad. In fact, the information obtained by means of such surveys provides information, which is more related to the number of active drivers and the use of the licenses.

However, due to the fact that data are available and easily accessible, driving license registers are the most widely used source of data on drivers and driving licenses. One additional advantage is that the derived exposure data can be classified per driver characteristics such as age, gender and driving experience. In the following Table 3.4, an overview of national driving license registers is presented on the basis of examples from seven European countries.

Table 3.4. Overview of driving license databases in the EU

	Norway	Greece	Portugal	Netherlands	France	Hungary	Denmark
Official database	Public Roads Administration	Ministry of Transport	Ministry of Internal Affairs	National Travel Survey	Ministry of Internal Affairs	Ministry of Internal Affairs	Police
New drivers	•				•	•	
New licenses	•	•	•	•	•		•
Deceased drivers							
Variables							
license category	•	•	•	•	•	•	•
driver age	•	•	•	•	•		
driver gender	•	•	•	•	•		
region	•	•	•	•	•	•	•
Other information			training, driving tests		Offenses	Offenses	

It is noted that the variables available may include license category, driver age and gender, as well as regional and time distribution. It is also interesting to notice that in some countries, in addition to driving license information, other information may be available, mainly as far as traffic offenses are concerned. More specifically, in several countries traffic violations and / or demerit system records are maintained. However, care must be taken using data from such records, as they may or may not be linked to the driving license records.

Summarizing, as vehicle fleet data are used to produce (national) exposure estimates per vehicle characteristics, driving licenses are used to produce (national) exposure estimates per driver characteristics. However, these exposure estimates concern only drivers and not all road users and may present some inaccuracies or bias complicating the international comparison.

3.6. Road registers

As mentioned in the introduction of the present Chapter, information on road length is available in most countries through some roads register, maintained by the national administration. As the methods for the estimation of road length are well covered in engineering literature and are considered to be sufficiently accurate for all practical purposes, these methods are not discussed in this Chapter. An overview of the various national sources of road length is presented in Table 3.5 below.

Table 3.5. Overview of road registers in EU countries

	Norway	Greece	Portugal	Netherlands	France	Hungary	Denmark
Official database	Public Roads Administration	Ministry of Environment, Physical Planning, Public Works	Highways agency	National Road Database	Ministry of Transport	National Road Databank	National Road Database
Digital or GIS mapping	•			•		•	•
National roads	•	•	•	•	•	•	•
Rural roads	•	•	•	•	•	•	•
Local roads				•	•		
Intersections	•	•		•		•	
Disaggregation Level	link /node	county		Link/node		Link	
Variables							
length	•	•	•	•	•	•	•
axle load restriction	•						
speed limit	•						
road type	•	•	•	•		•	•
pavement type						•	
Other sources	Local roads available at local authorities	Local roads partly available at local authorities	Local roads available at local authorities		Local roads available at local authorities		Local roads available at local authorities

It is interesting to note that the examples of data availability by country presented in Table 3.5 range from general aggregate information (e.g. total road length by county and road type for rural roads in Greece) to very detailed and accurate information (e.g. digital mapping of all national and regional roads in Hungary and the Netherlands).

One common feature of most registers concerns the unavailability of road length data on municipal and / or local and / or private roads. In some cases, this data is maintained (to some extent) by local authorities and, in some cases, the related figures are forwarded to the national registers. Another positive feature is that the information concerning the length of motorways, out of the total road length is available in most cases.

3.7. Other methods

3.7.1. Model estimation of vehicle kilometres based on fuel sales

As reported earlier, road traffic volume estimation is a complex task, involving the execution of expensive travel surveys and traffic counts, on a regular and systematic way. Traffic volume estimations in some countries are made with a combination of direct techniques (travel surveys and traffic counts) and indirect methods. Some of the later require the use of data on fuel sales and estimations of fuel efficiency of the vehicle fleet (when available).

In Germany, vehicle kilometres originated by gasoline consumption and vehicle kilometres originated by diesel consumption are estimated separately, using a common procedure. Periodically, the estimation procedure is re-calibrated by traffic census across the whole country (COST 329, 2004).

Five gasoline vehicle classes are considered: mopeds; motorcycles; buses; trucks; and cars. Periodic surveys are carried out to estimate the mean yearly distance travelled (km/vehicle) and the mean fuel consumption per kilometre (l/km) for the first four classes. The calculated values are combined with the vehicle fleet figures, to estimate the total amount of fuel consumed by these four vehicle classes. The fuel consumption of cars is then calculated. This figure is combined with estimates of mean consumption per kilometre and the number of cars to calculate the total distance travelled by cars (car kilometres).

A similar procedure is used to estimate vehicle-kilometres travelled by diesel vehicles. Six diesel vehicle classes are used: cars; buses; semi-trailers; agricultural tractors; construction and special vehicles; and trucks. Periodic surveys are carried out to estimate the mean yearly distance travelled and the mean fuel consumption per kilometre for the first five classes. The calculated values and the vehicle fleet figures per class are used to estimate the corresponding total diesel consumption by vehicle class. Subtracting the diesel consumed by these five classes and by non-transport related activities from the total diesel consumed gives the total diesel consumption by trucks. The total number of truck-kilometres is calculated using the total number of trucks, the total diesel truck consumption and an estimate of the truck mean consumption per kilometre.

In France a similarly comprehensive procedure is used for estimating the total number of vehicle kilometres. Data on fuel consumption, car-park, unitary fuel consumption and mobility surveys are used in applying the model for estimating purposes. A similar procedure is also used in Norway.

Several alternative simple assumptions may be adopted to estimate the number of kilometres travelled, when no direct method is available. One often

used assumption is that the travelled distance is directly related to fuel sales and may be represented by the amount of fuel sales (expressed as tons of fuel), a figure that is assumed to represent fuel consumption. To be dimensionally consistent, the resulting composite safety variables are expressed as ratios of occurrences (accidents or victims) per ton of fuel. However, the mentioned assumption is not entirely satisfactory, since it ignores the effect of the increasing trend in energetic efficiency of cars and trucks. This weakness may be a major handicap in time series analyses because it is well known that current vehicles are more efficient than vehicles produced ten or twenty years ago.

Within the scope of the COST 329 Action (Models for traffic and safety development and interventions) a method was developed for estimating the traffic volume of countries where such data is not available, using data on vehicle fleet and fuel consumption in the studied country and mathematical models fitted to existing data (from other countries) on fuel consumption, vehicle fleet and traffic volume. This seemed especially useful for some Eastern and Southern European countries. Estimates for yearly traffic volume in Portugal during the period 1982 - 1995 were calculated (COST 329, 2004). Afterwards, the method was further developed and updated estimates for the period 1980-2000 were produced (Cardoso, 2005).

The method is built on two basic hypotheses: 1. the gasoline and diesel sold (for road transport related activities) in a country are used in the production of its traffic volume; and 2. countries with comparable developing levels and great economic interaction have some similarities in the overall yearly fuel efficiency of their road transport systems.

Data on seven European countries (base countries) were collected and used to quantify those similarities, by statistical analysis.

It is assumed that there is a general time trend in the yearly average distance travelled per unitary amount of fuel, which is common to all base countries. This is the same as stating that there is a common time trend in the energetic efficiency of fuel consumption. Furthermore, in each base country, the yearly average distance travelled per unitary amount of fuel sold is supposed to be related to a factor specific to that country. Differences between energetic efficiency of each fuel type in each year and base country are captured by specific factors for each fuel type. It is also assumed that there is a common factor relating the energetic efficiencies for gasoline and diesel.

The general form of the model is given by the following equations:

$$TV_{it} = a_i \times f_t \times (GAS_{it} + k_{it} \times DIS_{it})$$

$$k_{it} = \frac{k_{truck} \times truck_{it} + k_{car} \times car_{it}}{truck_{it} + car_{it}} \quad \wedge \; k_{car} = 1 \; ; \; k_{truck} = 0.25$$

Where:

TV_{it} - traffic volume in country ***i***, during year ***t***;
a_i - country factor, dependent on vehicle fleet figures for country ***i***;
f_t - common trend for the unitary fuel efficiencies $\boldsymbol{b_{it}}$ and $\boldsymbol{c_{it}}$;
b_{it} - energetic efficiency in gasoline consumption in country ***i***, during year ***t***;
c_{it} - energetic efficiency in diesel consumption in country ***i***, during year ***t***;
k_{it} - common factor, relating the energetic efficiencies for gasoline and diesel, in country ***i***, during year ***t***;
GAS_{it} - gasoline sales in country ***i***, during year ***t***;
DIS_{it} - diesel sales in country ***i***, during year ***t***;
car_{it} - number of cars in country ***i***, in year ***t***;
$truck_{it}$ - number of trucks and buses in country ***i***, in year ***t***.

Factor $\boldsymbol{a_i}$ is mainly influenced by overall features of each country, namely the composition of its vehicle fleet, the road network characteristics and the overall road transport activity; it was fitted to each base country data, by non-linear regression analysis. Variable $\boldsymbol{f_t}$ reflects the overall development of fuel consumption efficiency in all base countries (the ones used for model fitting); it was estimated using principal component analysis. Factors $\boldsymbol{b_{it}}$ and $\boldsymbol{c_{it}}$ are used for uniformization of energetic efficiency of different types of fuel; their effect is combined by means of the common factor $\mathbf{k_{it}}$, which reflects the relative weights of the gasoline and diesel fleets.

In spite of the successful application of this method to the Portuguese traffic system, the generalization of the method to other countries should be preceded by research to confirm the applicability of the two basic hypotheses described above and to evaluate the suitability of the presented equations. Differences in basic characteristics of the vehicle fleet (vehicle types, age distribution, overall maintenance level) or in key features of the road networks (alignment characteristics, pavement types and traffic management systems) may exist, as compared to the base countries used in COST 329. These differences may require the fitting of new parameters to the basic equations. Also, taking into consideration the increase in the percentage of diesel passenger cars, assumptions concerning the ratio between consumption rates of gasoline vehicles and diesel vehicles may have to be revised, as well.

3.7.2. Odometer readings at regular vehicle inspections

Starting (after) 2002, the information on vehicle kilometres traveled in Denmark is based upon the mandatory vehicles technical inspection every second year. When a vehicle is inspected, the amount of kilometres travelled is registered and compared to the amount from the last inspection. Knowing the type of vehicle and the total numbers of that type, it is possible to give an estimate of kilometres travelled by type of vehicle.

Obviously, this type of procedure does not give any information on where the kilometres are driven. But the information is considered to be relatively reliable; at least for the time points on which the samples are taken. As mentioned above, it will be impossible to disaggregate the total number of kilometres into different road types or over time. Additionally, foreign vehicles are not included in the scheme. Unfortunately, no further information on this method is currently available.

An interesting development would be to include a questionnaire-based survey in the vehicle inspection process, so that the vehicle owner could report additional information on the type of travel and his personal characteristics, allowing for more disaggregate estimates.

3.8. Synthesis

In this Chapter, the main collection methods for "raw" exposure data were presented by means of examples from several EU Member States. The analysis focused on methods involving an intrinsic statistical error (i.e. sampling-based methods). However, other methods concerning more objective measurements (i.e. databases and census) were also described and assessed. Other methods for obtaining exposure data in a more aggregate form were also briefly discussed. 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.

Travel surveys are widely implemented to obtain different exposure data (vehicle- and person kilometres, time in traffic, number of trips) for several reasons, including cost effectiveness and the fact that, having persons as a unit allows for disaggregate analysis by combining driver, vehicle and road network characteristics. Moreover, surveys are carried out interactively, usually by telephone, allowing to reducing to some degree reporting errors.

However, the most critical problems encountered in surveys concern sampling, non-response and measurement errors. It appears very difficult to cover the entire target population and, in some cases the sample frame is not designated to cover the whole population (e.g children not included). In other cases the problem is more subtle; for example, telephone based surveys only contact persons that have access to a telephone. Moreover, larger families are more likely to be successfully contacted than small families, persons traveling more often are less likely to be contacted by a telephone survey and persons not traveling often are less likely to be contacted by a roadside survey.

It should be also noted that most of the errors discussed above are often uncalculated, not allowing for an assessment of the accuracy of the data. Another important limitation rises from the fact that, even when travel surveys are carried out at a systematic basis, time series of exposure estimates can not be obtained.

On the other hand, traffic count systems operating in most countries allow for continuous measurements of traffic volumes over time. However, this method is also sample-based, as the measurement sites may be more or less representative of the entire road network examined. Moreover, little is known on how exactly vehicle kilometres are derived from traffic counts. Two approaches seem to be in place; one based on weighted counts, in which a site is assumed representative for a number of other sites and another approach is to use models to estimate the counts for the non measured sites.

In any case, the estimated counts are multiplied by the length of the sections to obtain aggregate vehicle kilometres. When multiplying with average vehicle occupancy rates, person kilometres data can be produced form traffic counts.

However, an important compromise within this method is that it has vehicles as units, not allowing any classification per person characteristics. In some cases, the classification per vehicle and network characteristics may also involve difficulties.

Vehicle fleet and driving licenses databases are two other main sources of exposure data in most countries. The related exposure data (number of vehicles in use and number of drivers) can in principle be obtained though surveys. However, in most cases some approximations (number of registered vehicles and number of driving licenses) are derived from the respective official databases.

Obviously, the main problem with such registers is that only crude estimates of exposure can be derived. In countries where no other reliable source of exposure data is available, they allow for exposure estimates per driver characteristics (in the case of driving licenses) and vehicle characteristics (in the case of vehicle registers), however no combined analysis per driver and vehicle characteristics is possible. When combined with survey data on average distance traveled, average vehicle occupancy etc., vehicle- and person kilometre estimates can be obtained.

Both kinds of registers may share the problem of insufficient updating; although the introduction of new entries (vehicles or licenses) is accurately implemented, the removal of invalid entries (e.g. scrapped vehicles or deceased drivers) is not carried out systematically. Care should be taken that invalid entries are at least indicated as such in the databases. More accurate data on the actual number of vehicles in use and active drivers could be obtained by other registers, such as vehicle inspection databases (not available in most cases) and vehicle taxation or insurance databases (both not accessible in most cases).

Additionally, road registers, available in most countries in more or less advanced types, are used not only to extract (more or less disaggregate) road length information, but also to exploit this information for the calculation of vehicle kilometres from traffic counts. The usability of the data depends on the coverage of the road network by the register (usually only main national and rural roads are included).

Another approach that is in place in some countries is based on statistical modelling. In a sense, statistical modelling is also applied in some methods of obtaining vehicle kilometres from vehicle counts. This approach, however, relies even more on model assumptions. The basic idea is simple: if the average fuel consumption per kilometre is known for the vehicle fleet, the total

fuel consumption provides a rather reliable estimate of the total number of kilometres traveled. The method can be further calibrated when alternative sources are available (e.g. survey data). A problem with this method is that it is extremely difficult to distribute the kilometres traveled over time and in space. Therefore, the more aggregate the analysis is, the more reliable the results would be. It is interesting to notice that, technically, this method derives exposure data from other exposure data.

Finally, another emerging method for the estimation of vehicle kilometres is based on the use of odometer readings at regular vehicle inspections, providing the total number of kilometres travelled by a vehicle since the previous technical inspection. However, practically nothing is known as to where and when the vehicle travelled by means of this method, therefore the vehicle kilometres estimates are usable only in aggregate level. If combined with survey data, some of these issues may be partially resolved. It should be noted, though, that the main advantage of this method is that it could be used to benchmark or validate other methods.

The above methods to collect exposure data were highlighted by means of examples from seven representative countries: Denmark, France, Greece, Hungary, the Netherlands, Norway and Portugal. A basic finding of the review of common practices was that, although a lot of information on the features of each method is available, it is not always clear how the national exposure estimates are made, especially as far as vehicle and person kilometres are concerned. The availability and definitions of variables and values used (level of disaggregation) varies significantly among countries. Taking into account that each country uses various (and different) sources of "raw" data to calculate the national exposure estimates, it is obvious that a very good knowledge of the data collection and calculation processes of each country would be required in order to carry out international comparisons.

These comparisons are currently mainly carried out through International Data Files, gathering the available national exposure estimates of different countries over time. According to the analysis presented in this Chapter, it is interesting to investigate the availability, quality and comparability of the exposure data within these International Data Files. An exhaustive review on these issues is presented in the following Chapter 4.

References

1. Bjørnskau, T. (2005). Risk and Exposure Data in Norway. TØI Working paper. SafetyNet Task 2.1.
2. Bos, N; De Blois, C; Bijleveld, F. (2005). Report on Risk and Exposure Data availability, collection methodologies and use in the Netherlands. SWOV Working paper. SafetyNet Task 2.1.
3. Cardoso, J. (2005). Risk and exposure data availability, collection methods and use in Portugal. LNEC Working paper. SafetyNet Task 2.1.
4. Cardoso, J. L. (2005). The use of international data on fuel sales and vehicle fleet for the estimation of yearly national traffic volumes. Accident Analysis and Prevention, Vol. 37/1 (207-215).
5. Cochran, W.G (1963). Sampling Techniques. Second edition, Wiley, 1963, third edition, 1977
6. Cseffalvay, M; Holló P. (2005). Risk and Exposure Data in Hungary. KTI Working paper. SafetyNet Task 2.1.
7. COST 329 (2004). Models for traffic and safety development and interventions. Final report. Directorate-General for Transport, European Commission, Brussels.
8. FHWA, United States Department of Transportation - Federal Highway Administration, "Sources of Exposure Data for Safety Analysis", Publication No FHWA-RD-97-025, 1997.
9. FHWA, United States Department of Transportation - Federal Highway Administration, "FHWA Study Tour for National Travel Surveys", 1994.
10. Haddak, M; Bergel, R; Lejeune, P; Treny, V; Armoogun, J. (2005) Risk Exposure Data in France. CETE - INRETS. Working paper. SafetyNet Task 2.1.
11. Hemdorff, S. (2005). Risk Exposure Data availability in Denmark. DRD Working paper. SafetyNet Task 2.1.
12. Yannis, G; Papadimitriou E. (2005). Risk and Exposure Data availability, collection methodologies and use in Greece. NTUA Working paper. SafetyNet Task 2.1.

4. International data files with exposure data

4.1. General

Considerable efforts have been made since the early sixties towards setting up International Data Files (IDF) containing detailed data on road accidents and general transport system factors (mainly exposure) for different countries that may be used to evaluate accident risks and to compare the safety performance of different countries and regions.

The interest for IDF is explained elsewhere in this report, but at this stage mention should be made to the fact that the interest in international and inter-regional comparisons is not limited to the benchmark of safety performance (namely expressed as the number of accidents or victims divided by a suitable measure of exposure). From a national and regional point of view, provided that the appropriate level of disaggregation is available, these comparisons make it easier to identify less performing areas and overall safety issues; also, they make available a benchmark for what has already been achieved, and therefore sensible targets may be set. Another important aspect of international accident data files is the possibility to get some hindsight to the peculiarities of different national road systems that may affect the international transferability of national best practices and guide their adaptation to other states and regions.

In this chapter, a comparative presentation of risk exposure data (RED) in the IDF is presented, on the basis of information collected from contacts and interviews to the persons responsible for the IDF, as well as the related publications. The International Data Files examined are:

- Eurostat
- ECMT
- UNECE
- IRTAD
- IRF

The discussion is mainly concerned with the following main issues:

- data availability,
- definitions used,
- variables and values considered,
- collection methods
- data quality control.

The questionnaire used and the responses obtained are presented in the Annexes V and VI respectively. Following a brief presentation of the data files, the selected exposure data examined in the framework of this Report are analyzed in the scope of each IDF.

4.2. Brief description of the analysed IDF

4.2.1. Eurostat

The **EUROSTAT** (http://epp.eurostat.cec.eu.int) publishes since 1990 an annual publication (European Commission, 2005), with an overview of transport and energy statistics for the EU Member States. Data is collected by means of an official questionnaire to be filled by Member States. The objective is to provide the EU with high quality standardized data on transport.

Data is collected by means of the common EUROSTAT-UNECE-ECMT questionnaire, at the NUTS 2 level of aggregation. NUTS, which stands for "Nomenclature of Territorial Units for Statistics", is a five-level hierarchical classification of EU geographical space that includes three regional levels and two local levels. According to NUTS, each Member State is subdivided into a whole number of NUTS 1 regions, each of which is in turn subdivided into a whole number of NUTS 2 regions and so on.

Quality control of published data is ensured by the Members States through their official data providers. Therefore, no data quality control is carried out and no correction factors are applied, as the Member States have to comply with the common definitions.

All aggregate data is freely available on the internet (http://europa.eu.int/comm/dgs/energy_transport/figures/pocketbook/).

4.2.2. ECMT

The **European Conference of the Ministers of Transport** (**ECMT**) (www.cemt.org) publishes accident statistics since 1975. Between 1975 and 1984 these statistics were included in the Transport Statistics Yearbook (ECMT, 2004); since 1985 accident statistics are presented in a separate publication: the annual Road Accident Statistics Yearbook (ECMT, 2003). These publications are intended for supporting political decision-making concerning European transport policies. The ECMT road accident data file and the transport statistics database contain data on accidents and victims and on exposure related data that provides road accident related indicators (especially rates) that may be compared to similar indicators for other transport modes.

4.2.3. UNECE

The **United Nations** publishes since 1955, through its **Economic Commission for Europe** (**UNECE**) (www.unece.org), an annual publication

containing statistics on the road traffic system activity in Europe and North America (UNECE, 2005). Data on accidents and victims are presented, with data on road length, traffic volumes, number of registered vehicles and population. There are 55 countries in the UNECE data file; however the present analysis only focuses on the EU Member States.

The accident data in the database is concerned with injury accidents only. The data is collected from replies by member countries to a questionnaire and from official national and international sources. In fact, UNECE, the European Conference of Ministers of Transport (ECMT) and EUROSTAT have agreed on a common questionnaire concerning transport, with common definitions.

4.2.4. IRTAD

The **International Road Traffic and Accident Database (IRTAD)** (http://www.bast.de/htdocs/fachthemen/irtad) was established by the Steering Committee of the OECD Road Transport Research Programme, to provide a framework for the integrated collection of international aggregated data on accidents, injuries and exposure on a continuous basis.

This database is mainly research-oriented and its development was adjusted to the following objectives: scientific cooperation; collection of harmonized and timely aggregate accident and exposure data; improvement of data available for research and policy planning; harmonization of definitions; and identification of special safety issues deserving further research (Bruhning, 1995). The database was initially hosted by BASt, in Germany. The hosting has changed recently to the Joint OECD/ECMT Transport Research Center.

IRTAD was established in 1989. Annual aggregated data are collected for every year since 1970, on several safety related issues, such as accidents, casualties, exposure, including safety belt wearing rates. Currently, IRTAD has 50 member institutes and data is collected for all OECD countries, except Mexico and Slovenia (29 countries). Data are collected continuously, using electronic forms. Access (on-line or through diskette) to the database is only possible for members of IRTAD; however, a brief overview is available to the public on the internet.

Quality control of input data is performed, especially in what concerns recorded definitions and mathematical correctness. Corrective factors are applied to data that does not comply with the IRTAD standardized definitions. As an example the number of fatalities is adjusted to the Vienna Convention of Road Traffic “death within 30 days” definition. The total number of deaths according to the “dead within 30 days” definition is calculated multiplying the number of non-standardized fatalities by an adjustment factor. The following internationally accepted adjustment factors are used: 24 hours – 1.30; 3 days – 1.15; 6 days – 1.09; 7 days – 1.08.

4.2.5. IRF

The **International Road Federation (IRF)** (www.irfnet.org) is a non-governmental, not-for-profit international organization established in 1948 to promote development and maintenance of better and safe roads and road networks. Members include both private and public organizations, including some government agencies, from several countries worldwide.

Development of the IRF database started in 1958, and the first data tables were first published in 1964, concerning 20 countries. Data are collected annually, using paper and electronic forms. Aggregated data for 84 countries (up to year 2002) are presented in the 2004 data tables. On-line access to the data is provided to IRF members only.

No validation is performed on the provided data, since these are national official data. However, when needed, national representatives provide corrections to data previously sent. Even though, the overall impression conveyed by IRF experts is that further improvements in the collected data could be obtained.

4.3. Comparative analysis of the collected RED

In this section, selected exposure measures are compared among the examined IDF. The comparison is carried out by means of a ratio, of which the denominator is Eurostat figures for a particular exposure indicator, and the numerator is the respective figures of another IDF. Moreover, the ratios are calculated for all data disaggregation levels, i.e. for different categories or variables of each indicator.

It is obvious that, in theory, this ratio should be equal to one for all exposure measures and related sub-categories, indicating that all IDFs publish the same figures per country, per year and per exposure measure. In practice, however, ratios significantly higher or lower than one were calculated, indicating significant differences among the IDF figures per county, per year and per exposure indicator. These differences are graphically described and commented for each exposure indicator, and are summarized and discussed in the synthesis section.

4.3.1 Road length

As regards the collection of data on road length, the common questionnaire for EUROSTAT, UNECE and ECMT divides the roads in two major classes: motorways and other roads. The “other roads” class is further divided in three administrative classes, resulting in a total of four road classes. The length of international E roads is also collected. There is no specific reference to type of road environment (urban or rural), as a classification criteria.

Tables published by EUROSTAT and UNECE present the road length for each of the mentioned four road classes (UNECE, 2005a, UNECE, 2005b, European Commission, 2005).

Published tables by ECMT do not address specifically road length (UNECE, 2005). However, an additional road disaggregation is used in the common questionnaire specifically in what concerns accidents and victims, which indicates that additional data is being collected. Accident data are divided in four road classes (UNECE, 2005):

- motorways;
- roads in built-up area;
- roads outside built-up area;
- unknown type of road.

The IRF database contains information on road length by class of road (Motorway; Main Highways or National Roads; Secondary or Regional Roads; and Other Roads), type of operation (public or private), type of surface (paved

and unpaved) and condition (good, fair or poor). The IRF database includes data on road density (in km per km^2), as well.

The IRTAD database contains data on road length, according to four road classes: motorways; A-level rural roads; other rural roads; and roads inside urban areas. A-level roads are the primary national road network (IRTAD, 1998).

In summary, EUROSTAT, IRTAD and IRF consider four road classes in their published tables. IRTAD road classes differentiate urban roads from rural roads. Similarities can be found between the road classifications of IRF and EUROSTAT/ECMT/UNECE, as both include operational (motorway/non motorway) and administrative (main/secondary) criteria and do not take in account the type of road environment (urban or rural).

The road classifications adopted in different IDF are summarized in the following Table 4.1.

Table 4.1. Road classes considered in different IDFs

ROAD CLASS	UNECE/EUROSTAT/ECMT	IRF	IRTAD
Motorway	X	X	X
A level			X
Other rural			X
Urban road			X
Main Highway		X	
State highway	X		
Secondary highway		X	
Provincial Highway	X		
Other		X	
Communal Highway	X		

A comparison of 2000 and 2001 data from EUROSTAT and IRF is presented in Figures 4.1 to 4.3. Comparison is presented by means of ratios, where the denominator is EUROSTAT data and the numerator is the other IDF data. The EUROSTAT database tables do not provide data on the main and secondary road network lengths for 2001 (European Commission, 2005).

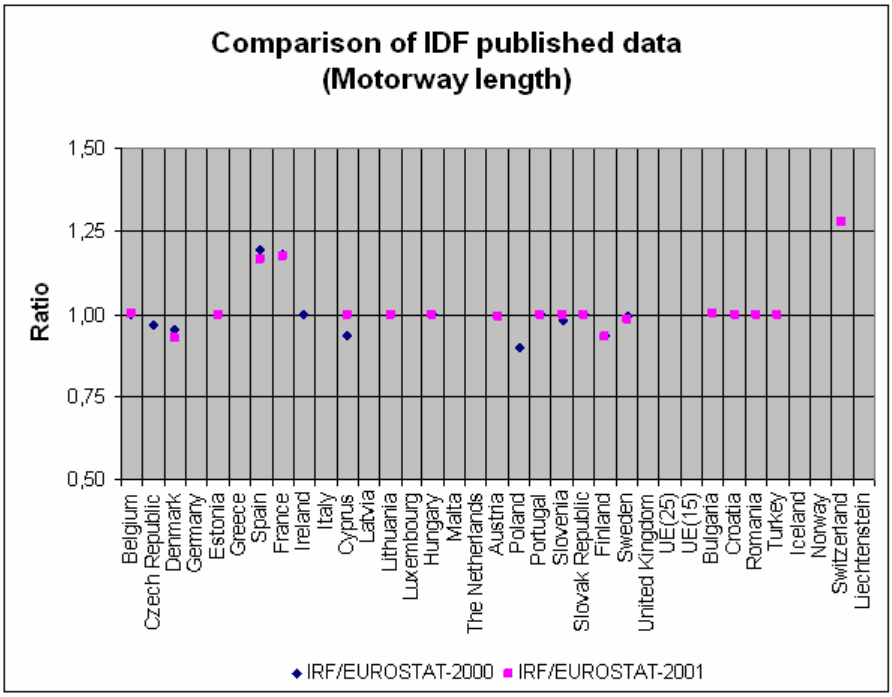


Figure 4.1. Comparison of motorway length data in EUROSTAT and IRF

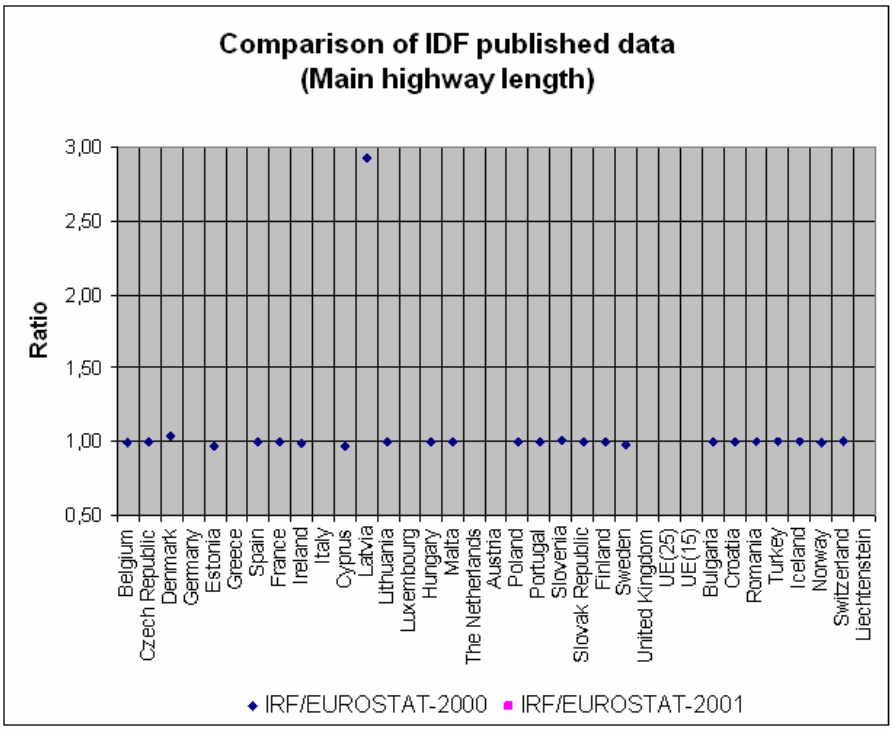


Figure 4.2. Comparison of main highway length data in EUROSTAT and IRF

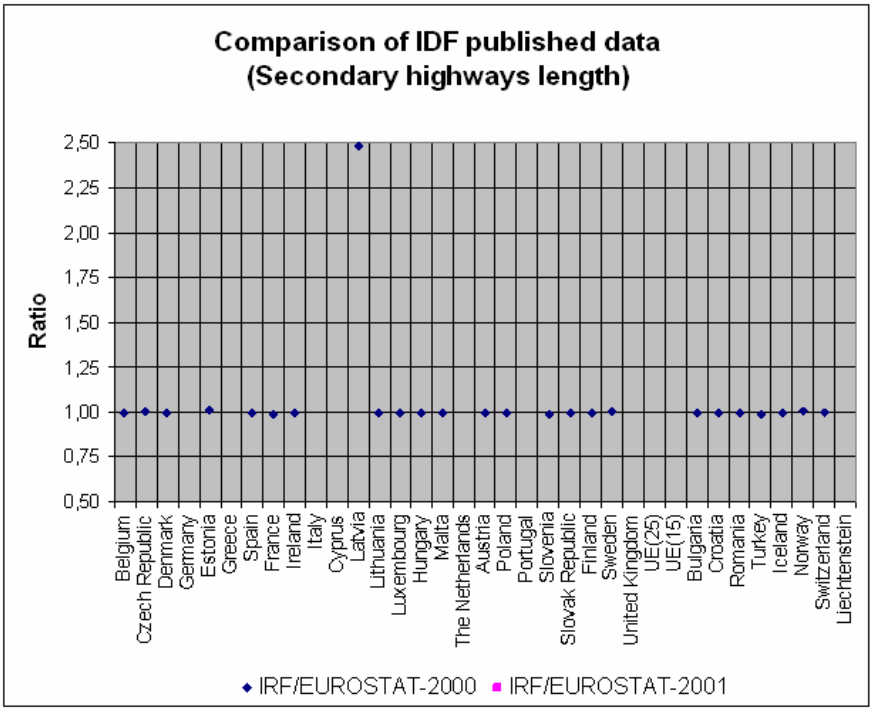


Figure 4.3. Comparison of secondary highway length data on EUROSTAT and IRF

In what concerns the comparison of EUROSTAT and IRF data, no major differences between the main and secondary highway length data were detected for the 25 EU countries. The only detected case is probably due to a typing error in the IRF database. In what concerns the data on motorway length, several cases of significant differences were identified (10 cases). The differences vary between -7% and +28%.

EUROSTAT and UNECE motorway length data show considerable agreement, except for one single country. Considerable disagreement is observed in main and secondary highway length data. This, in part, is due to differences in terminology, concerning the concepts of main, secondary, state, provincial and local roads.

4.3.2 Vehicle kilometres

As regards the collection of data on travelled distance, the common questionnaire for EUROSTAT, UNECE and ECMT is designed to collect data on four vehicle classes (motorcycles, passenger cars, buses, lorries and road tractors), irrespective of the road class.

However, published UNECE tables with yearly data provide travelled distance for five vehicle classes: mopeds; motorcycles; passenger cars; buses, coaches and trolley buses; and lorries and road tractors. In the last available

publication from UNECE (UNECE, 2005a), data for 31 countries are provided. However, several countries do not provide data for two wheeled vehicles.

The last EUROSTAT tables available do not provide any data on travelled distance by vehicles (Eurpoean Commission, 2005).

Yearly data for travelled distance are provided in the IRF database for four vehicle classes: motorcycles and mopeds; passenger cars; buses and coaches; and lorries and vans (IRF, 2005).

The IRTAD database contains data on travelled distance, according to the four road classes and to the six vehicle types considered (mopeds and mofas -mopeds with maximum speed of 30 km/h); motorcycles and scooters; passenger cars and station wagons; goods motor vehicles; buses; and other motor vehicles).

In summary, availability and disaggregation of travelled distance by road vehicles varies among the analysed IDF. UNECE, ECMT and IRF have information disaggregated by vehicle class; IRTAD has information disaggregated by road class and vehicle class; the most recent EUROSTAT tables do not contain any information regarding vehicle×kilometres travelled.

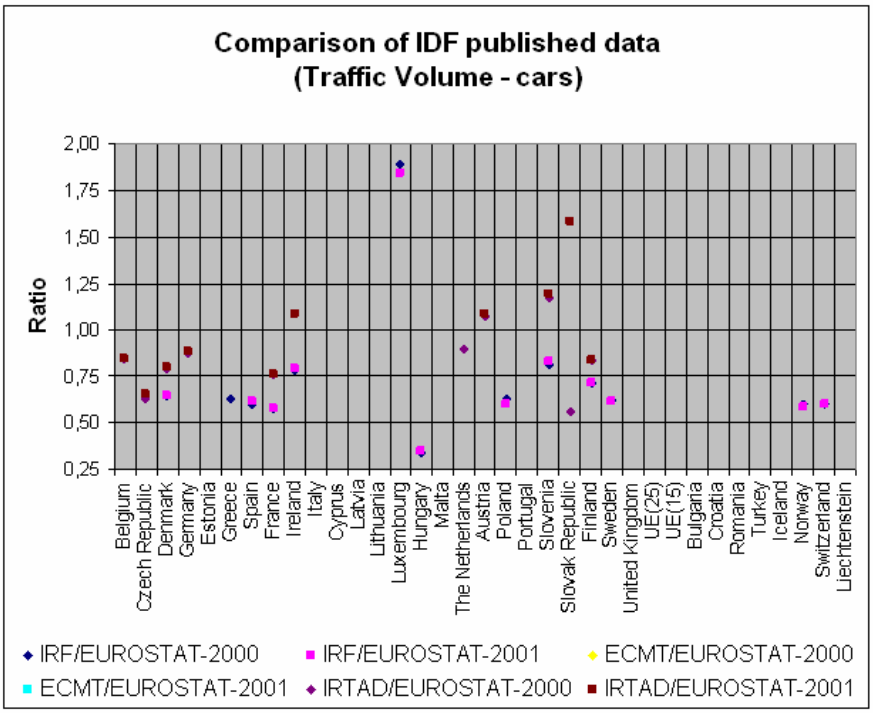


Figure 4.4. Comparison of distance traveled by passenger cars in EUROSTAT, IRF, ECMT and IRTAD

A comparison of 2000 and 2001 data from EUROSTAT, IRF, ECMT and IRTAD is presented in Figure 4.4. Considerable differences are detected: +/- 80%.

4.3.3 Transport measurement

As regards the collection of data on transport activity, the common questionnaire for EUROSTAT, UNECE and ECMT is designed to collect data on passenger travelled distance for three vehicle classes (motorcycles, passenger cars and buses) and on goods haulage distance for lorries and tractors. No disaggregation by road class is made. These data are included in the tables published by EUROSTAT (European Commission, 2005). Comparisons with other transport modes are published, as well.

UNECE tables do provide data on passenger travelled distance by vehicle type (motorcycles, passenger cars and buses) and goods haulage distance by type of traffic - national, international, loaded and unloaded, cross trade and road cabotage (UNECE, 2005a).

Yearly data for passenger kilometres are provided in the IRF database for road based private and public transport. Concerning the amount of surface goods transport, the IRF data tables contain the total transported tonnage (in tons) and the total hauled road distance (in tons×km).

IRTAD contains information about passenger kilometres for passenger cars and public transportation by year, but not for all countries.

In summary, as observed in the previous section, availability and disaggregation of passenger travelled distance and of goods haulage distance by road vehicles varies among the analysed IDF. EUROSTAT, ECMT and IRF have information disaggregated by vehicle class, even though the classes do not overlap completely; UNECE and IRTAD tables do not contain any information regarding ton×kilometres travelled; UNECE does not have data on passenger×kilometres traveled.

A comparison of 2000 and 2001 data from EUROSTAT, IRF, UNECE and ECMT is presented in Figures 4.5 and 4.6. Availability of this data in EUROSTAT is good, for passenger cars and for buses; the same cannot be said for two wheeled vehicles. On the IRF and UNECE data bases these data are missing for several countries. No ECMT data for 2001 was collected.

In what concerns the distance travelled by passengers of private transport, no major differences were detected between the two IDF. Data on passenger distance travelled by means of private transport does not differ very much with the IDF (-10%, +12%), except for a pair of cases that appear to be caused by wrong data input. Data on passenger distance travelled by means of public

transport show large variation, according to the originating IDF; there seems to have considerable differences in the way each the number of passenger-kilometres is considered in each IDF, in spite of the fact that EUROSTAT and UNECE data are collected with the same data form.

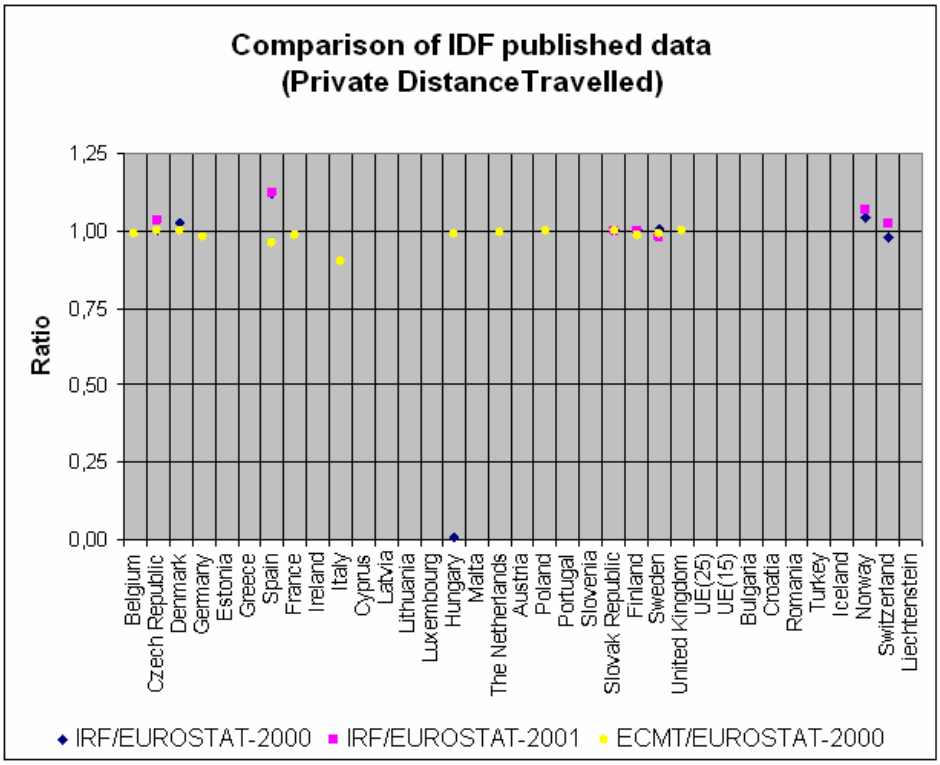


Figure 4.5. Comparison of distance travelled on private vehicles in EUROSTAT, IRF and ECMT

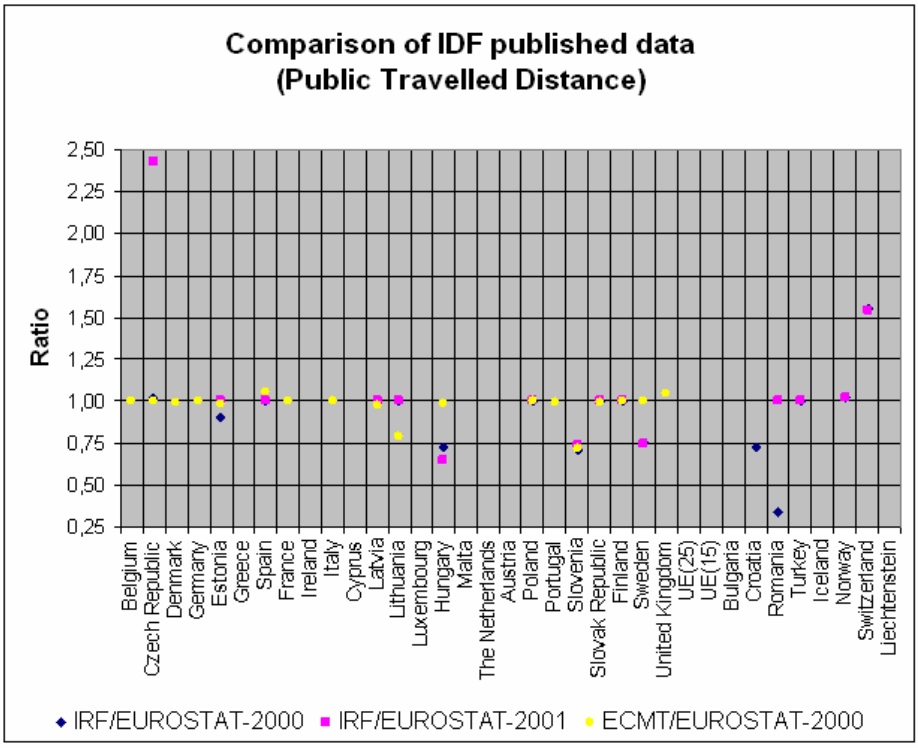


Figure 4.6. Comparison of distance traveled on public vehicles in EUROSTAT, IRF and ECMT

4.3.4 Registered vehicles

As regards the collection of data on the number of registered vehicles, the common questionnaire for EUROSTAT, UNECE and ECMT has nine road vehicle classes: mopeds; motorcycles; passenger cars; motor coaches, buses and trolley buses; trams; lorries; road tractors; semi-trailers; and trailers.

Motorcycles are further divided in two classes according to the engine size. Passenger cars are further divided by age (four classes), type of energy and engine size (10 classes), unloaded weight (four classes). Buses are divided by age (four classes) and type of energy (six classes). Lorries are divided by age (four classes), load capacity (eight classes), motor energy (six classes) and kind of transport (two classes). Road tractors are divided by age (four classes), type of energy (three classes) and kind of transport (two classes). Semi-trailers are divided by load capacity (five classes) and kind of transport (two classes). Trailers are divided by load capacity (five classes) and kind of transport (two classes).

However, yearly data tables for registered vehicles produced by UNECE database allow for eight vehicle classes, only: mopeds; motorcycles; passenger cars; buses, coaches and trolley buses; lorries; road tractors; semi-trailers; and trailers. In the last available publication (UNECEa, 2005), data for 43 countries are provided. Some countries do not provide data for mopeds.

The EUROSTAT tables for the yearly data on registered vehicles comprise five vehicle classes: mopeds; motorcycles; passenger cars; buses, coaches and trolley buses; and goods vehicles (European Commission, 2005).

The ECMT tables contain five vehicle classes: mopeds; motorcycles; passenger cars; buses, coaches and trolley buses; lorries and road tractors.

The yearly number of vehicles is provided in the IRF database for five vehicle classes: passenger cars; buses and coaches; lorries and vans; road tractors; and motorcycles and mopeds (IRF, 2005).

Yearly IRTAD tables provide vehicle registration data. Six vehicle types are considered: mopeds and mofas (mopeds with maximum speed of 30 km/h); motorcycles and scooters; passenger cars and station wagons; goods motor vehicles; buses; and other motor vehicles.

In summary, data availability on the number of registered vehicles is good in the analysed IDF. Two wheeled vehicles are separated from the rest of vehicles in all databases. Most IDF separate cars from buses, and both these vehicle classes from lorries and from road tractors. The only exception is the IRTAD database, which considers only goods vehicles (aggregating both lorries and road tractors in the same category).

A comparison of 2000 and 2001 data from EUROSTAT, IRF and UNECE is presented in Figures 4.7 to 4.9. Availability of this data is good in all IDF, for passenger cars, buses and lorries. As regards two wheeled vehicles, only IRF and UNECE do provide data for a considerable number of countries; for these vehicle it was possible to collected data on two countries in EUROSTAT.

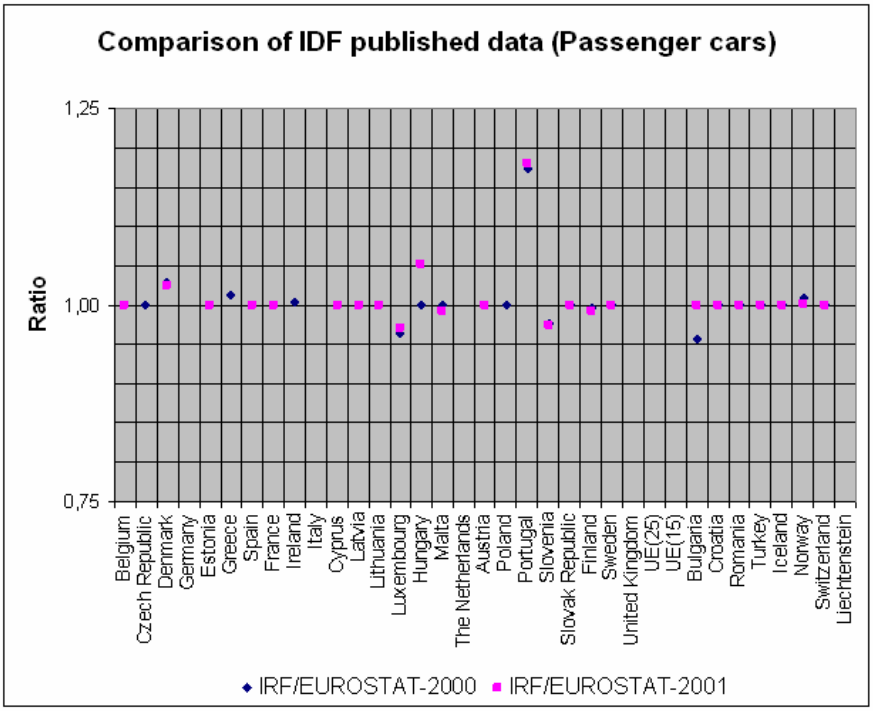


Figure 4.7. Comparison of number of cars in EUROSTAT and IRF

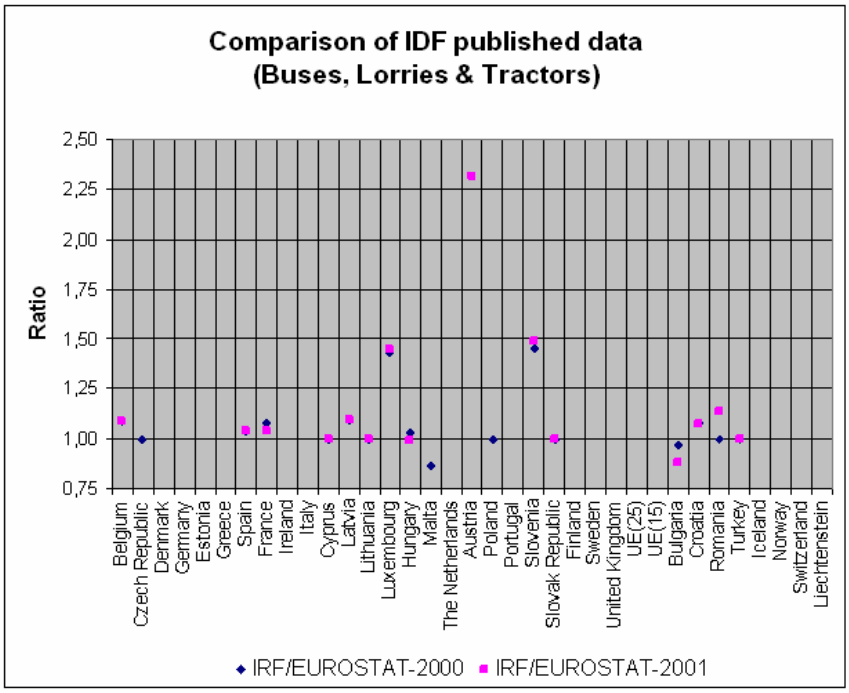


Figure 4.8. Comparison of number of buses, lorries and tractors in EUROSTAT and IRF

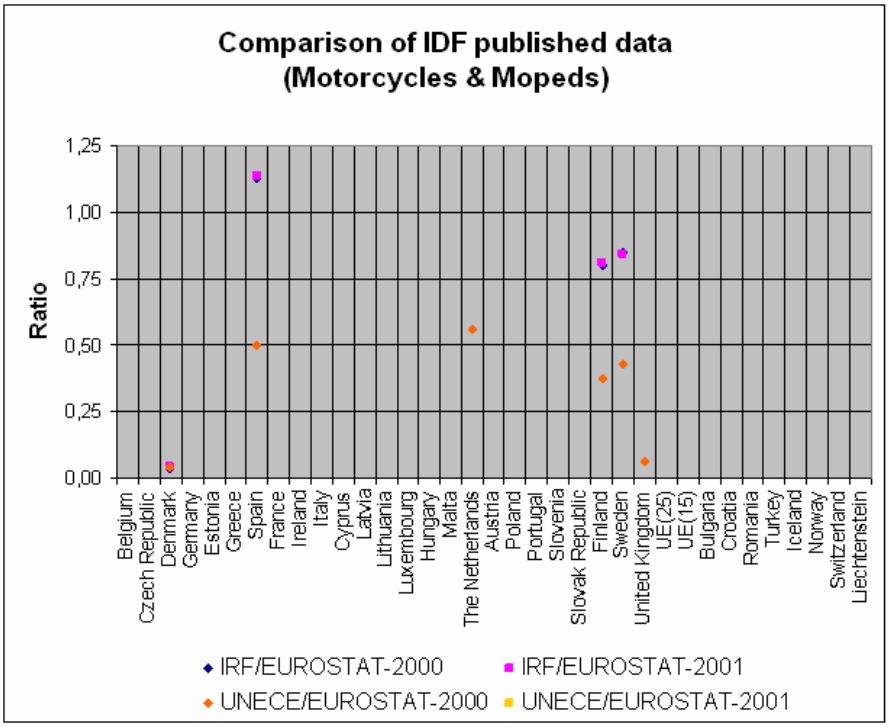


Figure 4.9. Comparison of number of motorcycles and mopeds in EUROSTAT, IRF and UNECE

Differences between the number of cars in each IDF are minor (largely within the +/-5% band). The registered numbers of buses, trucks and lorries present differences that vary considerably The number of comparable numbers of two wheeled vehicles amounts to very few cases.

4.3.5 Population

The UNECE/EUROSTAT/ECMT databases contain data on resident population, according to eight age classes: less than 6 years old; 6 to 9 years; 10 to 14 years; 15 to 17 years; 18 to 20 years; 21 to 24 years; 25 to 64 years; and 65 or more years old.

In the last available UNECE publication (UNECE, 2005a), data for 55 countries are provided. Overall there are no missing data; however, some countries do not provide the data in agreement with the standardized age group classification.

The IRF database does contain information on each country's total population, for the years since 1994. No disaggregation of population by age group is provided.

In IRTAD, population data is divided in twenty age groups: 0 to 5 years old; 6 to 9; 10 to 14; 15 years old; 16; 17; 18; 19; 20; 21 to 24; 25 to 64; 25 to 34; 35

to 44; 45 to 54; 55 to 59; 60 to 64; 65 to 69; 70 to 74; 75 to 79; and 80 years or more.

In summary, the importance of population as an overall accident risk indicator at the national level is recognized in all analysed IDF. Nevertheless, disaggregation of published data by age group varies with the considered IDF: IRF and EUROSTAT tables do not provide classification by age group. In the case of EUROSTAT, however, use of other EUROSTAT statistical tables, not directly related with transport, may overcome the absence of this information.

A comparison of 2000 and 2001 data from EUROSTAT, ECMT, IRATD and IRF is presented in Figure 4.10. No major differences were detected, apart from a (possible) typing error in IRF (concerning Hungary).

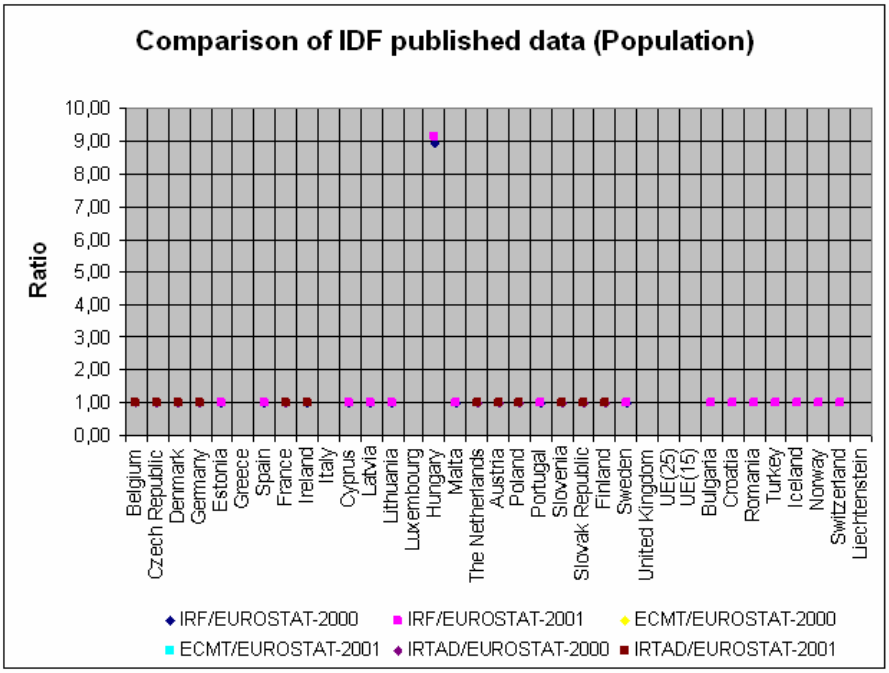


Figure 4.10. Comparison of the number of country inhabitants on EUROSTAT, IRF, ECMT and IRTAD

It should be also noted that it was not always possible to determine whether population figures correspond to a specific data of reference (e.g. population at 1st January, or 31st December). EUROSTAT presents average yearly figures; the ECMT presents population figures on 31st December of each year; No information was available on this issue as regards UNECE and IRF.

4.3.6 Driving licenses

None of the examined IDF (EUROSTAT, IRF, IRTAD, UNECE, ECMT) contains data on the number of driving licenses issued in each country.

It can be concluded that IDF administrators have not credited great importance to the number of driving license holders as a possible overall accident risk indicator at the national level. This lack of credibility may result from the difficulties in gathering updated data, as concluded in Chapter 3, at least for some countries.

4.3.7 Fuel consumption

The UNECE database does not contain data on each country's fuel consumption, or surrogate variables, such as fuel sales.

The EUROSTAT database does not contain procedures for collecting data on each country's fuel consumption in transport related activities. However, data from ECMT is used, for this purpose.

The IRF database contains information on the total amount of diesel and petrol consumed and on the percentage used by transport vehicles. Disaggregate data are provided for gasoline and for diesel.

The IRTAD database does not contain data on each country's fuel consumption, or surrogate variables, such as fuel sales.

In summary, data availability on fuel sales varies considerably with the analysed IDF. Information on this topic is only available in IRF and ECMT databases.

4.3.8. Number of trips

None of the examined IDF (EUROSTAT, IRF, IRTAD, UNECE, ECMT) contains data on the number of trips performed in each country.

4.3.9 Time in traffic

None of the examined IDF (EUROSTAT, IRF, IRTAD, UNECE, ECMT) contains data on the time spent in traffic in each country.

4.4. Synthesis

Table 4.2 below summarizes the overall situation as regards the published RED provided by IDF in publications available to general public, as described in the previous items. The availability in terms of countries and years varies significantly in the IDF. Moreover, the definitions of exposure measures, variables and values may differ among countries. However this information is not examined in this Table.

Table 4.2. Exposure data availability on the analysed IDF*

Exposure indicator		International data file (IDF)				
		EUROSTAT	ECMT	UNECE	IRTAD	IRF
Road length	(km)	●	-	●	●	●
Traffic volume	(vkm)	●	●	●	●	●
Transport activity	(pkm)	●	●	●	●	●
	(tkm)	●	●	●	-	●
Vehicle stock	-	●	●	●	●	●
Population	-	●	●	●	●	●
Driving licenses	-	-	-	-	-	-
Fuel sales	(t)	-	●	-	-	●
Number of trips	-	-	-	-	-	-
Time in traffic	-	-	-	-	-	-

● included
- not included
**availability per country and year not examined*

Concerning the issue of timeliness of published data, three IDF provided estimates for the average time delay for the publication of a reference year's data: one year for ECMT and IRF; and two years for UNECE. However, mention should be made to the fact that, as of July 2005, the most recent publications of the analysed IDF relate to years 2003 (EUROSTAT and IRTAD), and 2002 (ECMT, UNECE and IRF). Therefore, the delay in most IDF is over 2.5 years, which may be considered as not excessive. However, mention must be made to the fact that for some road data items, the most recent data provided in the latest EUROSTAT publication (2) relates to year 2002 and, in a few cases, to year 2000.

Some of the analysed IDF use common definitions and there is, to some extend, overlapping in the collected data and the corresponding published tables. This indicates that there is scope for combining the data collection

procedures in a common questionnaire. That has been already achieved to a great extent with the EUROSTAT-ECMT-UNECE common questionnaire.

Important RED are not collected in some IDF: that is the case for fuel sales (which may be used to estimate the amount of vehicle kilometres) and, especially the number of active driving licenses in each country. It is recognised, however, that fuel sales are not important if there is a direct indicator for travelled distances (such as vehicle.km). In addition, not all relevant disaggregated secondary variables are collected by all countries: this is the case for two wheeled vehicles, especially mopeds and bicycles, for which few countries consistently provide data.

As regards quality control, IRF acknowledges that they do rely on the quality control systems used by their data providers, as in most cases they are using official data. UNECE does not have internal quality control. EUROSTAT and ECMT have some routines for internally checking the data provided, especially in what concerns the coherence between partial and total values and with the values published on other databases. IRTAD checks the correctness of received data and, especially with new members, may resort to follow up actions to ensure correct use of agreed definitions.

The presented comparison of two years' data from EUROSTAT, UNECE, ECMT and IRF, highlighted the fact that differences in definitions may exist as regards some disaggregated basic variables such as motorway length, heavy vehicle and two wheeled vehicle fleets, and the distance travelled by public transport users. A few (most probably) typing errors were detected, as well.

The road classification definitions used in the EUROSTAT/ECMT/UNECE and IRF databases are slightly different from the ones used in the CARE database, which are very similar to the ones used in IRTAD.

The vehicle kilometres and transport measurement variables used in the EUROSTAT/ECMT/UNECE databases are not disaggregated by road class which limits the interest of its use in the context of CARE. Again, the information provided by IRTAD fits very well in the CARE database.

The vehicle classification used in the EUROSTAT/ECMT/UNECE common questionnaire can be applied in the context of the CARE database. However, mention is made to the fact that the published data is divided by less vehicle classes than are referenced in the questionnaire. Agreement between EUROSTAT and CARE vehicles class data is dependent on the registration of the disaggregated data on the database, instead of the aggregated published data. Again, the information provided by IRTAD fits very well in the CARE database.

Concerning the information on population, the EUROSTAT / ECMT / UNECE common questionnaire and the IRTAD database contain disaggregated data that can be used jointly with related information collected in CARE.

Table 4.3. Risk indicators in international data files

Risk indicator		International data file (IDF)				
		EUROSTAT	ECMT	UNECE	IRTAD	IRF
Accidents per inhabitant					●	
Accidents per vehicle-.km	General				●	●
	Build-up				●	
	Road class				●	
Fatalities per inhabitants	General	●	●	●	●	
	Age group	●		●	●	
	Age group and sex	●		●		
Fatalities per licensed drivers						
Fatalities per vehicles			●		●	
Fatalities per road user by type			●		●	
Fatalities per vehicle.km	General				●	●
	I/O build-up area				●	
	By road class				●	
Injured per inhabitants	General	●			●	
	Age group	●			●	
	Age group and sex	●				
Injuries per licensed drivers						
Injuries per vehicles					●	
Injuries per vehicle.km	General				●	
	Vehicle.km build-up area				●	●
	By road class				●	

Table 4.3 above summarizes the overall situation as regards the published road accident and casualty indices provided by IDF in publications available to general public, as described in a previous publication (DUMAS, 1998, NTUA, 1996). The corresponding risk exposure factors are highlighted in bold in the first two columns. Comparing Tables 4.2 and 4.3, it can be concluded that

most IDF related publications present risk indicators that use only a selected number of the possible risk exposure factors for which there is information in the corresponding database.

According to the descriptions provided, most IDF provide a single set of tables with a fixed array of queries on its data; only IRTAD allows for user defined queries of its content, making it, in fact, a user oriented IDF.

References

1. Bruhning, E. (1995). Five years of IRTAD: Review and perspectives. Seminar on international road traffic and accident data bases. OECD, Helsinki, 11-13 September 1995.
2. DUMAS (1998) Accident investigation: final report Work Package 4. Prepared by DENCO, Universitat of Brescia, INRETS, TRL, AUTh. Developing Urban Management and Safety (DUMAS).
3. ECMT (2003). Statistical Report on Road Accidents in 1999-2000. March 2003
4. ECMT (2004). Annual Report 2004. Joint OECD/ECMT Research Center. May 2004.
5. ECMT (2005). Trends in the Transport sector – 1970-2003. ECMT, Paris.
6. European Commission (2005). EU energy and transport in figures. EC, Luxembourg.
7. IRF (2005). World Road Statistics - 2004. IRF, Geneva.
8. IRTAD (1998). Definitions and data availability. OECD - RTR Road Transport Research Programme. BASt Germany, June 1998.
9. National Technical University of Athens (1996). Overview of international road accident data files. DGT-EC.
10. UNECE (2005a). Annual bulletin of transport statistics for Europe and North America. UNECE, Geneva.
11. UNECE (2005b). Statistics of Road Traffic Accidents in Europe and North America. Vol L – Tom L. UNECE, Geneva.

5. Conclusions and recommendations

5.1. Summary of the state-of-the-art

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.

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. However, in order to allow certain types of analysis at the EU-level, comparable data and definitions should be made available. 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 both at microscopic and macroscopic level.

However, reliable analysis of road accident data has to deal with a **series of limitations** related to the availability and the quality of essential information, both at national and European level. In particular, 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.

From the results of the present research, it is obvious that **comparing risk rates**, especially at international level, may be a very complex task. Both accident counts and exposure measures present some theoretical and practical limitations and are subject to estimation errors, which may compromise their usability.

Especially as far as exposure is concerned, 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.

In road safety analyses, different exposure measures are thus 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 be therefore noted that no general rule is available concerning the preferred measures of exposure.

However, it can be deduced that the most appropriate and recommended measurements of exposure appear to be **vehicle- and passenger-kilometres of travel**, as well as time spent in travel, the latter being less widely used in road safety analysis. These measures are conceptually closer to a theoretical definition of exposure and can be theoretically available to a satisfactory level of detail. However, they cannot be collected in the required level of detail on other than a systematic basis. In several EU countries, different systems exist and national exposure estimates are produced, whereas in some countries no data on vehicle- or passenger kilometres are available.

Therefore, **other exposure measures** are often used; namely the vehicle fleet and the drivers' population, the road network length, the fuel consumption, as well as the population figures. Although this data concern cruder exposure estimates and can replace the vehicle and passenger-kilometres only under specific conditions, they are widely used for the calculation of accident risk rates, mainly because they involve less complex collection methods and can more easily lead to comparable figures at EU level.

Table 5.1. Comparison of exposure measures

Road safety Outcome	Accidents / Persons	Persons	Accidents	Accidents / Persons	Accidents	Persons	Persons	Accidents / Persons	Persons
Amount of exposure	Vehicle - kilometres	Person - kilometres	Road Length	Fuel consumption	Vehicle Fleet	Population	Driver population	Number of trips	Time in traffic
Context of analysis	Traffic	Traffic - Mobility	Traffic - Infrastructure	Traffic	Traffic	Epidemiology	Traffic	Traffic - Mobility	Traffic
Temporal variation	•	•		•				•	•
Regional variation	•	•	•			•		•	•
Disaggregation level									
Road User category		•				•		•	•
User characteristics	•	•				•	•	•	•
Vehicle characteristics	•	•		•	•			•	•
Road network characteristics	•	•	•					•	•

Table 5.1 above summarizes the discussion presented in Chapter 2 of this Report, as far as the theoretical features of different exposure measures. It should be noted, however, that the features presented in the Table concerns the **theoretical properties of exposure measures**. In general, the availability, quality and disaggregation level of exposure measures may be compromised by limitations and particularities of the respective collection methods.

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** used in the EU (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. 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. 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 exposure 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, as complex mathematical models may be required.

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. 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. 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 (digital mapping, GIS etc.) is expected to improve data availability and quality in the coming years.

In Table 5.2 below, the main characteristics (exposure measures, variables and values) of the various RED collection methods in several EU Member States, as discussed in Chapter 3 of this Report, are summarized. It is obvious that the use and specification of methods varies significantly among the examined countries. Moreover, the availability, disaggregation and comparability of variables and values are also quite diverse. In particular, the **disaggregation level** theoretically possible for an exposure measure is seldom achieved in practice. Taking into account that even the theoretical disaggregation potential of exposure data is by far lower than the respective disaggregation level of accident data, it is obvious that the disaggregation potential of risk figures is mainly determined by the respective disaggregation potential of exposure data.

Additionally, as mentioned previously, **data from different sources (collection methods) may be 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.

Table 5.2. Exposure data collection methods and variables in selected EU countries

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

** more bullets indicate a more detailled classification*

*** the travel survey in not official; traffic counts system was operational up to 1993*

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 data and RED in the EU are the following:

- Eurostat
- ECMT
- UNECE
- IRTAD
- IRF

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 RED 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. Additionally, it is not always known whether the IDFs receive more (disaggregate) data than they publish. However, there is some evidence that **the more disaggregate national exposure data are not exploited at international level**, at least within the context of IDFs.

In the following Table 5.3 an overview of the IDFs examined in the framework of the present analysis (Chapter 4) is presented, focusing on the availability of exposure data and the related disaggregations.

Table 5.3. Overview of exposure data in the International Data Files

	Eurostat	ECMT	UNECE	IRTAD	IRF
Location	Luxembourg	ECMT, Paris	UNECE, Geneva	BASt, Koeln	IRF, Geneva
Contact person	Mr. H. Strelow	Mr. M. Barreto	Mr. M. Jovanovic	Dr. A. Schepers	Mrs. M. Mudbary
Data File description					
Number of countries	25	50	55	29	84
Available time series	1960-	1960-	1960-	1970-	1995-
Transport statistics	•	•	•	•	•
Accident statistics	•	•	•	•	•
Other statistics	•			•	•
Data collection method	Common questionnaire			questionnaire	questionnaire
Disaggregate/Aggregate data	aggregate	aggregate	aggregate	aggregate	aggregate
Access to the data	free/on-line	free/on-line	free/on-line	members only	members only
Publications*	•••	••	••	•••	•
Data quality control	limited	limited	limited	limited	
RED availability**					
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	•		•	•	•
Fuel consumption		•			•

**more bullets indicate more publications*

***availability by country and by year is not examined*

It should be noted that data availability in the different IDFs does not always imply comparability. Apart from the intrinsic comparability issues due to the national collection methods, as discussed above, other issues may further compromise the comparability of exposure data in the IDFs, not only among countries, but also among IDFs. In the framework of the present analysis, it was demonstrated that **differences in the published exposure estimates are observed among the IDFs**, these differences being more significant for the more "sophisticated" exposure measures (i.e. vehicle and passenger kilometres).

These differences may be attributed to the fact that some of the exposure estimates in the IDF may be based on crude national estimates, whereas the actual data source may not always be known. Additionally, another reason may concern **insufficient data quality control**, which may be either not carried out at all, or limited to the correction of only obvious mistakes by checking the totals and comparing with other 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. 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.

5.2. 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 RED limits significantly the comparability of the detailed national data**. On the other hand, the International Data Files containing road safety related data, including RED, 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 RED at European level. It is also obvious, from the discussion presented in this Report, that **the most useful RED are the least available**. Consequently, combining different sources of data could allow assessing the reliability of exposure estimates.

Further work and research should also 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 should be underlined that, from the results of the state-of-the-art survey, it is clear that the different exposure measures present different advantages and limitations, according to their properties and the context of the analysis. However, 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, both at national and EU level. However, given the importance of an improved RED 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 RED 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**. The harmonization of the definitions of exposure measures, variables and values between countries (at the most disaggregate level), in accordance to the existing accident data (i.e. CARE), as well as the current and future exposure data needs, would be an important first step to improve comparability of the existing disaggregate data. Moreover, the harmonization of definitions, variables and values among the International Data Files, and the improvement of the data quality control process in the related databases would also contribute to the improvement of the comparability of the aggregate data at international level.

Annex I - Expectation and variance of a ratio

Expectation and variance of a ratio.

Taken from Rice (1995, p 153).

If Z = Y/X then:

$$\mathrm{Var}(Z) \approx \frac{1}{\mu_X^2}\left(\sigma_X^2 \frac{\mu_Y^2}{\mu_X^2} + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y \frac{\mu_Y}{\mu_X}\right)$$

where

μ_X is the expected value of X and μ_Y is the expected value of Y,

σ_X is the standard deviation of (the error in) X and σ_Y is the standard deviation of (the error in) Y and

ρ is the correlation between the errors in X and Y.

Usually, the correlation between the errors in X (exposure) and Y (accidents) is assumed to be nil. Obviously then, if the relative error in X, (σ_X / μ_X) is small, then $\mathrm{Var}(Z) \approx \sigma_Y^2 / \mu_X^2$.

Sometimes this approach is extended to correcting the expected value of the ratio as well. See also Rice (1995, p. 153) in the references of Chapter 2.

Annex II - Examples of practical implementation of surveys in the EU

France

Over time, the researchers have set up investigations on road safety and mobility. For example, studies on the risk exposure for specific road users were conducted on motorcyclists, on young drivers (MARC survey) or in a specific region (Île de France).
The data of these investigations are recorded in databases. We will quote those which deal with the RED issue.

The National Passenger Travel Survey

The next large national survey on "transport" will take place in 2007. The fourth survey, entitled "transport & communications", was led in 1993-1994. Under a slightly different title, it follows the same line of the other "Transport" investigations carried out since the sixties (the previous surveys being 1966-67, 1973-1974 and 1981-1982). The main objective is to describe all the trips made by a panel of households living in France, whatever their purpose, mode of transport, length, period in the year or time of day. It aims to improve the knowledge of the mobility of the households, their behaviour and their use of public transport and private transport, for example by analysing the changes since the last survey. It includes questions about the possibilities of access to public transport, the ownership of private vehicles by each household and the individual means of transport available to the households as in unexcavated description of the vehicle stock (3 types of vehicles are listed: private cars, light commercial vehicles and carriers). One of the strong points of this investigation is to collect all the means of transports and to provide an overview of behaviours and practices of French households. The aspect "communications" is treated in terms of contacts carried out by the households in situation of mobility and terms of equipment in means of telecommunications. A randomly chosen household member (over 5 years old) is asked to record all trips made the day before and the previous weekend. The sample of more than 14,150 respondents (among a sample of 20,000 residences selected) was spread over 8 waves from May 1993 to April 1994 in order to neutralise the seasonal effects.

The Household Living Conditions Survey

A device of permanent investigations into "Household living conditions" is annually set up, since January 1996. It aims to study the annual trend of social indicators harmonized in the whole of the European Union . All the indicators are divided into three groups, each one being the object of an annual investigation in January, May and October. These surveys are carried out with a sample of about 8,000 households. These investigations consist of a "fixed part" known as "social indicators" and a "variable part" intended to punctually apprehend a particular social problem related to the living conditions. Thus, the variable part of the 1998 investigation is devoted to the "environmental practices and sensitivity to the problems of the environment"; these questions are related to the means of transport used by the households for their various displacements (work, leisure etc.) The 1999 investigation includes questions related to the "departures on holiday".

The survey PARC Auto: INRETS-SOFRES

It is a study on the automobile equipment of the households, and on the use of their cars. It comprises socio-economic data on the households and the use of their automobile park; communal census data from 1982 and 1990; characteristics of the vehicles; price of the updated fuels each year. The data is currently available from 1984. It is an annual investigation, based on the panel of SOFRES. This panel of 10,000 households is defined from the territorial database of INSEE and from the technical file of the APSAD vehicle database.

The MARC Survey

The MARC Survey on mobility, attitudes, risk and behaviour of the young drivers started in 2003. The questionnaires have been devised in order to adjust for three approaches to road risk; of sociological, psychological and economic type. The first results of data analysis are now available.

The road passenger Survey (TRV Survey)

This survey aims to measure the volume of collective passenger road transport. It is led every year, permanently (230 questionnaires being sent every week of the year), since 1978.

The road freight transport Survey (TRM Survey) and the Shipper's Survey

The TRM survey aims to measure the volume of road freight transport, and has been carried out every year since 1965. It is the only survey which produces regular information on freight transport, which is often excluded in the field of the travel surveys.
Two shipper surveys have been conducted in 1988, and in 2004. The first results of the 2004 survey should be available at the end of the year: around 3000 shippers have been questioned and about 9500 shipments have been studied.

Local surveys

Local surveys are regularly carried out in France, whether at town level (The Household travel Surveys - methodology of CERTU), or at regional level (Île de France Global transport Survey).

Portugal

The National Statistical Institute (INE) is responsible in Portugal for the population census. This procedure exists since 1864, and for some time has been made once every ten years.

The last census was made in 2000, and the individual dwelling questionnaires distributed included some questions related to travel habits:

- Place of work or study (in the same zone of home; in another zone within the county, in another county, in another country).
- Time spent on the journey to work or to the place of study (none, less than 15 minutes, from 16 to 30 minutes, from 31 to 60 minutes, from 61 to 90 minutes, more than 90 minutes).
- Main transport mode used in the journey to/from work – on foot, by bus, by tram or subway, by train, by school or company bus, private car (as a driver or as a passenger), by motorcycle, by bicycle, or by other mode).

Information thus gathered may be useful for some safety studies in urban and metropolitan areas, as far as the commuting travel is concerned. However, other types of travel (such as interurban and leisure/business related travel) are not covered, which is a serious barrier to more comprehensive studies.

One obvious issue that is introduced in Portugal is that population figures and this type of travel information will be correlated. It is the same person in the population figure that is also in the travel survey. For example when a risk for a certain population group in a certain area appears to be high, it may turn out that that population group in that area had a remarkably small amount of travel kilometres. It may also turn out that there were only a small number of people in that population group in that area. Care must be taken assuming that the presence of a small number of people according to the survey explains the small amount of travel kilometres, as it can easily be the same sampling error.

Greece

In the framework of a research project in progress, a travel survey was carried out in 2004 by NTUA for the Ministry of Transport and Communications. The project "Accident risk investigation of drivers with high participation in road accidents" aims to examine the accident risk of drivers with frequent accident involvement in Greece, focusing on two-wheelers and young drivers. One of the main objectives of the project was the collection of detailed exposure data (vehicle-kilometres of travel) for the above driver categories, through a nationwide travel survey.

The survey was carried out in 2004 by Computer Assisted Telephone Interviews (CATI) based on an extensive questionnaire. The survey sample was a simple random sample (SRS), quoted by geographical area in relation to the population density of each region. The geographical classification adopted was the following:

- Athens
- Thessaloniki
- Large urban areas
- Other urban areas
- Suburban areas
- Rural areas

Survey participants were selected among respondents on the basis of certain characteristics. In particular, only individuals of at least 16 years old (minimum age for moped license), who are moped, motorcycle or passenger car license holders and who are active moped, motorcycle or passenger car drivers were interviewed. Active drivers were defined as drivers who have used their vehicle in the last 12 months. The above constraints in survey participants, in addition to the relatively low survey response, resulted in a sample of 2.500 fully exploitable questionnaires out of 6.000 contacts in total.

The following definitions for the parameters examined were considered:

- Vehicle type: mopeds, motorcycles, passenger cars.
- Days: typical weekdays, weekends (special events excluded).
- Lighting conditions: day (during daylight), night.
- Trips: urban (<100 km distance), interurban (>100 km distance).
- Road type: motorway (yes/no).

In particular, the questionnaire included questions on:

- Driver characteristics such as age, gender, nationality, driving license category and date of issue, number and type of vehicles used
- Vehicle characteristics such as vehicle type and make, vehicle age, cubic capacity, total number of kilometres traveled (from vehicle counter)
- Travel characteristics such as number and distance of interurban trips in the last 3 months, kilometres traveled the previous day, kilometres traveled the previous weekend, breakdown by area type (inside/outside urban area), road type (motorway yes/no), by lighting conditions (daylight/night), by use of seat belt/helmet.
- Other characteristics such as number of accidents with fatalities/material damage

It should be noted that the vehicle-kilometres collected are fully disaggregated in relation to driver and vehicle categories (i.e. vehicle-kilometres per gender and age and vehicle type and vehicle age and so on), whereas vehicle-kilometres are not fully disaggregate in relation to road and trip categories (i.e. vehicle-kilometres per road type or per lighting conditions or per area type and so on). Therefore, each classification of vehicle-kilometres can include more than one driver or vehicle variables, but only one road or travel variable.

The main limitations in the exploitation of the survey rise from the examination of car and two-wheelers drivers only and the relatively small sample. However, in the framework of the above project, the results of this survey have allowed for the calculation of reliable and analytical accident risk rates for the first time in Greece.

Norway

The Institute of Transport Economics (TØI) conducts National Travel Surveys every 4th year. The survey runs throughout a whole year with telephone interviews to a sub sample every day. People are asked about all the travels they made the previous day. In this way, the national travel survey gives a representative total of all travels conducted throughout a year.

TØI has calculated exposure and risk figures based on each of the National Travel Surveys, including figures for pedestrians and bicyclists. These surveys are for the moment the only data set available in order to estimate exposure and risk figures for these road users. In addition, it is for the moment also the only data set available to estimate risk (killed/injured per person kilometre) for different age/gender groups.

There are, however, limits to these data. Only private travels are in principle covered by the survey, as in most travel surveys, but in the last survey professional drivers in the sample were also asked how many kilometres they drove as professional drivers during the register day.

Another problem is that it is impossible to calculate reliable exposure data for transport means that are rarely used like motorcycles. The data set is normally not large enough to estimate risk by age and gender for other road users than car drivers and pedestrians. A specific problem connected to car passengers has been encountered several times; for some reason car passengers underreport their amount of travel in the travel surveys.

Finally, the survey also only covers people over twelve years, so it is not possible to say anything about exposure and risk for children by use of these data. This last problem is now being resolved by a separate travel study for children that will take place together with the ordinary National Travel Survey in 2005.

Netherlands

Since 1978, the Dutch national statistics bureau (CBS) has been carrying out the National Travel Survey (NTS). The purpose of the survey is to describe the personal travelling behaviour of the Dutch population.

For each trip, the following is registered:
- place of origin,
- place of destination,
- time of departure,
- time of arrival,
- modes of transport,
- trip purpose.

Among other things, age group, sex, and province are also known.
The vehicle numbers for 'slow' traffic (mopeds, light-mopeds, and bicycles) are derived from the NTS.

Data collection since 1999

In 1999, the research method was altered drastically. The New NTS is based on the German Neu KONTIV Design (NKD). It is designed as a simple, written survey, with a telephone conversation to motivate the respondents, and any follow-up surveys for more detailed questions per sub-group.
One of the biggest differences with the old NTS is that diaries are sent first, after which telephone calls are made to motivate people to fill in the questionnaire and trip diary.
If a household has not yet filled in/returned the questionnaires for the first fill-in day, a telephone call is made exactly a week later. Telephone and written reminders are carried out to a maximum of 4 times.
Households without a telephone receive the questionnaires, but are no longer 'accompanied' by the interviewers. They receive reminder letters after one and after two weeks if the CBS has not received their questionnaires. At the third reminder (after three weeks), new questionnaires are sent. Finally, after four weeks, if necessary, a fourth reminder is sent.

Data Processing

In the data processing, the category 'rest' and 'unknown' are not used. The coder tries to make an estimate, or rings the respondent.

Sample: nature and size

	1985-1993	1994	1995	1996	1997	1998	1999	2000
Approached	16		70				90.5	92.3
Households	10	34.5	68.4	62.8	60.1	58.0	63.3	64.2
Individuals	25	82.8	167.9	152.5	144.0	138.0	146.3	146.5
Trips	78	304	611	554	525	498	465	455

Table: Sample sizes and trips (numbers x 1000).

In the New NTS, the sample is drawn from the Municipal Basic Administration (GBA). Because of this, those who live at the drawn address receive a personalised mailing, which leads to a better response.
Households without a telephone or with an unlisted number are now also approached.
In 2000, 92,261 households were approached. Eventually, 64,240 household questionnaires and 146,528 individual questionnaires were sent out.

Reliability, accuracy, and completeness

Many mistakes can be made when filling in a trip diary. Trips are forgotten (too short, unimportant, before-and-after transport in public transport trips), they are deliberately omitted (nobody else is allowed to know), the times of day are rounded off, the distance travelled is estimated incorrectly, etc. In 1978, SWOV checked the distances in the questionnaires. This resulted in correction factors of –8% for car trips and –20% for bicycle trips. The CBS uses these factors to calculate the total distances travelled in the Netherlands.

Weighing and extrapolation

Weighing is compensating for the under- and over-representation of certain groups, e.g. degree of urbanisation, age, and car possession by registration year. The trips are also weighed by day of the year.

Sampling error

The Dutch National Travel Survey uses households per day as sampling units. This has a consequence that, in order to assess the sampling error of the population estimate of the number of kilometres in a certain travel mode, as a starting point the variance over the households has to be computed. Households that did not have any travel kilometres with a certain travel mode that day should then be treated accondingly: they had 0 km.

This effectively means that travel modes that are relatively rare (and therefore are not often used, resulting in 0 km, or, like motorcycles, when they are used, they are used in a large number of kilometres) may have a substantially larger sampling error (within households variance) than modes that are used approximately equally as much by ordinary households. The common sense argument for this is that if by chance one motorcycle driver more is selected in place of someone else, it has a larger effect on the total number of kilometres as mostly the expected kilometres on a motorcycle of the alternative selected would be small.

Such a result extends to for instance holiday trips. People rarely take holiday trips, but if the do, they are most often quite long.

The following table (table 1) gives the relative error in the in the travel kilometres broken down in travel mode and age. Unfortunately, the white figures are worst. Some of them actually exceed a relative margin of 100%.
NB: a 'sloped' refers to what used to be called the 'light-moped' (and 'slopedists' are what used to be called 'light-mopedists' or 'riders of light-mopeds').

Mode of Transp. → Age group	Pedestrian	Bicycle	Sloped	Moped	Motor/ scooter	Car (driver)	Car (passenger)	Bus/Tram/ Metro/Train	Rest	Total
0 – 5y	231	480	2	1	0	0	6636	182	28	7560
6 – 11y	239	817	2	4	8	0	5249	367	25	6711
12 – 14y	82	1296	0	1	1	0	1832	588	23	3823
15 – 17y	94	1132	10	270	26	0	1820	1476	45	4873
18 – 19y	73	505	7	91	8	1000	1307	2594	28	5613
20 – 24y	162	764	11	59	193	4706	3097	4168	75	13235
25 – 29y	219	786	2	59	161	8765	3294	2468	103	15857
30 – 34 y	249	915	2	52	150	12501	3537	1598	83	19088
35 – 39y	259	1091	33	50	109	12863	3179	1840	99	19524
40 – 44y	247	1120	11	56	202	11469	2955	1357	99	17517
45 – 49y	250	1016	18	33	233	10669	2701	1478	86	16484
50 – 54y	236	1077	12	20	67	9428	3042	1200	106	15188
55 – 59y	215	872	16	15	39	7746	2934	923	101	12863
60 – 64y	212	739	16	10	4	4654	1946	748	29	8358
65 – 69y	179	525	5	7	11	2694	1432	616	37	5506
70 – 74y	127	379	1	2	5	1720	1189	398	51	3871
75 – 79y	91	200	6	6	1	1052	639	280	7	2281
80y +	59	93	7	2	4	478	628	234	6	1511
Total	3226	13807	161	738	1222	89745	47417	22515	1029	179861

Table : Kilometres travelled in NTS 2003 by age group and modal split.

Margin <10%	Margin < 20%	Margin >=20%

Denmark

Person kilometres estimates in Denmark are based on a survey asking people how they travelled the day before (Transportvaneundersøgelsen – TU).

TU is based upon telephone interview of people between 16 and 80 years of age. They are asked how they travelled the day before.

The problem with TU is, that it is a limited portion of the Danish population, that is asked and only people with a telephone. Also, no information is obtained about people younger than 16. Walking distance and the use of public transport however is available.

TU is conducted regularly and gives information about age and traffic mode. Yearly figures can be seen on www.vejdirektoratet.dk or www.statistikbanken.dk.

Annex III - Survey errors

Measurement error

By nature, surveys do not produce 100% accurate figures for the population. Always an estimate is obtained that is not completely head on the true population figure. Ideally, the estimate would be right on average. However, when setting up a survey, in practice choices have to be made that will render the estimates to be on average a little lower than the true population figure, whereas other choices may have an opposite effect. The resultant effect is mostly unclear, at least numerically, as otherwise the surveys would be corrected for them. Errors tending the result in one specific direction can be called biases. They occur both on the respondent side and the statistical institute side (survey side). The respondent side error is considered to be the measurement error - in the case the interviewed respond -, whereas the survey side error is considered to be a sampling error.

Respondent side (measurement) errors

Many kinds of errors – as seen from the position of the survey -- are possible on the respondent side of a survey. A very important category is probably the error due to misunderstanding the question(s) and/or the procedures. This is likely the reason why interactive surveys are so often used, most of which telephone based. In such a setup the respondents are guided though the process of completing the survey if not completely released from that task. Additionally, respondents can be helped not to make common errors or omissions, which often include:

- Respondents ignoring 'trivial' transport. Respondents tend to recall the major trips only. The minor ones are deemed insignificant and therefore forgotten. For instance, if all results from the Dutch National Travel Survey could be fully trusted, many people in the Netherlands actually live at train stations, and go to work at other train stations: respondents forgot that they for example took the bus, tram or bicycle, or went by foot to the station and did similar so at the station of arrival. Respondents tend to recall that they went 'by train' in such cases. Also return trips may be forgotten (in some surveys return tips as considered part of the original trip). A big advantage of interactive surveys is that such errors may be avoided because the interviewer guides the respondent through the questions, for instance by asking 'how did you get to the station?'.
- People apparently have problems estimating time and or distance. This is reported in the Norway national report, restricted to car passengers (including drivers?) and travel amount. The Dutch National Travel Survey uses both data as well as information on starting point and end point of trips to verify distances travelled and or time involved. A physical interviewer may know distance tables and may help the respondent.
- Sometimes people would not complete a survey because they did not travel the given day. Alternatively, they may complete the questionnaire for a day they in fact did travel. Such errors can also be avoided by using an interactive survey too.
- For information based on trips that took place longer in the past, problems memorizing trips may arise as huge trips are often better remembered than smaller ones.

Missing data

Another particular problem is missing data (sometimes called item non response). A common problem, particularly with paper questionnaires is that respondents do not complete the questionnaire for all questions. For this reason it is often called a non-response error. There is two potential reasons for this to happen.

- The respondent does not know the answer (or does not want to give it)

- The respondent for various reasons unintentionally did not answer the (part of the) question.

In many cases, respondent side errors are encountered by feedback from the interviewers, in order to correct missing data or apparent errors. In some cases a scheme is implemented that corrects for errors. In case of missing data, it is often called imputation. Some practices in this respect used for RED in the EU are discussed below.

Survey side measurement error

The predominant survey side error is due to the sample (for instance the actual set of people interviewed) not being representative for the population the survey is intended to be used for. In some cases this may be caused by the fact that the survey is or cannot be set up properly for the target population. More often the problem will be that the survey sample is representative for the target population, it is not representative for the population it is used for.

A trivial theoretical example; you have somehow obtained a sample of clients from an insurance company. It turns out that everybody has a motorcycle! It turns out that this insurance company specializes in motorcycle insurances. Also other information, for example age or gender, private car possession will be difficult to generalize to the national population from such a survey.

Another very common type of error is introduced with the way the survey is conducted. This is also a representativeness problem, and it is very common. Which ever way the statistical institute chooses to communicate with the respondents, some potential respondents are not able to respond using that way of communication:

- Some people do not have a (fixed) telephone line
- Some people do not have a fixed address
- Some people do not have internet
- Some people do not have e-mail
- Some people do not want to respond to surveys
- Some people are away from home so much that they can hardly be reached via (fixed) telephone lines
- Some people leave the house so seldom that they can hardly be reached by means of roadside surveys.

If all persons either stay at home all day or travel all day, a telephone survey would reveal that nobody ever travels, while a roadside survey would reveal that everybody travels all day.

The state of the art in surveys in the EU appears to be that telephone surveys are the rule, but interview-based surveys are also found. What is crucial however is the interaction between the interviewer and the respondent. Even in the paper based questionnaire used in the Dutch National Travel Survey, respondents are contacted when errors or implausible responses are encountered or no response is given. Then, effectively, this survey too becomes a telephone based survey.

Sampling error

As described above, an important property of a survey is that a selection out of a population of individuals is asked for information rather than all possible individuals in the population. This means that if another selection of individuals is asked for the same information, this information is to some extent likely to be different from the information from the original selection of individuals. In terms of travel kilometres, the average kilometres driven may vary with the selection of individuals actually interviewed.

In most cases, the selection is done at random, meaning that all individuals in the population have had a prescribed probability of being included in the survey sample (usually at least assumed to be without replacement). By some chance process (being the driver of 10th vehicle passing since the previous interviewed, being the one actually picking up the phone) some individuals were actually chosen. This contrasts the alternative where a usually small selection of individuals is chosen which are supposed to be representative for the population[1].

The important difference between the two approaches with respect to sampling error is that in the latter approach, (almost) exactly the same selection should be chosen whereas in the first method an (almost) entirely different selection will be chosen.

In practice, most surveys in a sense mix both approaches in that samples are taken from selected subpopulations. This approach is called stratified sampling. This way it is ensured that sufficient information is available for the selected subpopulations, yet the sample is useful as a survey sample for the entire population. This approach is used for instance in the Dutch National Travel Survey as well as the NTUA survey in Greece where it was seen to that a sufficient number of individuals from specific geographic areas (the main cities, the main urban areas, rural areas, regions of the country) were contacted. When roadside surveys are performed a similar procedure is used as usually a well chosen set of locations is used were the actual interview will take place.

The subject of the remaining of this section is how to asses the consequences of this random approach. In a sense this is also how it is determined how much a sufficient number of individuals is for specific purposes.

The following statistical details are taken from "W.G. Cochran, Sampling Techniques, second edition, Wiley, 1963, third edition, 1977". This book, as many similar books will, describes details on how estimates are obtained from the samples for simple figures as the total sum (f.i. travel kilometres) as well as ratios (f.i. the percentage of car owners that also own a motorcycle, the ratio of two quantities that both vary by respondent, as opposed to a mean, where the denominator is fixed). Cochran devoted an entire chapter to Ratio Estimates (Chapter 6). This chapter is too involved for this presentation but when ratios are to be estimated, the chapter of one in a similar book is strongly suggested reading.

One issue of particular interest is what Norway states as "Another problem is that it is impossible to calculate reliable exposure data for transport means that are rarely used like motorcycles. The data set is normally not large enough to estimate risk by age and gender for other road users than car drivers and pedestrians." A similar conclusion can be drawn from Dutch data. Additionally, information of the sample error is available for such cases.

Non-response error

Cochran (1963) defines non-response as the failure to measure some of the units in the selected sample. In our cases, predominantly surveys directed at the public, this failure can be further specified as: non-coverage, not at-home, unable to answer, the "hard-core" (again Cochran, 1963). These definitions are with respect to 'units', mostly persons in road safety surveys.

Different types of non-responses may occur:

- Unit non-response refers to the failure of a unit (a household or an individual) in the sample frame to participate in the survey.

- Item non-response refers to the failure to obtain complete information from a participating unit.

There are two main ways for to cope with non-response:
- Weighting, which consists of an expansion of the respondent's weight, is commonly used for total (unit) non-response.
- Imputation, which replaces a non-response by the response of a respondent, for item non-response.

As an example, it is quite likely that item non response is treated similarly to obvious response errors. All non-interactive surveys will encounter the occasional not-yet-born car driver, because the respondent filled in the current year instead of his or her actual birth year. Quite often also the age of the respondents is asked. In such cases, the maintainer of the survey may impute the birth date of the respondent: deriving the birth date from the age. The opposite also happens, when the answer to the age question is missing, it may be imputed from birth date and the day the questionnaire is completed, issued or received by the maintainer of the survey.

It is important to know about the presence of imputed variables in your data. For instance in the Netherlands, if the number of bicycles owned by the family is larger than the number of respondents from that family, it is assumed that all respondents over the age of four owns a bicycle. (which is not that strange in the Netherlands). However, in the same survey, roughly speaking, if a scooter owner is aged over 34, it is assumed that the scooter is of the slow-moped type, whereas if the owner is under 35, it is assumed the scooter is of the moped type. It seems it is best to be aware of such potential issues. This issue could also be discussed in the next section.

Survey error handling

In practice, it is not feasible to correct for all types of survey errors. The measurement error, for obvious reasons, can not be measured easily. The sampling error, the extent of which can be measured, can only be reduced by enlarging the sample's size. At last, as has just been seen before, the non-response error can on the contrary be corrected for, and it actually is the case for all surveys.

In the case of the national passenger travel surveys, for instance, the rate of 18,3% of nonresponse was found in the description of the daily trips made by persons living in France as well as their use of public and private transport means which is a rather satisfying result.

Correcting for nonresponse: the daily trips in the French NPTS

The example of correcting for non responses for daily trips in the French NPTS of 1993-1994, mixes both methods mentioned, weighting and imputations.

A two stage technique was used to cope with total non-response: first through a post-stratification according to the households' characteristics explaining response behavior, and second in correcting for sampling error by calibration on margins. It was demonstrated that, when data is collected by face to face interview, the nonresponse mechanism can be considered as unconfounded (inside small homogeneous population groups, travel behaviour of non respondents does not differ significantly from the behaviour of respondents).

Usually, weighting is used to cope with unit non response, while imputation is implemented to correct for item non response. In this example, trips weights were also modified to

compensate for memory effects. It increased average mobility according to omitted trips and did not seem to introduce an overly large bias in trip distributions.

Documentation

Probably the best source of information on survey errors is the documentation maintained for the survey. The respective survey maintainers publish documentation from which details on reliability can be found.

Annex IV - Examples of practical implementation of traffic counts systems in the EU

France

In France, apart from annual censuses which allow for estimating, in an exhaustive but heavy manner the annual aggregate traffic volume on the main road network (A-level roads, and motorways), a sample design has been retained and a special procedure applied, since January 1986, for estimating rapidly the aggregate traffic volume on a monthly basis, on this main road network. Monthly indexes are thus available with about one month and a half delay (Note d'information n° 7 du SETRA, mars 1986, Ministère de l'Equipement).

A special attention is now paid to the methodology used for deriving vehicle-kilometres from vehicle counts, once the vehicle counts have first been measured by means of the SIREDO-system.

The SIREDO-system: A panel of 3000 road network sensors, placed on both main roads and motorways, is used to measure the traffic volume on the interurban network. 1500 traffic count points are collected by the SETRA in France, and 1500 other points are collected by toll motorway companies and added to this database. The sensors measure the numbers of vehicles passing per time unit, but also measure their speed and the percentage of vehicles driving faster than the authorised limit; aggregate information is systematically gathered, but more detailed information can also be available at a local level.

Weights are determined for each of about 250 permanent counting sites, in order to guaranty the representativeness of these counting sites (each weight is a sort of "fictive length" related to each counting site - which is to say for a road section of about 10 kilometres - and takes account for the road capacity, the traffic volume level, and the geographic situation of the counting site. The total number of vehicle-kilometres, on an aggregate network (whether A-level roads, toll motorways or public motorways), for a given month, is then calculated by summing, on all the road counting sites of the aggregate network, the weighted number of vehicle counts of each site, for this given month.

Special attention is, for evident reasons, paid to the consistance of the monthly indexes with the annual measure provided by the annual census.

Norway

Traffic counts in Norway are only done regularly on national and county roads, at specific points on the road net. Whether these points are representative for the total Norwegian road traffic is not known. Traffic counts may give good estimates of annual daily traffic (ADT) on stretches of roads, but there are practical problems involved in calculating vehicle kilometres from ADT. Such calculations are made, but because traffic is only counted regularly at a limited number of points, one has to add traffic counts from points where counting is done less often.

The result is that it is difficult to give precise estimates of total traffic volumes for one year based on the traffic counts. In addition, local roads are left out, and accordingly one will not get the full picture of road traffic by this method. In addition, traffic counts are not suitable to

distribute exposure according to age/gender groups, or car types (apart from “long” vs. “short”). Also mopeds and motorcycles are often left out because their weight is too low to be registered automatically (by sensors underneath the road surface).

The Public Roads Administration carries out the traffic counts on national and county roads. Normally they do not count bicyclists and pedestrians. Such counts are therefore only done locally in the municipalities and thus they vary greatly between districts. There is a system for regular bicycle counting in Oslo, which can be used to estimate whether bicycle travel is increased from one period to another, but which cannot give total bicycle exposure in Oslo.

Hungary

The system of national road traffic census

For the well-established implementation of the tasks relating to the official decision-making, designing, and research in road management and maintenance, it is indispensable to know the road traffic data, the traffic composition by vehicle categories, and the pavement loading effects caused by vehicles.

In Hungary, on the network of the national public roads, of course with continuous updating, the unified system of traffic censuses has been operated since 1927. Whereas the first counts have been considered rather as historical preliminaries, the yearly comprehensive national census of 1955-56, in its essentials, was carried out already according to the principles of today.

The two main elements of the traffic census, essentially based on one another, and correlated with each other are:

- continuous monitoring of the road traffic, and
- comprehensive traffic census enlarged with a program addressed to secondary stations.

Extensive traffic counts are cyclically repeated on the almost 30 000 km long national road network. Monitoring is integrated in the extensive traffic census.

The extent of the cyclic character of the extensive traffic census, in the course of years, has changed several times in conformity with the regrouping requirements of the labour force and equipment. Full-scale national counts being carried out in the same year on the whole national public road network took place every 5-year until 1980. Subsequently, the national census was carried out two times (between 1984 and 1986, as well between 1990 and 1992) in a 3-year-rotation system, when in each year the road network of 6 or 7 counties was involved.

From 1995 on, unlike in the previous years, the national road traffic census has been made in a so-called “rolling” system. With the introduction of the new traffic counting system, a comprehensive traffic census is carried out in subsequent 5-year cycles. Traffic counts are annually made on about one-fifth of each county’s road network. Besides the data of the year concerned, the annual average daily traffic data are formed using the traffic changing multipliers calculated for the four-fifth of the country on the basis of the data of the every year operating stations of the monitoring network. The introduction of the “rolling” system made feasible for the profession to be in command of the relevant traffic data and the national summaries relating to practically the whole national road network.

The following table offers a brief description on different basic data types, the present system of the data flow, and those responsible and involved in execution.

Types of the basic data and the present system of dataflow

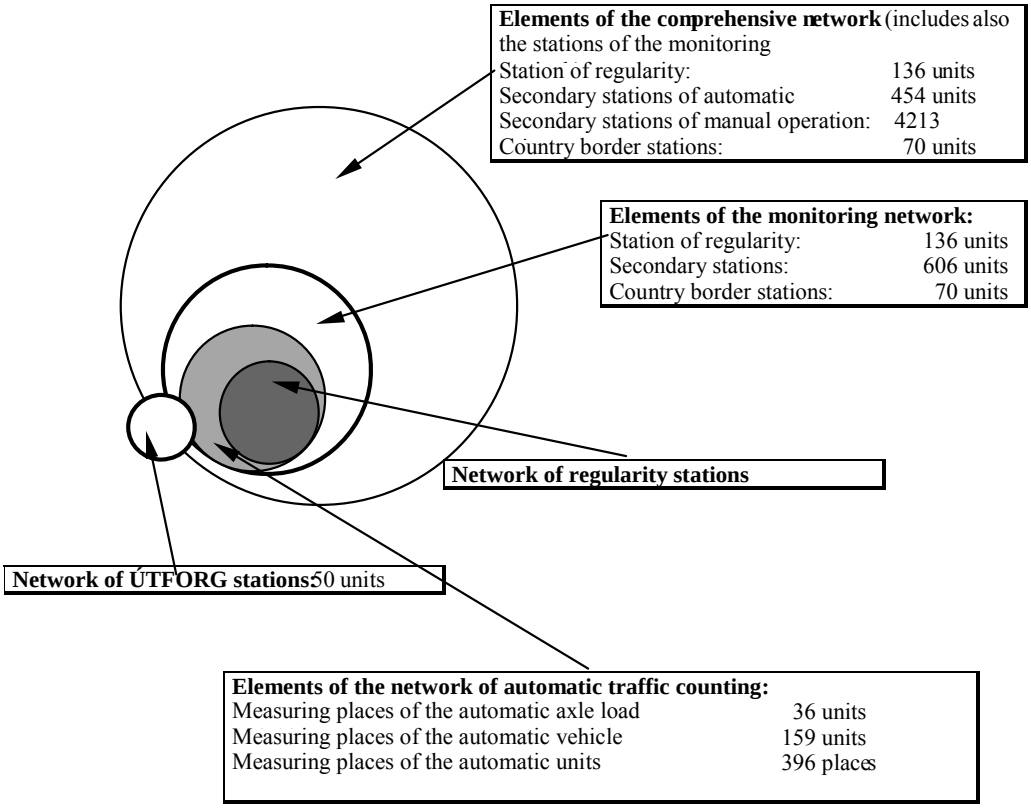
MANUAL DATA RECORDING Executed by: road administrator, or its client	AUTOMATIC DATA RECORDING (operation of instruments) Executed by:
Section 1.01 DATA RECORDING Executed by: road administrator or its client *Section 1.02 Possible modes: by directions (with special code), and by hour intervals* Summarised for cross section by hour intervals Summarised for cross section and day period (usually between 6 and 18 hours) File type: dBase, excel, text, paradox	MNL 2032 (unit counter) (Hungarian) road administrator, or its client ADR-2000 (vehicle classifier) (English) road administrator, or its client RAKTEL 8000 (vehicle classifier) (South-African) Hungarian vendor HESTIA (classifier+axle load weighing (French) ÁKMI MVOs, or Hungarian vendor QLD-6CX (vehicle classifier) (Slovenian) road administrator, or Hungarian vendor For processing the data of the national road traffic counts, the road administrator submits according to schedules the data of the systems built for continuous motorway monitoring (MARABU, MAESTRO, LANTASTIC). The possibility and the method to integrate the data provided by the stations ÚTFORG (WEISS automatic vehicle classifier) in the national traffic census is just being investigated.
DATA TRANSMISSION If the entrepreneur, in compliance with the agreement concluded with the road administrator, carries out the traffic census, then this is usually transmitted to the commissioner, and to ÁKMI (from 2002 to the central data processor) in the form of CD, or e-mail. In another case it is the road administrator's responsibility to transmit the checked data for central processing.	
DATAPROCESSING Entrepreneurs commissioned by the road administrators and for special tasks the ÁKMI Kht. also carry out preliminary processing. Central processing is implemented at the organisation commissioned by the ÁKMI Kht.	
PUBLICATION OF DATA Centrally processed data are included in annual publications and CD-s. Yearly ADT-s after processing are passed to the National Road Data Base, too.	

The National Office for Statistics supplies data concerning the country border stations for Statistics

The network of traffic counting stations

As it has been mentioned in the Introduction, continuous monitoring and comprehensive traffic census are the two integral and closely related elements of the road traffic census. Stations of the **monitoring network** are integral parts of the **network of the comprehensive counting** stations. The basis of the present network of the traffic counting stations (this concerns both the monitoring and the comprehensive network) was created in the early seventies by KTI with the involvement of the Road directorates and their experts. Of course, over a period of years, for several times, the network has been revised and modernised. However, the basic principle did not change according to which: if possible (and the relocation of the counting place is not justified), for the sake of comparability, the counting stations should remain in the same transverse profiles.

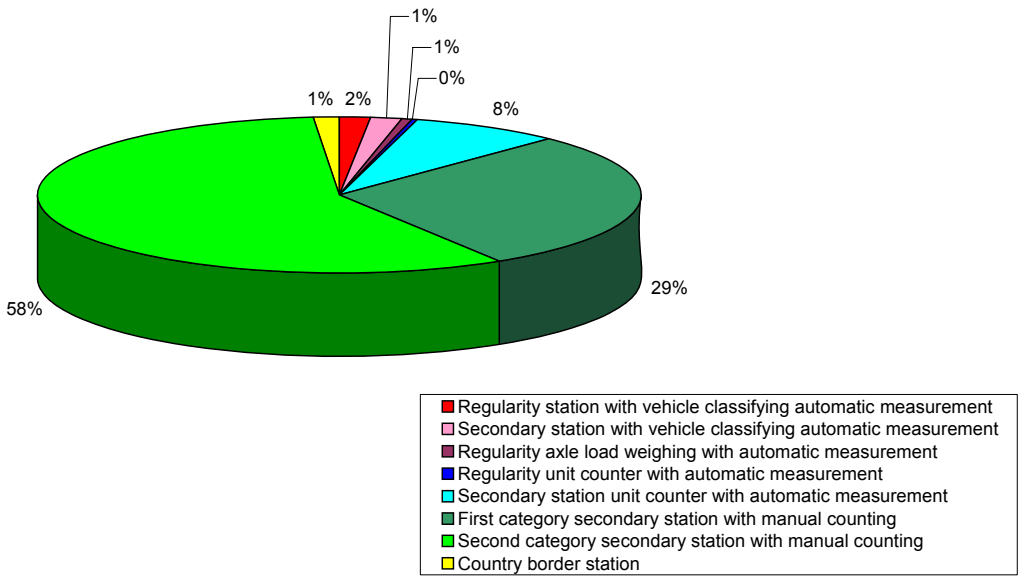
Until now only two networks have been mentioned, but the different counting stations are also the elements of smaller networks on the basis of their roles played in the traffic census. In the following, a figure is used in order to present the connection between the different networks, their relation to the comprehensive (full scale national) network, as well as their main characteristics.



Types and centrally determined programs of the traffic counting stations

The circle diagram below shows the types and distribution of the counting stations operating within the framework of the national road traffic census.

Distribution of the different stations on the national road network



Road Technical Regulation ÚT 2-1.109 specifies the minimum counting and measuring programs of the different traffic counting stations in the function of the type of the station and the traffic character of the road section.

	Prescribed program	
	Automatic	Manual
Regularity (or main)stations	12 weeks/year (hours peak station: 24 weeks/year)	12 days/year
Secondary stations	Pending on traffic character 2-4 weeks/year	Pending on traffic character 2-5 days/year

However, the implementation of the counting program prescribed in the Technical Regulation changes by years and counting places. In several cases only a portion of the prescribed counting programs is implemented, or only some part of the measurements can be used in data processing. Nevertheless, it occurs that in high priority cross-sections, a counting program of a larger order of magnitude than the prescribed one (in some cases, a through-year continuous) is scheduled and implemented.

The present system of data processing

The data, which during one year have been recorded manually and by automatic measurements at the traffic counting stations, are processed in several levels. ÁKMI Kht. (Technical and Information Services on National Roads), the counties and their clients are also engaged in carrying out the recording of the data of manual censuses and in certain preliminary processing tasks.

However, central processing is needed for

- the comparative analysis of the data (related to the traffic the counties' joining road sections, the time series of the previous years),
- carrying out different traffic analyses and the summarising of the national statements,
- the computerised presentation and publication of the tables of the documents of the traffic census with regard to the whole national road network.

The following tables are included in the yearbooks with the traffic census results:

- the AADT of the main road network
- the AADT of the connecting- and secondary roads
- AADT and traffic performance on the road categories of the public road system (country- and county-wide)
- distribution by intervals of the traffic volume of the public road network (country- and county-wide)
- monthly ADT (at regular stations and some secondary stations operating according to special program)
- peak-hour survey (at peak-hour stations and stations working with special program)
- share rate of traffic of motor vehicles with foreign registration at determined stations
- analysis of traffic development (at regular stations and the country border stations)

Application of automatic counters in the national road traffic census

On the national network of counting stations consisting of 4900 cross sections until 1970 only manual data recordings were carried out. The demand to exchange the high volume manual counting with automatic measurements emerged at that time already. In fact the automatization of the traffic counting began in mid-seventies in Hungary, which since then went through a significant qualitative change.

The table below gives a short chronological presentation of the different periods of the process of automatization.

Period	Automatic counters	Measuring capacity
1975-1986	Rubber hose	Axle counting
1984-1989	MK3 (English)	Vehicle categorisation (4 categories)
1988-until now	MNL 2032 (Hungarian)	Unit counting
1995-until now	ADR 2000 (English)	Vehicle categorisation, speed measurement
1995-until now	HESTIA (French)	Vehicle categorisation, speed measurement, axle weighing
1997-until now	RAKTEL 8000 (South-African)	Vehicle categorisation, speed measurement, axle weighing at 1 measuring station
1999-until now	QLD-6CX (Slovenian)	Vehicle categorisation, speed measurement
2001-until now	WEISS (ÚTFORG)	Vehicle categorisation, speed measurement

Greece

In Greece there is currently no national traffic counts system. Some attempts in the past have provided traffic volume data for the main interurban network, allowing for origin-destination and vehicle-kilometres estimates. However, no system is currently operational and consequently significant limitations in road safety analyses are encountered.

The System of Traffic Counts of the Ministry of Environment, Physical Planning and Public Works (up to 1989)

In the Ministry of Environment, Physical Planning and Public Works there are data on the Annual Average Daily Traffic (AADT) of the main interurban road network of Greece. This traffic data was collected each year from 1979 to 1989 by selected traffic counts stations according to the guidelines of the National Traffic Counts System described in Report No 17 of the Louis Berger study (Organization Study for Road Design and Maintenance, 1976).

The traffic counts were carried out by means of the standard methodology (FHWA, 1997) with both loops and observers on all the arms of all the main intersections of the national and main interurban road network, and on other selected locations of the national road network. The stations were divided in the following three categories: permanent stations, control stations and covering stations.

There were in total 8 permanent stations where traffic volume was continuously measured. The main purpose of these measurements was the estimation of traffic coefficients (for peak hours, special events etc.) to be applied on the counts obtained from the control stations of the particular area of the permanent station.

Additionally, there were in total 58 control stations throughout the country. The main purpose of these measurements was the estimation of the monthly traffic variation coefficients to be used in the estimation of the AADT.

Finally, in total 836 covering stations were operating on each intersection branch and a twelve-hour traffic composition measurement was obtained on typical weekdays from 7:00 a.m to19:00 p.m.

Unfortunately, since 1989 the system is no longer operating. The lack of traffic data is still the most important constraint in road safety analysis in Greece.

The National Origin - Destination Survey (1993)

In 1993 the last National Origin-Destination Survey was carried out for the Ministry of Environment, Physical Planning and Public Works. The survey was carried out on three distinct periods on 228 roadside stations. Each survey period included two measurements (morning and evening) of 10-12 hours each. Additionally, two 24-hour measurements were also carried out. The survey resulted in the calculation of the AADT on the basis of a process much similar to the one described above.

In particular, a traffic volumes database was developed, including fields such as: department, region, road code, road kilometer, station serial number, intersection's branch and vehicle type. In order to calculate the AADT, the traffic counts obtained from covering stations were weighted in relation to the monthly traffic variation obtained from control stations in the particular area. Unfortunately, since 1993 no traffic data are available in Greece.

Data from the National Highways' tolls

Some very useful traffic data could be extracted from the national toll stations' operating services of the Hellenic Highway Fund. This data could be fully exploited for the calculation of vehicle-kilometres of travel on the main interurban road network of Greece by road type, by vehicle type and by day/month/year.

These data are available on paper form and aggregated since the mid-eighties. Since 2002, a project on the creation of the related databases is under way, allowing for availability of the very useful traffic data. However, as the access to the database is restricted and no publication is available, these data remain unexploited.

Portugal

Traffic censuses are performed on the National Road Network in a systematic and regular way since 1955. Until 1990, traffic counts were made once every five years, using the method recommended by ECE-UN. Since 1995, a new method has been used; this method is basically an improvement of the ECE-UN method, to suit it to Portuguese conditions.

This new method is used only in the part of the National Road Network where the management of operation is not leased to private companies.

Basically, the system comprises 61 permanent automatic counting sections (365 days×24h), 162 main counting sections (14 days×24h), 86 main counting sections (14 days×16h) and 403 secondary counting sections (5 days×8h). Counts are made in all these sections once every five years; in the intermediate years, counts are made in the permanent sections and in one third of the main counting sections (in this way, traffic counts in main counting sections are made twice every five years).

The algorithm used for estimating the average annual daily traffic (AADT) is the one proposed by ECE-UN. Conversion factors are applied to AADT to calculate the average annual weekend traffic, the average annual night traffic, the average annual daylight traffic, the average summer daily traffic and the average winter daily traffic.

Vehicles are classified in eight classes: bicycles, mopeds, motorcycles, passenger cars, vans, buses, rigid trucks with two axles, rigid trucks with three or more axles, articulated trucks, and other vehicles.

Error in the AADT calculation is estimated to be less than 13%. In absolute numbers, it is less than ±15 vehicles for two wheeled vehicles; between ±350 and ±900 for passenger cars; and less than ±100 for the other vehicle classes.

Estimates of traffic volumes are made by running the traffic modelling software EMME2 on a simulated NRN.

From past experience, there is a minimum delay of one year between data collection and the availability of the results.

As mentioned before, the operation of some part of the NRN has been leased to private companies. These stretches of the NRN are mostly motorways. In some motorways, users pay a toll for travelling along the road; in others, it is the responsibility of the Portuguese State to pay a shadow toll, depending on the amount of traffic in the road.

In both types of these leased roads a sophisticated system for continuous (365×24 h) measuring of AADT in each road stretch is used.

Shadow tolls are paid using the eight vehicle classes of the rod administration's basic classification. Direct tolls, however, are paid by road users on the basis of the classification of their vehicles in one of five vehicle classes: motorcycles; cars; vans, buses and trucks with two axles; trucks with three axles; trucks with four or more axles.

As a result of the counting system used, very accurate estimates of traffic volumes are possible in leased roads, simply by summing up the traffic volumes in all leased stretches.

Denmark

Concerning traffic volume there are two different sources. For state and county roads the information can be obtained from the Road Databank (VIS). This information is based on stationary monitoring stations on the road. Information is also available for some municipality roads. The total amount of traffic has up to 2002 been estimated using a lot of different sources. Now the information is based upon the mandatory check of cars every second year. When coming to check the amount of kilometres travelled is registered and compared to the amount from last check. Knowing the type of vehicle and the total numbers of that type, it is possible to give an estimate of kilometres travelled by type of vehicle.

Using the two sources it is possible to divide traffic by administration and road category. Traffic volume on municipality roads is regarded as the difference between traffic amount from the Road Databank and the total traffic amount.

The traffic information from the Road Databank can be divided into traffic on motorway, semi-motorways and other roads. It is also possible to distinguish between inside and outside urban areas. For municipality roads the distinction between inside and outside urban area is not directly possible. Some estimation has been done, but it is not very reliable.

The information about traffic is also available for different vehicle types. In the case of division by vehicle type, there is some problem with the accuracy of the amount for mopeds.

Annex V - IDF data gathering questionnaire

OVERVIEW OF INTERNATIONAL DATA FILES (IDF)
FOR ROAD TRANSPORT STATISTICS

General Information

1. Name of the data file
2. Owner Organization
3. Address
4. Name of the data file responsible
5. Name of the interviewee
6. Telephone
7. Fax

Information about the organization

8. What is the profile of your organization?
9. Which are the official relations with the members of your organization?
10. Which is the objective of your data file?
11. Is the management of the data file carried out internally or externally?

Data file history

12. Which was the year of creation of your data file?
13. How many years did the IDF development phase lasted?
14. Which was the first year of the IDF operation?
15. Which was the year of the first publication?
16. How many system updates did you carry out up to today?
17. How many members did you had in the beginning?
18. How many members do you have today?

Data collection

19. For how many countries do you collect data?
20. Which are these countries?
21. For how many countries do you have available data?
22. Is your IDF dedicated to road transport statistics or it concerns a wider/narrower area?
23. Which are the data collected?
24. Do you collect data in aggregated or disaggregated form?
25. How often do you collect data?
26. In which form do you collect data (paper or electronic form)?
27. When does the earliest data transmission for the reference year take place?
28. When does the latest data transmission for the reference year take place?
29. How are your data sources obliged to provide you data?
30. How do you accelerate the data collection process?
31. Do you have any plans for improvement of the data collection process?

Data processing

32. How do you carry out the data quality control?
33. Do you use correction coefficients?
34. How do you develop these correction coefficients?
35. How do you consider the data quality of your system?

System specifications

36. How do you carry out the data input to your data file?
37. Are all data collected introduced in the data file?
38. Which kind of hardware do you use?
39. Which kind of software do you use?
40. Is the programming of the data file carried out internally or externally?
41. Do you have any plans for improvement of your system?

Publication

42. Do you issue a publication with road accident statistics?
43. Is your publication available to everybody?
44. How often is this publication issued?
45. Which is the reference period of your publication?
46. How long after the reference year is the related publication issued?
47. How much does the publication cost?
48. How many publications do you issue?
49. How many subscribers do you have?
50. How many publications you send by mail?
51. Which are the basic categories of publication "clients"?
52. Do you promote your publication and in which way?
53. Are the publications sales carried out internally or externally?
54. Do you have any plans for improvement of the publication?

External access to the data file

55. Which are the eligibility criteria for providing data in electronic form?
56. Do you provide data in electronic form to your data suppliers?
57. Do you provide data in electronic form to outside people?
58. Do you provide on-line access to your data suppliers?
59. Do you provide on-line access to outside people?
60. Do you have any plans for providing data in electronic form?

Overall comments

61. Which do you think are the main advantages of your data file?
62. Which do you think are the main deficiencies of your data file?
63. Which are the limits of your data file?
64. Do you have any plans for improvement of your data file?

Material to be provided

65. We will be grateful if you could supply us with the following documents (if any):

- Extract of the statistics publication
- List of data contained in the IDF (road accident and exposure data)
- List of definitions used for the IDF data
- Description of the system
- Any other useful document

Risk Exposure Data information

66. What type of exposure data is available in the IDF and in which format?

- Vehicle-kilometres and passenger-kilometres of travel (by road network type, by vehicle type, by vehicle age, by driver/passenger age and gender)
- Number of registered vehicles (by vehicle type and vehicle age)
- Number of driving licenses (by driver age and gender)
- Fuel consumption (by vehicle type)
- Population (by age and gender)

67. What is the availability of the data (number of countries and number of years)?
68. Which variables and values are used?
69. Which are the sources of the data in each country?
70. Are there any coefficients or mathematical formulas used for the estimation or correction of the data?
71. How are these coefficients/formulas developed?

Annex IV - IDF responses to the questionnaire

	EUROSTAT	CEMT	UNECE	IRTAD	IRF
Information about the organization					
8. What is the profile of your organization?	European Commission	International, non - profit	Inter – governmental organization (collects what the governments want). UNECE strives to foster sustainable economic growth among its 55 member countries. To that end UNECE provides a forum for communication among States, brokers. International legal instruments addressing trade, transport and the environment and supplies statistics and environmental analyses.		Non governmental, worldwide members, profitable, private - public sectors, committed to road development
9. Which are the official relations with the members of your organization?	Legal acts, Gentlemen	Voluntary basis			Some governmental and some private organized. Data comes mostly from government, Ministries
10. Which is the objective of your data file?	Provide EU high quality	Help policy of road safety, consultants for studies			Mainly as a reference tool for road professionals. Statistician – research tool. UN economic sector I world bank I institutions.
11. Is the management of the data file carried out internally or externally?	Mixed (contractors)	Internally	internally		Internally
Data file history					
12. Which was the year of creation of your data file?	Transport 1990	1997 – 1998.	1977 (1951) – (1954)		1964
13. How many years did the IDF development phase lasted?		Almost 6 months			1958 -1964
14. Which was the first year of the IDF operation?		1998			1964
15. Which was the year of the first publication?	Goods transport very - Overview mid-90's (01	1972 (hand written) - the one produced by the database: 2001 (including 1998 data)		Not applicable	1964
16. How many system updates did you carry out up to today?		2			Updates
17. How many members did you had in the beginning?	Following the number I	17		1990: 6 national coordinating institutes and 3 additional institutes	20
18. How many members do you have today?	Following the number I	43		55 50 member institutes	186 provided data for 2004

	EUROSTAT /ECMT	ECE-UN	UNECE	IRTAD	IRF
Data collection					
19. For how many countries do you collect data?	EU-25 + EFTA-4 (+broader).	50	55	29 countries	84
20. Which are these countries?		available on the website		All OECO countries (except Mexico) and Slovenia.	
21. For how many countries do you have available data?	Yes	48	55	29 countries	
22. Is your IDF dedicated to road transport statistics or it concerns a wider/narrower area?	Yes	Databases per activity, not per subject (road safety data may be maintained in different databases)	wider	Road accident data and relevant exposure data	Wider Taxations - road expenditure motor fuels
23. Which are the data collected?	Common questionnaire broad but not so deep. NUTS 2 Level aggregate data (regional). Road legal act: micro data on road goods transport (veh-km tor goods vehicles). Passenger mobility: unharmonized surveys from Member States (to be revitalized). Ad hoc data on modal split		Injuries; killed; accidents per thousand motor vehicles; killed per population; alcohol test results; accidents by location (motorways, inside / outside built up area), time of occurrence (month, day, week), road condition (dry, wet), light condition (daylight, twilight, darkness); by nature of accidents; - killed or injured by category of user (driver, passenger etc) and age group (0 - 9, 10 - 14, 15 - 17, 18 - 20, 21 - 24, 25 - 64, 65 <); - road traffic accidents involving one or more persons under the influence of alcohol; pedestrians - drivers under the influence of alcohol; road vehicle fleet; veh-kms traveled (estimated); population by age group (not from Transport Statistics - demography).		
24. Do you collect data in aggregated or disaggregated form?		Aggregated form.	Aggregate	Aggregated data only	Aggregated form into Access
25. How often do you collect data?	Yearly, except Road legal act (monthly)	Annually	Annually	Continuously	Annually
26. In which form do you collect data (paper or electronic form)?		Electronic	Common Questionnaire for Transport Statistics (Eurostat/ECMT/UNECE). Intend to make a common questionnaire for road accident statistics with ECMT.	Electronic form	Both (Questionnaires)
27. When does the earliest data transmission for the reference year take place?		Send questionnaire June, collect September - October		March for the annual accident data (earlier for the monthly data)	Summer previous year
28. When does the latest data transmission for the reference year take place?				Not applicable	January - February (March publication)
29. How are your data sources obliged to provide you data?		9	Moral obligation	The national coordinating institutes sign a Memorandum of Understanding	Good - will _provide CD-Rom to whoever provides data
30. How do you accelerate the data collection process?		Sending reminders			Need to hire personnel more efficient questions
31. Do you have any plans for improvement of the data collection process?	Constantly (several projects) e.g. communicate separate data sources	Yes, Web - applied Questionnaire	Web - questionnaire	Annual survey into the likely time of data release Continuous provision of tailored data questionnaire for individual data categories to the individual national data providing institutes at the time specific data are released	WHO _ accidents ; GTZ _ motor fuels -> better structure ; Working with World Bank ; -> revised Questions ; -> funding ; -> more personnel

	EUROSTAT /ECMT	ECE-UN	UNECE	IRTAD	IRF
Data processing					
32. How do you carry out the data quality control?	Every time (sources)	Internally: checking totals.Comparison with other data files.	No, only obvious mistakes	Checking of the data for mathematical correctness and recording of the definitions; if necessary, follow up with data providers and revision of data (especially for new member countries)	No validation - Use of official data ; the countries send corrections
33. Do you use correction coefficients?	No standard correction coefficients. Legal acts impose to the Member States to comply with the definitions (* CARE)	No. These are applied by countries.		Input of data (including adaptation of correction factors and database extensions)	No. Each country with differences are noted as footnotes in publication
34. How do you develop these correction coefficients?				Death within 30 days as a result of the accident. All fatality figures in IRTAD have been corrected to conform to the standard definition. Conforming to the Convention of Road Traffic, Vienna, 1968. In the IRTAD guideline of 1989 the following international correction factors are mentioned:Internationally, fatalities are defined as "death within 30 days":24 hours:1.30; 3days:1.15; 6days:1.09; 7days: 1.08.The correction factor might be changed by the national data provider. For example, this was the case for France where a new correction factor (5.7%) was introduced in 1993, based on extensive national research.	
35. How do you consider the data quality of your system?			i) We are happy to receive data Basically ok, based on the fact that the countries have to comply with the definitions	ii) High - However, there are some countries (e.g. Turkey) where we know that there are problems	Need to be further improved

	EUROSTAT	CEMT	UNECE	IRTAD	IRF
System specifications					
36. How do you carry out the data input to your data file?		Automatically processed after validation.		Manually	Manually in Access
37. Are all data collected introduced in the data file?		All	Yes	Yes (exception: during the initial phase of the collection of new variables, the data are not released)	All
38. Which kind of hardware do you use?			Only aggregate data		PC
39. Which kind of software do you use?		Microsoft Access		PHP and Oracle, Excel	Windows - office software - Access
40. Is the programming of the data file carried out internally or externally?		Externally - consultant		Internally	Internally
41. Do you have any plans for improvement of your system?		Web - applied database (downloading with queries)		Yes	More efficient PC's Pentium 4
Publication					
42. Do you issue a publication with road accident statistics?		0ne	Statistics of Road Traffic Accidents in Europe and North America	Annual flyer with overview of selected road accident and traffic data	Yes. 'World Road Statistics"
43. Is your publication available to everybody?		CD freely - publication, hardcopy on charge	Yes	Yes	Yes. sold - CD-Rom; Hardcopy ; PDF file
44. How often is this publication issued?		Annually	Annual bulletin	Once a year	Annually
45. Which is the reference period of your publication?		One year behind	1992 - 2002	Yearly	Every five years - for 2005 two years
46. How long after the reference year is the related publication issued?		Minimum of 12 months	2004 (2 years later)		One year later
47. How much does the publication cost?		27€ - 35 €	available on-line	Free	Hardcopy + CD - Rom _ 400 CF CD - Rom _ 300 CF (in MS Excel) PDF _ 300 CF
48. How many publications do you issue?			Transport statistics		
49. How many subscribers do you have?				No fixed subscribers	Standing order 150 ; Others?? Total: 300 for 2003
50. How many publications you send by mail?		One (CD) to 500 people approximately	Free, on-line access	Also available on public web site	Ali
51. Which are the basic categories of publication "clients"?		Governmental, Statistic offices, Research institutes, consultants	Agents, booksellers, universities, libraries, institutes		Academic Institutions ; Bookstores Researchers ; UN I Economic CIA ; World Bank
52. Do you promote your publication and in which way?		OECD / marketing - (newsletters)			Since it is old, it is well- known. But needs further improvement. Intent to hire a person to promote sales
53. Are the publications sales carried out internally or externally?		Externally (OECD)			Both (through bookstores)
54. Do you have any plans for improvement of the publication?		Yes, improving content - easily available through web.	Standard Publication (Wanted to reduce the number of Tables, according to data availability but the countries were reluctant). Motor Traffic Census (5-year), used for spatial planning. The issue for 2005 will be available in 2007.		Maybe more topics or analyses later. Maybe it will be published in India (Cheaper)

	EUROSTAT /ECMT	ECE-UN	UNECE	IRTAD	IRF
External access to the data file					
55. Which are the eligibility criteria for providing data in electronic form?	On-line system. All aggregate data are free, everything (except confidential)	Freely.			Provide everybody a part from libraries + bookstores
56. Do you provide data in electronic form to your data suppliers?		Yes.	Not send, we put it on the internet electronicaly	Excel spreadsheets	Yes if they pay - ask
57. Do you provide data in electronic form to outside people?		Yes.	No	Yes	Yes if they pay - ask
58. Do you provide on-line access to your data suppliers?		Not yet - working on that - End of March	No	Yes, if they are IRTAD member institutes	No
59. Do you provide on-line access to outside people?		No restriction - on demand it will be possible	No	Yes, if they are IRTAD member institutes	No. Only to IRF members
60. Do you have any plans for providing data in electronic form?			No	Data are provided in electronic form already	Already exists
Overall comments					
61. Which do you think are the main advantages of your data file?	Common questionnaire: Only broad data collection	Easily (Excel file). Web application	Everything that can reasonably be collected, is there. However, many blank cells (to be removed? countries said no)	Easy access, timely and reliable data, user friendliness, service	Global combination on road + vehicle (only ones to provide this)
62. Which do you think are the main deficiencies of your data file?	Lack of data quality, lack of data	Not detailed enough	Empty cells		Look of more countries
63. Which are the limits of your data file?		Could be detailed		Data availability in the individual countries	More countries providing data (Not all members provide data: 185 members, less data)
64. Do you have any plans for improvement of your data file?				Yes	
Material to be provided					
65. We will be grateful if you could supply us with the following documents (if any):					
Extract of the statistics publication			Extract of the statistics publication;		
List of data contained in the IDF (road accident and exposure data)			List of data contained in the IDF (road accident and exposure data);		
List of definitions used for the IDF data		List of definitions used for the IDF data _ Glossary - download it	List of definitions used for the IDF data		
Description of the system					
Any other useful document					

	EUROSTAT /ECMT	ECE-UN	UNECE	IRTAD	IRF
Risk Exposure Data information					
66. What type of exposure data is available in the IDF and in which format?					Vehicle-kms by: road type / vehicle type ; Road network length by: road type / operation (public-private) / road surface type (paved-unpaved) / road condition (good-fair-poor)
Vehicle-kilometers and passenger-kilometers of travel (by road network type, by vehicle type, by vehicle age, by driver/passenger age and gender)	Yes. Vehicle I passenger - km yearly and per country. - Goods vehicles maybe more detailed Sources: statistical authorities	- Vehicle - kilometers and passenger - kilometers of travel (by road network type, by vehicle type, by vehicle age, by driver / passenger age and gender) Website	Vehicle-kilometers and passenger-kilometers of travel (by vehicle type)	Vehicle-kilometers (by road network type and by vehicle type) and passenger-kilometers of travel mode	Passenger-kms (only for roads public transport - private transport and rail)
Number of registered vehicles (by vehicle type and vehicle age)	Yes	- Number of registered vehicles (by vehicle type and vehicle age)	- Number of registered vehicles (by vehicle type)	Number of registered vehicles (by vehicle type)	Fleet of vehicles by: vehicle type / first registration and import / production and export
Number of driving licenses (by driver age and gender)	No	- Number of driving licenses (by driver age and gender) x		Number of driving licenses (by driver age and gender) not available	
Fuel consumption (by vehicle type)	ECMT data	- Fuel consumption (by vehicle type) _ statistics short term trends _ website		Fuel consumption (by vehicle type) not available	Fuel consumption (only total numbers for diesel and petrol)
Population (by age and gender)	Yes	- Population (by age and gender)		Population (by age) Yes	The population pyramid of users will be available from next year
67. What is the availability of the data (number of countries and number of years)?	Data availability Table	since 1980 (according to country availability)		Depending on the national availability when it comes to kilometrage data, historic time series back to 1970	
68. Which variables and values are used?		43 countries			
69. Which are the sources of the data in each country?					
70. Are there any coefficients or mathematical formulas used for the estimation or correction of the data?	Avoid the work of correction coefficient (don't really check the sources)	No			
71. How are these coefficients/formulas developed?	Avoid the work of correction coefficient (don't really check the sources)			Seat belt use (rate); Network length (by network type)	