Deliverable 2.5 - Risk Exposure Data – Recommendations for collection and exploitation

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:
CETE-SO – Centre d'Etudes Techniques du Sud-Ouest

Due Date of Deliverable: **31st Oct 2008**

Submission Date: **19th Nov 2008**

Report Author(s):
Gilles Duchamp, Vincent Treny, **CETE-SO,**
Stig Hemdorff, **DRD,**
Mouloud Haddak, **INRETS,**
Peter Hollo, **KTI,**
Joao Cardoso, **LNEC,**
Eleonora Papadimitriou, George Yannis, Antonis Chaziris, **NTUA,**
Frits Bijleveld, **SWOV,**
Torkel Bjørnskau, **TOI,**
Thomas Leitner, **KfV**

Project Start Date: **1st May 2004** **Duration:** **4 .5years**

Project co-funded by the European Commission within the Sixth Framework Programme (2002 -2006)		
Dissemination Level		
PU	Public	X

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

Table of Contents

Executive Summary

This report is the conclusion of the activities of Work package 2 of the SafetyNet project, which explored road safety risk exposure data.

In previous work, Risk Exposure Data appeared as a fundamental need for road safety policy makers and researchers, in order to determine if a given road safety issue is a local (national or regional) problematic, or an issue shared by all EU countries.

However, WP2 work highlighted several complementary needs in order to make RED use optimal:

- Need of data harmonisation
- Need of improvements and quality controls
- Need of new harmonised data,
- Need of harmonised existing, and new, collection methods

As a progressive answer, the present report gives firstly a set of general recommendations for all people involved in road safety, from policy makers to experts of the domain, dealing firstly with existing data to enlarge its matter to new data.

Moreover, it focuses on specific recommendations for data collection and use, in order to draw the general framework of a future road safety observatory, namely ERSO, which should, as far as benchmarking of different geographical areas may be compared, take into account risk exposure data indicators:

- Population
- Driver population
- Vehicle fleet
- Road length
- Vehicle-kilometres
- Three traffic and mobility exposure data : person-kilometres, number of trips and time in traffic

1. Introduction

The present deliverable, as the last of work package 2 of SafetyNet dealing with risk exposure data, is the conclusion of four years work.
In the light of the progress made from the State of the art ([Ref 1]) to the common framework ([Ref 3]), including the practical case of the pilot study ([Ref 4] , its aim is to give recommendations and guidelines to collect and use risk exposure data. The potential users of this guide represent a large group; policy makers at EU and national level, statisticians and experts as well as other stakeholders in the field.

To achieve the aims, these recommendations on data collection and use will be structured in two parts.

In the first one, **general recommendations will be given**; these will be distinguished two steps:

- one concerning the data already available, for which recommendations will suggest what to do in order to harmonize data and collection methods
- one exploring the way of obtaining new data

In a second part, more **specific recommendations** will be given, for each exposure indicator selected since the state of the art report:

- Population
- Driver population
- Vehicle fleet
- Road length
- vehicle-kilometres
- And the three traffic and mobility exposure indicators; person-kilometres, number of trips and time in traffic

2. General recommendations

2.1. Introduction

Significant efforts are made at national and international level for the enhancement of risk exposure data in Europe, for the improvement of data availability, disaggregation and reliability. In particular, most European countries have established and use specific systems and methods for the collection of more or less detailed exposure data. Moreover, several International Data Files (EUROSTAT, ITF, UNECE, IRTAD, IRF etc.) gather and publish the national aggregate exposure data in a systematic way, and are currently the only sources allowing international risk exposure comparisons.

Despite these important efforts, **a lot of work is still required** for reliable risk comparisons across the European countries. In particular, several weaknesses in the national data need to be addressed, namely poor data availability, incompatible definitions and inappropriate disaggregation (in relation to accident data), as well as limited accessibility. An important lack of balance can also be observed in the fact that the most useful exposure data are the least available in most countries or the least compatible at European level. Furthermore, different calculation methods and combinations of data (and sources) are used in different countries, to produce a national exposure estimate. As a consequence, the aggregate data made available through the International Data Files are unlikely to be comparable across countries, whilst data quality control within the Data Files is also rather limited.

SafetyNet provided an in-depth assessment of existing exposure data, elaborated the first transformation rules and tested a group of selected countries

Within SafetyNet WP2, several activities allowed for the **identification of the main problems concerning risk exposure data in Europe** and the respective steps required. In particular, an exhaustive state-of-the-art survey on risk exposure data was carried out, covering data definitions and properties, collection and analysis methods. Moreover, a detailed assessment of risk exposure data availability and compatibility was carried out, in which transformation rules were proposed for the improvement of data compatibility, where possible. Finally, a pilot study with actual exposure data from six European countries was carried out, in which the data was used for risk estimates of different driver, vehicle and road network categories across Europe.

SafetyNet revealed several limitations in the use of exposure data in Europe and confirmed the need for a future European framework

From these activities, a number of issues were identified regarding the future potential and current limitations of existing risk exposure data in Europe:

- Although the different exposure measures have different properties, advantages and limitations, vehicle- and person-kilometres, as well as the time spent in traffic are closed to the theoretical definition of exposure. However, this data is seldom available and/or comparable across European countries.
- Vehicle fleet, driver population and road length are useful alternative exposure measures available in most countries. However, the definitions used for the variables and values available are often not compatible.
- Moreover, vehicle fleet, driver population and road length data are usually stored in national registers, which are in most cases insufficiently updated. For instance, scrapped vehicles or deceased drivers are seldom removed from the respective registers.
- Several countries have travel / mobility surveys for the collection of vehicle- or person-kilometres data. These surveys have several problems:
  - First of all, these surveys have a purpose other than to provide exposure data, in the form required for accident risk analysis.
  - In each country, the travel survey may be of different type (e.g. telephone, roadside, diary etc.), unit (e.g. person, household etc.), target population (e.g. including pedestrians or not), coverage (e.g. rural areas included or not etc.), sample size, duration, and respondents' length of time covered (e.g. one day, one week etc.).
  - Additional sample limitations may be involved (e.g. age limitations, geographical limitation etc.).
  - A different degree of non response may be observed, measurement errors etc. Moreover, the magnitude of these errors is seldom known.
  - Although these surveys have important potential, in the sense that road user, vehicle and road network characteristics can be combined; the actual number of variables available is rather limited, whilst the definitions of these variables are often incompatible between different countries.
- Traffic counts systems are operational in several countries, allowing the collection of time series data for vehicle-kilometres. Common limitations of these systems include:
  - The coverage of the road network, as the systems are usually established on the national and main interurban road network, urban or rural areas being seldom included.
  - The vehicle classification; in most systems, two wheelers are not always detectable or properly classified.
  - The methods used for calculating vehicle-kilometres from the traffic counts, which may differ considerably among countries.
- In general, the data sources, the combinations of data and the calculation methods used for obtaining a national exposure estimate are not well documented. In most countries, a combination of data is used. A typical example concerns the combination of data from a travel survey for private transport with data from transport operators for public and freight transport. Another example concerns the various methods used for obtaining a national vehicle-kilometres estimate from a representative sample of traffic counting stations.
- This information is also not provided to the International Data Files gathering the national data. Therefore, the published figures may not be comparable.

- In most cases, the data requested from the International Data Files should comply with specific definitions. However, this is not checked and validated within the Data Files, given that the efforts are focused on receiving data in the first place.
- Moreover, most International Data Files focus on the collection of those exposure measures that fall within the field of transport statistics (i.e. vehicle and person-kilometres, road length and vehicle fleet), whereas the other exposure measures (i.e. driver population, number of trips, time spent in traffic) are seldom collected.

It was revealed that the **lack of a common European framework** for the collection and exploitation of risk exposure data limits significantly the comparability of the detailed national data. The need for harmonisation of exposure data across Europe was further confirmed by the in-depth assessment of existing data, in which guidelines for transformation rules have been proposed, as well as by the results of the pilot study. However, the appropriate concrete transformation rules can only be defined by working with the actual data, namely the entire national datasets, which is currently possible only within the respective national administrations. In the long term, a new framework for the collection of exposure data would allow for the direct collection of data in accordance with specific definitions and methods.

In this chapter, a **two-step process is proposed** for the improvement of risk exposure data in Europe, as shown in Figure 1. More specifically, recommendations for the improvement and harmonisation of existing exposure data are initially proposed. Furthermore, a framework for collecting new harmonised exposure data on a systematic basis across Europe is presented. These sets of recommendations can be examined by the European Commission, as well as by individual countries wishing to improve their data collection systems.

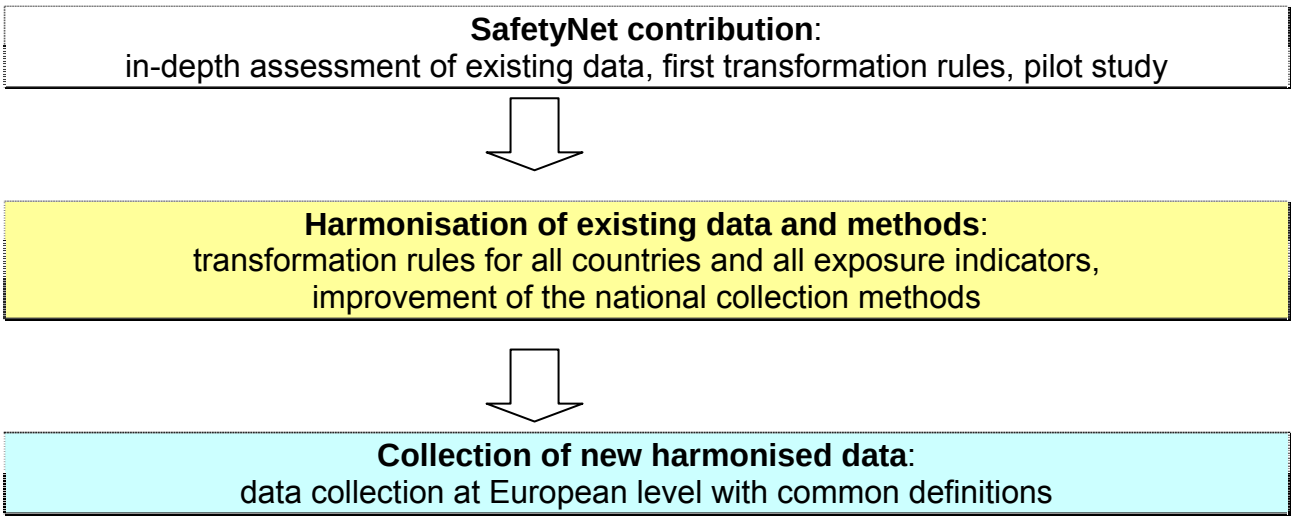

***Figure 1**. A two-step process for the improvement of exposure data in Europe*

2.2. Harmonisation of existing data and methods

2.2.1. National definitions and transformation rules

A major limitation of the existing exposure data is their incompatibility with CARE accident data, which is due to the fact that these data are mainly collected for other purposes than to provide road accident risk estimates. These incompatibilities mostly arise from **differences in national definitions of variables and values**.

For the improvement of data compatibility, appropriate transformation rules need to be developed, in order to align the existing data definitions with the respective CARE definitions, as presented in the CARE Glossary. A similar process was implemented during the development of the CARE system, where national variables and values were gradually harmonised for specific variables by means of transformation rules (e.g. CAREPLUS1 and CAREPLUS2 projects).

A "CAREplus" type of process could be implemented within the national administrations for each exposure indicator

However, **implementing a CARE-like process for exposure data** may be much more complicated, given that the work concerns several different exposure measures, ranging from vehicle and person-kilometres to road length and population. The problem is further complicated when considering that the different exposure measures are maintained by different national authorities, each one using a different definition for the same variable (e.g. for road type). For these reasons, it is recommended that this process is handled within the national administrations maintaining the data files, and only coordination and assistance should be provided centrally at EU level. The harmonised data should be then provided to EUROSTAT and ERSO as well.

Within SafetyNet WP2, the overall type of transformation rule required was identified for each particular case of data incompatibility. However, the specific transformation rules can only be developed by the competent national authorities maintaining the data, as only these authorities may have sufficient knowledge of the data properties. These **transformation rules** may be of different types:

- **Aggregation**: for example, in a country where road length is available for motorways, main interurban roads, secondary interurban roads and urban roads, an aggregation of the "non motorway" values would render the data compatible with the CARE accident data per "motorway / non motorway".
- **Conversion**, by means of appropriate coefficients: for example, if a country does not store mopeds in the vehicle register, other data sources may be used (e.g. surveys, vehicle taxation data) to estimate the proportion of mopeds in the total vehicle fleet, and develop a correction coefficient for the total vehicle fleet.
- **Selection**, where the data are limited to a specific subset of compatible values: For example, if a country collects person-kilometres for persons older than a given age

(i.e. such a limitation is imposed to the travel survey sample), it may be specified that this data can be used only for the same sub-group of the CARE casualty data.

- **Other**, more complex transformation rules may need to be developed in particular cases.

Four types of transformation rules need to be considered: Aggregation, Conversion, Selection and Other

It is noted that selection-type transformation rules are proposed as an alternative way of exploiting at least part of the existing data. In cases where it is not possible to yield a concrete transformation rule (e.g. conversion) this allows the entire dataset to be exploited in a reliable way.

It is also underlined that the transformation rules may concern the definition of the exposure indicator itself, e.g. vehicle-kilometres. In this case, the EUROSTAT definitions, as presented in the Glossary for Transport Statistics, need to be adopted. Therefore, the transformation rules for exposure data should be based, where applicable, on the EUROSTAT **definitions** for the global exposure indicators, and on CARE definitions, for the variables and values available for each indicator:

EUROSTAT definitions should be applied to the global exposure indicators, CARE definitions should be applied to the variables and values

In previous SafetyNet WP2 activities, the need for compatible exposure data was identified, resulting in an important number of indicators and variables to be harmonised. Moreover, given the number of different national authorities to be involved in the process, it would not be feasible to implement the process for all countries / all indicators and variables at the same time. However, SafetyNet WP2 identified specific **priorities in data needs**. These priorities may be used as a starting point for harmonizing the existing exposure data among European countries. In particular, they include:

- Road length data per road type, area type and region
- Vehicle fleet per vehicle type and vehicle age
- Driver population per driver age and gender
- Vehicle-kilometres per vehicle type, age, road type, area type and year
- Person-kilometres per person class, age, gender and year

The priorities in data harmonisation need to be considered first

It is noted that **driver population** data are currently not collected nor published by any international organization, although the quality of this data is considered to be quite satisfactory. Therefore, particular focus should be placed on collecting and harmonising this data, which may be very useful in person-oriented traffic risk analyses.

Once these priority variables have been harmonised, **additional data needs** may be tackled, as regards the following additional indicators and variables, which were identified as important ones:

- Vehicle fleet per engine size and region
- Driver population per nationality and experience
- Vehicle-kilometres per engine size
- Person-kilometres per nationality and experience
- Number of trips per person class, age, gender and vehicle type
- Time spent in traffic per person class, age, gender and vehicle type

In terms of **implementation process**, for each indicator the responsible national authority should be identified and a single contact person should be selected for each exposure indicator in each country. For that purpose, existing groups and networks can be used. For the exposure indicators related to transport statistics, the existing networks of EUROSTAT should be exploited. Other contact persons could be sought in the National Experts groups of the European Commission (e.g. CARE Experts). In any case, closer cooperation between the CARE Experts group and the EUROSTAT network would be extremely useful for strengthening the link between road safety needs of exposure data and current and future transport statistics.

The respective national authorities should subsequently be provided with guidelines for the development of transformation rules for the selected variables. Then, as mentioned above, national authorities would be in charge of the development and validation of the appropriate data transformation rules. It is noted that it might be preferable to not collect actual data while developing the transformation rules. Data collection may be initiated once the first set of transformation rules has been developed and validated.

The cooperation between existing official representatives for road safety and exposure data should be exploited and strengthened (e.g. the Eurostat network for Transport Statistics, the CARE Experts).

Then, for the harmonised data, **data collection** should be carried out by means of a common collection form (e.g. standard Tables). Certainly, an interesting additional task towards the exploitation of the existing data is the selection of data cross-tabulations i.e. combinations of variables, resulting in an efficient and realistic degree of disaggregation for the harmonised exposure data. A useful example may be found in the data collection forms used in the SafetyNet WP2 pilot study.

2.2.2. National collection methods

As mentioned in the introduction to this chapter, there are other issues jeopardising the comparability of exposure data among European countries, which can not be addressed through the harmonisation of exposure data variables and values in accordance with the CARE data variables and values. These additional issues are mainly related to differences in national **exposure data collection methods**.

More specifically, a major limitation rises from the fact that often **different methods or different combinations of methods** are used to provide a national exposure estimate.

This problem mainly concerns vehicle and person-kilometres of travel, as well as the number of trips and the time spent in traffic. Dealing with this type of limitation requires a lot of effort and resources, and is considered to be far beyond the scope of the harmonisation of existing data. Instead, recommendations on these issues are presented in sub-chapter 1.3.

On the other hand, there are other issues arising from the collection methods, which may be adequately addressed together with the development of transformation rules. These mainly concern the exposure data that are collected at **national registers** i.e. road length, vehicle fleet and driver population. The types of improvements of these registers may include:

- **Updating**; this is mainly the case of vehicle fleet and driver population registers.
  - As regards vehicle fleet, scrapped vehicles need to be removed from the register; vehicle taxation data may be a useful data source for correcting for this overestimation (e.g. vehicles for which tax is not paid may be considered as scrapped ones).
  - As regards driver population, deceased drivers need to be removed from the register.
- **Handling duplicate entries**; this is also mainly relevant to vehicle fleet and driver population registers, and may lead to additional overestimation of the actual number of registered vehicles / drivers.
  - As regards vehicle fleet, in several countries a new entry is introduced for vehicles changing license plates or for vehicles changing owner.
  - As regards driver population, given that the registers are driving license registers, a new entry is often introduced for drivers upgrading their license.
- **Linking with other data sources**; this is mainly the case of vehicle fleet and road length registers, for which in some cases subsets of data are available in different databases. The national authorities responsible for the main register should collect the additional data, at least in grand totals, and be able to provide the total aggregate figures.
  - For example, in a few countries, mopeds are stored in a different database (often maintained by a different authority), given that the licensing procedures are looser than the ones for cars and motorcycles.
  - As regards road length, it is common that regional or local authorities are responsible for local roads (including urban roads) and consequently these are not included in the national register. In a few countries, this data may not be available in electronic form; however, nearly all European countries are already almost fully covered by electronic navigation maps, which may be used for collecting reasonably accurate relevant length data on local roads.

The actions needed for the improvement of the national road, vehicle and driver registers include: Updating, handling duplicate entries and linking with other data sources

Addressing incompatibilities rising from the data collection methods becomes more complicated when examining travel surveys and traffic count systems, which are used

for the estimation of vehicle and person-kilometres of travel. Both these methods are **sampling methods** and are therefore subject to various types of errors (e.g. sampling, measurement, response etc.). However, in the detailed assessment of data collection methods carried out within SafetyNet WP2, very few countries reported known errors in the form of specific figures.

It would therefore be extremely useful if all countries using these methods undertook the necessary effort to provide an estimate of these errors, preferably in the form of **confidence intervals for the exposure estimates**. Certainly, this is not an easy or straightforward task; however related statistical tools exist and could be applied for that purpose.

Each country should calculate and provide indicators of reliability (e.g. confidence intervals, sample representivity, etc.) for the sample-based exposure data (vehicle and person-kilometres)

Finally, for the particular cases of a **combination of methods** being used to provide a national exposure estimate, SafetyNet WP2 experience demonstrated that the data sources, methods and calculations used in the different countries are quite diverse. Harmonising these combinations of methods would be an enormous task; however, each country should at least provide a **complete and comprehensive description of the data sources, methods and calculations** implemented. A lot of useful information on this topic can be found in the Appendices of SafetyNet Deliverable D.2.3. The continuation of this work, in a way to obtain systematic and comparable descriptions would be very important. Although the relevant data compatibility issues can not be addressed through this process, gathering the related information and providing the data users with a clear and concise documentation would assist in a better use of the data and an enhanced interpretation of the road safety results.

Each country should provide a comprehensive description of the data sources and calculations used for vehicle and person-kilometres

2.2.3. International Data Files

As mentioned in the introduction of this chapter, the International Data Files with exposure data are useful data sources, where the systematic efforts of several years allow road safety data users to access an important amount of aggregate data. However, although most of the data files require that the collected data comply with specific definitions (e.g. EUROSTAT/ ITF/UNECE), this is seldom cross-checked in practice. In general, data quality control is limited to correcting obvious errors in total figures.

In previous SafetyNet work, it was demonstrated that the exposure **figures published from different International Data Files for the same variable may vary significantly**. This is particularly the case for vehicle-kilometres, for which the differences in published figures among different International Data Files may reach 80%. Even for less sophisticated indicators (e.g. road length) important differences of around 15% were reported in published figures.

Obviously, once national data are harmonised according to the methods and processes described above, the quality of the published figures will be improved as well. In the meantime, **data quality control procedures** should be intensified within the International Data Files.

Data quality control procedures should be intensified in the International Data Files

Part of the differences in the published figures may be attributed to the fact that the different International Data Files may use different national sources. The identification of a **single national contact person for each exposure measure**, as is the case for road accidents national experts forming the CARE Experts group of the European Commission, would be extremely useful in this framework. In each case, the name of the contact person and the related national organisation should be available in all publications of the International Data Files.

2.2.4. The role of ERSO

The European Road Safety Observatory (ERSO) could have a central role in the process for harmonising the existing exposure data. On the one hand, the ERSO could be used as a gateway for bringing together all those involved in the process, for exchanging knowledge and for providing detailed guidelines where required. On the other hand, the ERSO will gain important benefits from this process, as it would be the main source of the truly harmonised data and the related meta-data.

ERSO should serve as a gateway for enhancing and monitoring the data harmonisation process

In particular:

- The European network of exposure data experts could be strengthened, through the closer cooperation between the existing groups (e.g. the Transport Statistics network of EUROSTAT and the CARE Experts group of the European Commission). The contact details of the official representatives for exposure data should be available at the ERSO. A password-protected area could be considered for communication within the group.
- A lot of knowledge on existing exposure data definitions, availability and quality is currently available at the ERSO, mainly in the SafetyNet section; these should be used as a starting point for the harmonisation of the existing data.

- Accordingly, the knowledge on exposure data collection methods, available through the SafetyNet section of the ERSO, should be fully exploited for the harmonisation of the existing data.
- The EUROSTAT definitions for exposure indicators and the CARE definitions for variables and values applicable to exposure data should be available at the ERSO.
- Comprehensive and detailed guidelines for the development of transformation rules should be elaborated, in order to assist the national administrations in the harmonisation of their data. These could be also published at the ERSO. Certainly, these guidelines could not replace the close and ongoing communication of the European Commission with the national authorities, for dealing with the particularities and specific problems of each country.
- Guidelines for estimating confidence intervals for sample-based exposure estimates could also be considered. These could be also included in the ERSO.
- During the existing data harmonisation process, a lot of important documentation on detailed national practices will be gathered. A section in the ERSO including a complete outline of these issues would provide valuable insight to the data users.
- The harmonised national figures that will gradually become available should be published, making ERSO the first international data source with compatible exposure data.
- Several ERSO outputs, such as the Annual Statistical Reports, the Basic Fact Sheets and other published aggregate data should be gradually updated and improved with more and more compatible exposure and risk estimates.

Eventually, ERSO would become the first international data source with truly comparable exposure data. Information should of course also be available at the EUROSTAT homepage

2.3. Collection of new harmonised data

The harmonisation of existing exposure data proposed in the previous chapter is an important process, whose implementation can be initiated in a quite straightforward way in the imminent future, and may result in major improvement of road safety analyses at European level. However, this process can by no means address all the needs for improvement of exposure data.

In particular, the development of transformation rules by national administrations, according to EC guidelines, may be most efficient for data collected in a systematic basis, namely those stored at national registers i.e. vehicle fleet, road length and driver population. On the other hand, the development of **transformation rules for the sample-based exposure estimates would not be sufficient**, i.e. for vehicle and person-kilometres, number of trips and time spent in traffic. The calculation of confidence intervals for this data would only at least allow the user to assess the magnitude of the problem. Besides, very few countries have sufficient and reliable data on these indicators, although some of these data are top priority for road safety data users.

For these reasons, it is obvious that comparable exposure data on these indicators can be made available only through **a pan-European data collection system**. Such a system should focus on the collection of disaggregate time series of advanced exposure data by road user, mode and network characteristics, and should be organised to provide data in a consistent and systematic way.

A pan-European data collection system should be established, focusing on vehicle and person-kilometres and including different data collection processes

Moreover, this system should include **different data collection processes**, each one presenting different features and advantages. 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 count 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.

In any case, existing efforts for data harmonisation, methodological guidelines etc., such as those of EUROSTAT and UNECE, should be thoroughly examined. The specific elements of the calculation process of exposure measures would be an important and complex task, and only some details are presented in the next sub-chapters.

2.3.1. Data selection

The pan-European data collection system should primarily focus on vehicle and person-kilometres of travel. Number of trips and the time spent in traffic are useful additional indicators to be collected if possible. The data priorities identified within SafetyNet should also apply in this future process. However, given that an important amount of effort and resources will be devoted anyway once the process is initiated, it is recommended that data collection aims to meet all the data needs and not only the priorities.

The pan-European data collection system should aim at meeting as many of the data needs as possible

The initial list of data to be collected includes:

- **Exposure indicators**: vehicle and person-kilometres, number of trips, time spent in traffic.
- **Variables**: vehicle type, vehicle age, road type, area type, month, day, hour, person class, person age, gender, driving experience and nationality.

This data can only be collected by sample-based methods, as mentioned above. Moreover, it is important to ensure as much data disaggregation and cross-tabulation as possible.

2.3.2. Collection methods

2.3.2.1. Travel surveys

A pan-European travel survey, having persons as units, would allow for the collection of exposure data cross-tabulated per person, vehicle and road network characteristics. The achievement of the desired degree of cross-tabulation for an important amount of variables would not be an easy task, however given the **added value of such data**, it is important to pursue this objective.

A pan-European travel survey would provide disaggregate person-kilometres, cross-tabulated per person, vehicle and road characteristics.

Existing efforts, mainly those of the UNECE and EUROSTAT related working groups, have provided useful insight as per the type of data required and the collection methods recommended, and these efforts can be considered as a first step towards setting up a pan-European survey system. However, a lot of work is still required in order to identify the specific needs and features of such a system, especially as regards administrative and financial aspects, which should be handled by the competent European authorities.

Moreover, the implementation of a pan-European travel survey involves several **methodological and practical issues**. For instance, the type of survey (e.g. telephone, roadside etc.), the sample size, the target population and the duration, are issues that need to be carefully examined. The creation of a common survey questionnaire and the development of a common methodology for calculating exposure indicators from the survey data, together with their confidence intervals, would also be very demanding tasks.

Obviously, a concrete proposal on these issues can only be elaborated after extensive consultation with road safety experts, and is certainly beyond the scope of the present analysis. It is noted, however, that although the survey design should be carried out at European level, on the basis of existing best practice, the implementation could be handled either at European level or by the national authorities. For example, the SARTRE survey methods could be examined. In any case, very close **cooperation** between all parties involved will be required, and the coordination of the whole process would be a critical point.

The pan-European survey design should be carried out at EU level,
but implementation could be handled
either at EU level or by each individual country

It is underlined, therefore, that the following steps are required for a pan-European travel survey:

- Review and best practice analysis of European travel surveys.
- Pan-European travel survey design (common questionnaire, survey features).
- Implementation of the survey design in each country.

Eventually, the pan-European travel survey should be carried out on a systematic basis, e.g. every five years.

2.3.2.2. Traffic counts systems

As mentioned in the introduction, travel surveys have the major advantage of providing exposure data combined per person, vehicle and road characteristics, but can not provide **continuous exposure measurements over time**. This can only be achieved by means of the exploitation of traffic counts data. However, the harmonisation of traffic count systems across Europe, or in the long term the implementation of a pan-European traffic count system would be extremely complex. This section presents some basic steps that can be gradually considered, given the usefulness of having comparable traffic counts data at the European level.

Traffic counts at the European level would provide
continuous traffic measurements over time,
which could be used for monitoring exposure.

Not all European countries currently have traffic counts systems; however, several countries have very extensive and advanced traffic counts systems. The **implementation and operation of a traffic counts system** requires significant infrastructure and management. In any case, sufficient coverage of at least the main interurban road network should be achieved, and vehicle classification in an accurate way should be ensured. It is important that all countries are encouraged and supported in order to implement such a system. Good practice examples can be found in several countries (see SafetyNet Deliverable D.2.1).

During the harmonisation of existing exposure data, it was recommended that all countries provide documentation about the way traffic counts are implemented in their country. From this information, a framework can be created for **comparing this data at the European level**. Eventually, all countries should send specific traffic count data to the European Commission (e.g. through the ERSO), so that exposure can be monitored at the European level.

However, a yet more useful solution would be the implementation of a **European traffic counts system**. This very ambitious project would require that all existing national traffic counts systems should be adapted and incorporated into this pan-European system, and that all remaining countries would install a traffic counts system according to appropriate European guidelines.

Guidelines for a European traffic counts systems should be elaborated. The system could be initially implemented on the Trans-European Road Network (TERN) and expanded gradually to lower level roads.

Given the obvious difficulties involved in such an effort, it is recommended that the European traffic counts system should initially include only the main European road network, and it should progressively expand to lower level roads. In parallel, it is important that a **common method for producing exposure estimates** on the basis of traffic counts systems becomes available at European level.

2.3.2.3. Driver and vehicle registers

Although the existing data harmonisation process, including the development of transformation rules and the improvement of the national registers, will be an important contribution to the improvement of vehicle fleet and driver population data, the development of **European registers** would be an even more valuable long term contribution.

The creation of European driver and vehicle registers with selected disaggregate data should be examined.

First of all, such European registers should include the **disaggregate data**, allowing for transformation rules, updates and other improvements of the registers to be

implemented centrally. Moreover, gathering this very useful exposure data at European level would be a major step for improved road safety analysis.

Some difficulties may rise from the fact that national registers often include personal / confidential information (e.g. person name and address, vehicle plate number etc.). It could be arranged that only a **selected number of variables** for each entry of the national register to be available in the European register.

Some first steps for the harmonisation of European vehicle fleet registers are carried out within DG-ENV of the European Commission. Cooperation among all the interested authorities (including ERSO and EUROSTAT) would therefore accelerate the process and prevent duplication of effort.

2.3.2.4. Other methods

Other methods could complement the proposed future European framework for exposure data.
In particular, **odometer readings** from regular vehicle inspections can be a very useful way of obtaining total yearly vehicle-kilometres, possibly per vehicle type and vehicle age, even if often limited to vehicles older than a given age. These figures can be used not only for country comparisons, but also for monitoring the various exposure estimates produced by means of other e.g. sample-based methods at national or European levels.
Other new domains to explore are ITS domain and satellite applications to mobiles, especially linked to GALILEO and KOPERNIKUS. Indeed, using instrumented and "intelligent" vehicles, at least for specific vehicles as HGV, as well as satellite observations, as images analysis for mobility studies, should invite road safety researcher to develop new data collections and processes in order to complete road safety RED.

2.3.3. Contribution to the ERSO

Certainly, the establishment and operation of a European data collection system for exposure data would be a complex and time-consuming task, which would also involve a significant effort and cost, both at national and EU levels. Moreover, this system should comprise different sub-systems and procedures, as outlined above. Despite these complexities, given the importance of improved exposure data availability and quality, to support and monitor an efficient road use and safety policy at EU-level, **it is necessary to promote its development**.

As in the case of harmonisation of existing data, the development of a European exposure data collection system should exploit the knowledge and possibilities provided by the ERSO, having as final goal the enhancement of the ERSO. The use of ERSO as the gateway for all those involved in the development and implementation of the system could be considered. The publication of new comparable exposure data at European level through the ERSO would be of **major added value to all ERSO users**, including

researchers and policy makers.

3. Specific recommendations

In the following chapter, more specific recommendations concerning each RED are presented. These recommendations will primarily address the problem of lack of harmonisation in national collection methods. However, they may also be seen as a guide should any national register require alteration. This would lead, comparison on European level to become easier.

3.1. Population

As recorded in the common framework (D2.3) population is a common exposure measure used in road safety related analyses. According to the needs of risk exposure calculations population data are useful to European countries for the assessment of public health risk on a population level. In particular three variables are important; person age, gender and nationality. Population data are also available on a regional level therefore risk calculations can also be developed by matching road safety data with population data on a regional level (NUTS III).

Its most important advantage is the high availability of population data for all countries in Europe offered by free data download from EUROSTAT: All 28 European countries (EU-27 plus Norway) collect population data in national registers and update it on a regular basis by nation wide population censuses. Other data sources can also be consulted because population data is in most cases easily comparable and can be used without transformations, as no significant inconsistencies are observed.

All variables and values (in particular person age, gender and nationality) included in the population registers have a straightforward meaning. Therefore their definitions and their compatibility could easily be assessed and used for any risk calculation in matching with population based road safety variables and values in CARE.

Population data are generally available for a long time series. Population censuses are carried out on a regular basis (normally every 10 years). Data on intermediate years are estimations, based on the results of the censuses, therefore the population registers are updated annually. That way the change of risk exposure indicators can be described for various time periods and furthermore allows possible trend estimations for different countries.

Attention should be given to the character of population data: In general, EUROSTAT and also other sources on national level provide average population data and population as of the 1st January of every year. To avoid misleading results, population data with the same characteristic should be used.

A risk calculation based on population data only is not significant for international road safety comparisons in the case of different levels of motorisation, traffic density, etc. For international comparisons, additional exposure data has to be consulted for risk calculations (e.g. vehicle fleet, vehicle kilometres, etc.).

Summary regarding the use of population data for risk exposure analyses:

- High data availability for all countries
- Data from different sources can be easily compiled and compared
- Risk analyses can be developed in a differentiated way by doing analyses based on several variables (age, gender, nationality, etc.)
- Population data allows risk calculation for time series and possibly trend analyses
- Attention regarding the different characteristics of population data (specific times of the year, yearly average)
- Limited usability for international comparisons without using any further exposure data for risk calculations.

3.2. Driver population

In general information about driver population is collected in all countries following the information from Deliverable 2.1. State of the Art Report on Risk and Exposure Data.

Basically the information in all countries is based upon a driver license database.

In the case of driver population geographically distribution has no meaning (NUTS levels) as the distribution of driver licenses will be connected to where drivers live, which is different from the location of accidents.

3.2.1. Registration problems

In some countries, people are not omitted from or marked in the database when they die. Therefore the number of drivers can be overestimated. Also the registration of moped drivers can give some problems as there are differences throughout the countries in whether a license is required or not.

3.2.2. Data availability problems

All though data is registered in the country the information may not be available (published) in a way that can be used in the framework of RED exposure data. As an example Denmark has a license database covering all licenses by age of issue for each category of license by individual. But the published statistics only covers newly acquired licenses, lost licenses and recovered licenses. So although the information basically exist the collected statistics do not allow one to draw conclusions about driver population in general.

3.2.3. Errors and misinterpretations

When individuals who have died or who are no longer licensed are not marked or removed from the register there may be an overestimation of the number of drivers.

As it is of interest to know the number of licenses by category some drivers will be registered under more than one category. Therefore it is important to have the total number of licenses as well, as the sum of categories will be larger then the actual number of drivers. On the other hand it can be important to know the information by category for special surveys as accidents with heavy goods vehicles or motorcycles.
Also having a driver's license for a specific category such as for a lorry does not necessary imply that one frequently drives a lorry. The licence could have been acquired many years previously but no longer used. It is the same case for elderly people who have a license for car but no longer drive. In both cases this can give an overestimation of the population. Some correction factors might be established if a question about driver's license and use is introduced as part of travel surveys given the extra possibility of errors that might introduce.

Information about the licenses of foreign people is not known. This can lead to an underestimation of driver population concerning comparison to accidents. There is no way to cope with that problem. In the past, that problem was perhaps limited to few countries within the EU having traditionally a large number of foreign drivers (one is Luxembourg); but it can become more and more significant as travelling abroad may become a more and more usual behaviour.

3.2.4. Recommendations

Based on the discussion above we will conclude that in spite of the methodological problems involved national driver license databases are still the best source of information for driver population.

Based upon the national databases it is important that following information is made available:

- Total number of active driver licenses (that is excluding people who have died or people who have lost their license to drive (whether temporarily or permanently).
- The information of number of licenses should also be available by license group and preferably by age group of license holder.
- The information should be available by year at least (this could be by 1st of January).

3.3. Road length

Road length data is a practical exposure variable for the estimation of traffic risk at the network level, as evidenced in the Common Framework SafetyNet report [Reference].

In the mentioned report, road class, area type and region (NUTS 2 level) are regarded as important and useful variables, as they are also included in the CARE database.

Road class (motorway vs. non-motorway) was considered as a very usable variable in most countries.

However, in the pilot study several limitations were detected in the availability of the value “non-motorway”, which restrict considerably the usability of the variable. These limitations are worsened by low availability of the variable area type, as few countries provide data on non-motorway roads in urban or rural areas.

Furthermore, availability at the NUTS 3 or county levels is not widespread. Such a high level of disaggregation is not especially important for motorways; however, in the case of non-motorway roads potentially interesting comparisons between small geographical areas are impeded by the absence of this disaggregation, which restricts the efficient use of existing accident data.

There is the possibility that in several countries not yet providing data for non-motorway roads, the relevant information needed for providing road length by road class, area type and NUTS 3 (according to EUROSTAT and CARE definitions) is already available in local or national authorities responsible for the operation and maintenance of lower class roads. In those cases, procedures for aggregating the information through several road authorities are needed, such as a national questionnaire.

Where relevant data are not yet available, it is important that national authorities carry out the required activities for collecting this information, which is important for supporting decisions on safety issues and on other transport policies. Road length data may be collected on-site, using vehicles equipped with odometers, or with maps. In both cases, care must be taken in order to adequately handle intersection areas and avoid double measuring their length.

3.4. Vehicle fleet

The vehicle fleet – as exposure data – can be used well for international comparisons of road safety, because it is in correlation with the level of motorisation. To make valuable international comparisons, the level of motorisation has to be taken into account.

Previous work [Ref 6] has shown that, even if country and time dependant, there is a relationship between the level of motorisation and road safety (fatality risk).

According to research results, the best estimation of the exposure is the number of vehicle kilometres, but these data are not always available and they are very expensive. Moreover, the available data are not always reliable. Therefore, – in lack of reliable vehicle kilometre data, – the vehicle fleet is the so-called “second best” exposure data. Of course, the number of vehicles and the number of vehicle kilometres are in correlation with each other, even if this is not a linear relationship.

Taking into account that the fatality risk is entirely different in a bus or in a car, or on a bike, etc., it is necessary to make the comparisons in the respect of different vehicle categories.

3.5. Vehicle-kilometres

3.5.1. Summary of current potential

The number of vehicle kilometres is probably the most appropriate exposure indicator for the estimation of accident risk. Vehicle kilometres are a direct measure of traffic volume and can be available in a significant level of disaggregation, i.e. time, vehicle type, road type, driver characteristics etc. Vehicle kilometre data allow for very meaningful analyses to be carried out, especially when used in conjunction with accident counts in order to produce accident risk estimates.

However, in practice, the availability and the level of disaggregation of vehicle kilometres varies significantly and is strongly dependent on the type and features of the collection method used in each country. Moreover, the calculation of the exposure estimate is not consistent throughout countries resulting in a low overall compatibility. Vehicle kilometres are estimated by several methods, most of which include data collection by surveys and traffic counts. Furthermore estimations are also carried out by the use of statistical models and combinations of methods.

Ten European countries use surveys for the estimation of vehicle kilometres. The variables and values used in each survey vary significantly; however, the road type and the vehicle type seem to be the most commonly used variables between countries. Apart from surveys, traffic counts are common for the collection of vehicle kilometre data across Europe. Eleven European countries collect vehicle kilometre data by traffic counts alone, while some use both methods (surveys and traffic counts) for this estimation. Again, the vehicle type variable which is widely used for the data collection is also the most useful for traffic risk analyses as in most cases each country's and the CARE definitions are compatible. The road type variable is also widely used among EU countries for the collection of vehicle kilometre data, mostly by motorways and non-motorways, as well as for other road type groups.

As far as international data files are concerned, collection methods vary according to the type of data and the responsible authority collecting these. As regards the collection of data on distance travelled, EUROSTAT, UNECE and ECMT use a common questionnaire to collect data for four vehicle types (motorcycles, passenger cars, buses, lorries and road tractors), while yearly data for distance travelled by vehicle type are also available in the IRF database. The IRTAD database also contains data on distance travelled, by four road classes and six vehicle types. Data from these files should be used with caution, given that the data sources are not always known, and the published vehicle-kilometres figures are different in different data files.

3.5.2. Indicator, variable and value sets and definitions

In order to obtain a common and compatible risk exposure measurement unit, the definition of the indicator should be uniform between all countries. In the Glossary of Transport Statistics (UNECE, EUROSTAT, ECMT, 2003) a definition of vehicle kilometre is proposed, which could form the basis for a common pan-European definition.

Vehicle kilometre:
"Unit of measurement representing the movement of a road motor vehicle over one kilometre. The distance to be considered is the distance actually run. It includes movements of empty road motor vehicles. Units made up of a tractor and a semi-trailer or a lorry and a trailer are counted as one vehicle".

Vehicle kilometre data are most useful for traffic risk analyses related to the vehicle and the road network. The vehicle type, vehicle age, vehicle engine size and road type seem to be the most important variables for the estimation of traffic risk on vehicle level, while the vehicle type, area type, road type and region variables are most important for the estimation of traffic risk on network level.

3.5.3. Data collection methodologies and exposure indicator estimation

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 source, which practically can provide continuous exposure measurements over time. Both methods, separately or in conjunction, can be used efficiently in order to collect and estimate vehicle kilometre data.

As far as travel surveys are concerned the adoption of a pan-European travel survey could provide harmonised data and enable risk comparisons between countries. The adoption of a common survey and the development of a common methodology for the estimation of vehicle kilometres from the survey data are the most important tasks that need to be tackled.

On the other hand, traffic counts can provide reliable and complete information for the calculation of all useful traffic estimates by means of (automatic or manually operated) counting stations. These stations can provide systematic and concise information for the direct measurement of the traffic volume. Traffic counts aim to obtain a representative sample to be exploited for the extrapolation of a national or regional traffic estimates. Within a typical methodology, average annual daily traffic (AADT) data for homogenous links would be multiplied by the respective road length and the appropriate period of time in order to calculate the respective vehicle kilometre data.

In the short or medium term, it is necessary that all countries carrying out surveys or traffic counts are able to provide documentation on the survey or count's system design and features, as well as on the methods used to calculate vehicle kilometres from the "raw" survey or traffic counts data. For countries using a combination of methods and data sources, the need for well-documented processes and calculations is even more important. Moreover, all countries should put some effort in calculating confidence intervals for the sample-based vehicle kilometres estimates.

As long as the data obtained by means of surveys and traffic counts cover a sufficient time period, they could be further integrated in appropriate statistical models. The

forecasts of vehicle kilometres by using older time series can prove useful when more recent data are not available.

3.6. Traffic & Mobility Risk Exposure Data

Among Risk Exposure Data, three of them are linked both with traffic and mobility: **Person-kilometres, time in traffic and number of trips**.

Certainly countries collect those RED either by use of surveys, by use of traffic counts and occupancy rate estimates and/or by use of passenger registers.

But for those RED, traffic counts and occupancy rate estimates give very limited added value to ordinary vehicle kilometres data. The use of passenger counts etc. is also of limited value as exposure data due to the fact that person variables (age, gender) and values are normally not available from such registers. In addition passenger registers are naturally restricted to public transport modes. These modes of travel have generally low risk and are thus not the most relevant as exposure data for road traffic risk estimations.

On the contrary, use of surveys for traffic & mobility RED can give information for non-motorised road users (bicycles and pedestrians) and also cross tabulated data for age/gender groups of road users (both motorised and non-motorised).

So, as they give more detailed data than other methods, surveys seem accordingly to be the preferred method to achieve those RED, and the specific recommendations about those RED presented here will thus be concentrated on the way to collect relevant data by use of surveys.

3.6.1. Errors surveys for traffic & mobility RED are faced with

Two main types of survey will be discussed; travel surveys (not on road) and road-side surveys. When addressing these two survey types it is important to bear in mind the major methodological errors that concern surveys in general i.e. sampling errors, non-response errors and measurement errors. These errors are described as follows ([Ref 1] p. 38):

- 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.

3.6.1.1. Sampling errors

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.
Two appropriate approaches are possible:

- Either 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.

- Or a small selection of individuals is chosen which is supposed to be representative for the entire population

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, surveys for mobility should be what are called stratified sampling, which mix both approaches, in order to ensure that sufficient disaggregated information is available for the selected subpopulations of the sample to extrapolate them to the entire population. So, sampling errors should normally not be a major problem.

However there can nevertheless be a number of issues and problems at hand.

One question is: What should be the sample unit? Persons, households, vehicle owners, license holders?

Currently all type of sample units are used in travel surveys. The advantage of using person (vehicle owner or license holder) as unit compared to household is that it will in many cases be easier to ensure a representative sample and also that the borders of households may be somewhat arbitrary (e.g. children with divorced parents living 50 per cent of their time with each parent).

On the other hand there may be advantages of using household samples, notably because household representatives may be able to report trips made by children etc. who are not in the position to report these trips accurately themselves. Still, from a methodological point of view person samples are to be preferred given that samples can be drawn from adequate and updated registers. In some countries such registers are missing and thus a household sample based on addresses etc. is the only sample possible.

For road-side surveys the sample unit will obviously be vehicle drivers and/or vehicle passengers, but also pedestrians can be the unit in road-side surveys. The main potential sampling error involved is connected to whether the time and place selected for interviews is representative for the target population. If the aim of a road-side survey is to obtain representative person kilometre data for the whole nation, the survey will have to encompass all types of roads, times of day and times of year in order to be representative. Thus to use road-side surveys to give national estimates will present rather substantial methodological (and financial) challenges.

3.6.1.2. Non-response errors

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.

Mobility surveys will be faced with a very common type of error introduced by the way the survey is conducted. Whatever the way the statistical institute chooses to communicate with the respondents, some potential respondents are not able to respond using that way of communication because they do not have (a (fixed) telephone line, a fixed address, internet, e-mail…). That bias is not specific to mobility surveys

Conversely, some non-responses are directly linked with mobility itself:

- 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.
- Some people do not want to respond to any survey in given circumstances as “never stop to answer while driving”…

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.

Travel surveys conducted by use of telephone interviews (or paper/web questionnaires) are victim to a serious non-response error due to the fact that people who travel a lot will not be available at home (on the phone) to the same degree as people who do not travel frequently. Such non-response errors may produce serious bias and underestimate the true amount of travel. There is no fool-proof way around this problem, but it may to a large extent be met by extensive call-back procedures.

For road-side surveys the flip-side of this problem is apparent: one will not reach people who seldom or never travel. Accordingly, to compute average travel distances by use of road side surveys may over-estimate average travel distances. There is no obvious way to correct this possible non-response error.

Another problem one often meets when conducting telephone surveys (or paper/web questionnaires) is that response rates differ systematically by population age and probably also by other variables. Often young people are under-represented among responders on telephone and paper surveys, but they may be over-represented on web-questionnaires. To solve such problems one needs to carry out analyses of non-responses and to adopt weighing procedures when analysing data. In road-side surveys one will meet a similar problem, namely that people being in a hurry will refrain from answering. This may constitute a serious non-response problem because such people may very well travel more than the average road user.

A more general non-response problem encountered in surveys is that respondents having an interest in the subject and/or have been involved in the activity in question are more inclined to answer than those not interested or not involved in the activity in question. However in order to have correct estimates of the amount of the activity it is important that also the more passive responders respond, if not the estimated averages will be biased.

Surveys targeting specific groups or activities by use of random population samples will thus often get responses from the most active ones and accordingly estimates of averages will be biased. Neutral travel surveys are therefore in many cases preferable to more specific surveys. A telephone-survey to a random population sample entitled "bicycle survey" will for instance be answered more by those riding bicycles than those not, and thus estimates of travel lengths by bicycle for the whole population will be biased upwards. This is in a way a similar problem that one encounters in road-side surveys; only those involved in the activity will respond.

Nevertheless, for uncommon travel modes or vehicle types, general and neutral travel surveys will often give very little data and accordingly road-side surveys or specific targeted surveys is the only viable way to obtain data. For motor cycle trips and travel this will be the case in many countries. A good alternative is to draw samples from motor cycle owners or motor cycle license holders and obtain travel data from them. Such samples will be representative for motor cycle riding in the total motor cycle population.

It is important to note that there is a principal distinction between such register-based surveys targeting a specific group and general population-based surveys targeting specific groups or activities. In the bicycle example, the problem arises because one uses a random population sample to ask about a seldom activity. In that case there is reason to believe that the sample responding is not representative.

In the motor cycle case, on uses a sample from registered owners or license holders and thus one does not ask people who are not somehow associated with motorcycles about their motor cycle riding. Thus for transport modes where reliable vehicle fleet registers and/or driver license registers exist, a survey of the owners or the persons holding a license about the use of the vehicle can be an attractive alternative to normal travel surveys. In some cases such a procedure is also preferable because data about the use of rare transport modes/vehicles will be scarce and subject to large random variations in ordinary travel surveys.

3.6.1.3. Measurement errors

Measurement errors may constitute serious problems in travel surveys.

Firstly, there is the general problem of the misunderstanding of the question and /or the procedures. This is likely to be the reason why interactive surveys are better. In such a setup the respondents may be 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.

Secondly, concerning specifically mobility surveys, other difficulties are upraising:

- Many people have great difficulties when it comes to estimating kilometres travelled or time spent in traffic and thus the data obtained may give wrong estimates. Moreover,
- There is reason to believe that the ability to give correct estimates varies systematically by distance travelled.
- Time of day, day of week etc. may be difficult to remember and such measurement errors will accordingly be more widespread when travel surveys are made by telephone interviews than by road side interviews.
- Respondents are ignoring 'trivial' transport. Respondents tend to recall the major trips only; the minor ones are deemed insignificant and therefore forgotten. For instance, based on results obtained in some surveys, many people should be supposed to live at train stations, and go to work at other train stations: Respondents forgot that they took the bus, tram or bicycle, or went by foot to the station and did similarly at the station of arrival. Respondents tend to recall that they went 'by train' in such cases. 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?'.
- Return trips may be forgotten (in some surveys return tips as considered part of the original trip).
- People apparently have problems estimating time and or distance. 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.
- For information based on trips that took place longer in the past, problems memorising trips may arise as huge trips are often remembered better than smaller ones.

For both type of surveys there are of course also the normal sources of error caused by human or technical errors when the information is registered.

One possible solution to avoid measurement errors in surveys is to adopt tests of logic and reason to check answers. If some respondent gives an extreme answer to one question, for instance how many kilometres he/she travelled the previous day, the answer ought to be doubled-checked: “You have reported xxxx km of travel yesterday. This is very much and it amounts to having driven the distance from A to B. Are you sure that you have travelled that far or do you want to go back and correct it?” It is also possible to use digital maps etc. in order to locate trips and to make it easier to have correct distances registered.

3.6.2. General Recommendations for surveys

Based on the discussion above the conclusion will be that, in spite of the methodological problems involved, telephone-based travel surveys are to be preferred to other types of data collection for the purpose of obtaining representative data for calculating risk per the three traffic and mobility RED. When respondents are contacted at home one will

also get responses from infrequent travellers and surveys at home allow for more detailed questions. For traditional travel surveys, telephone interviews are (still) to be preferred to paper questionnaires and web-based questionnaires. Paper questionnaires would have to be long and complicated and web-based questionnaires will not be adequate for elderly people (without internet access). For more specific surveys for example to motor cycle owners both paper questionnaires and web-based questionnaires may be good alternatives.

Normal travel behaviour will be well reported in ordinary travel surveys and will thus in principle also be suitable for calculations of person kilometres. Lesser used transport modes will however not be well reported, and often more tailor-made studies will be needed. If these modes are made up of specific vehicles that are registered in some vehicle register a survey explicitly directed to map the use of the vehicle in question can be administered to the owners (or license holders) of these vehicles. However, a survey explicitly directed to map the use of less common travel modes or vehicles should not be directed to random population samples because those not using the vehicle or travel mode in question will tend not to respond.

In order to avoid errors and obtain good traffic & mobility data, national travel surveys are recommended. If necessary such surveys ought to be supplemented with more tailor-made surveys for rare transport modes or vehicles – if there are registers making it possible to identify users. This will typically be the case for motor cycles, but not for bicycles, and not for mopeds in many countries. In the latter cases the better way to get exposure data is to include the vehicles/transport modes in the general travel survey and expand the sample.

Travel surveys are currently the most promising method available in order to have adequate data on person kilometres distributed by age/gender/road user and thus it is important to design the surveys in ways that allow for relevant risk calculations to be made. It is therefore recommend that travel surveys are conducted as follows:

- For risk exposure purposes travel surveys ought to be nation wide. Travel surveys in particular areas are less suitable because it is difficult to know how representative the area is, what the exact area covered is and it may be difficult to have precise correspondence between exposure data and accident data.
- Travel surveys ought to have sub samples distributed over a whole year (for instance sub samples every day) in order to account for seasonal travel variations
- Travel surveys ought to include data also for professional drivers and travels conducted as part of work in addition to private travels
- Travel surveys based on person samples often lack data for children. A possible way to obtain some data for children is to ask car drivers about age and gender of passengers. It is also possible to conduct travel (and activity) surveys among children using parents as responders
- It is important to distinguish between travel made in a road traffic environment and travel made outside the road network (in the mountains, forests etc.). For pedestrians and cyclists this is particularly relevant.
- In order to reduce the problems with inaccurate reporting of distances and time one should adopt tests of logic and reason to check answers.
- In addition to distance travelled one ought to try to register travel time as well.

- 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.

3.6.3. Recommendations for person kilometres and number of trips

Even if none of IDF (EUROSTAT, IRF, IRTAD, UNECE, ECMT) contains data on the number of trips made in each country, person kilometres and number of trips can be regarded as similar and very useful risk exposure indicators:

- As the number of trips may be informative when person kilometres are available, both are based on the same sources, and consequently the same level of disaggregation will be available.
- As person kilometres and number of trips are normally registered by the same travel surveys, the main important recommendation is to adopt a unique definition of a trip, and to collect all data whatever its length or its duration is.

In order to make respondents remember small trips, typically by foot or bicycle, it is important that the interviewers are aware of those points and that they actively encourage respondents to include all kinds of trips in their responses.

3.6.4. Recommendations for time in traffic

Time in traffic is one of the most relevant risk exposure indicators, adapted for all the road user categories; pedestrian, cyclists, public transport passengers or motorised road users. It is a risk exposure indicator defined as the total time spent in travel by persons, regardless of their age or their mode or means of transport.

It could be defined also as the total time spent in traffic. The background idea is more general: Only while being involved in traffic – moving or halted – one is exposed to being involved in an accident.

However, for the present time, this RED indicator is not used because it is not easy to obtain this information. Data on RED time in traffic are very rare, partial and not available for all the road users.

In general, time in traffic data are available from travel surveys, carried out for mobility research purposes and not for risk exposure purpose. So;

- These mobility surveys, when they are available, are often more adapted for motorised trips than for non-motorised trips. They are not adapted for the youngest children …
- these surveys are often very specialised: some deal only with daily trips, others with holiday trips …

Finally, these surveys give a partial overview of the traffic which under-report the non-motorised trips.

To conclude, difficulties may be encountered in the disaggregation of time spent in traffic, especially as regards comparisons; comparing the time spent in traffic between different road users, or by different age groups.

Harmonised European travel surveys must deal with all of the population, all of the trips (daily or holiday trips) and all of the types of road users or means of transport. All countries must have the same definitions. They may also work with data of the same year.

With such a sampling survey, RED time in traffic could be broken down by:

- National/Regional level
- date and time
- Vehicle type
- Transport mode
- Motorway/No Motorway
- Rural/urban area
- Age
- Gender

References

[Ref 1] SafetyNet, Deliverable 2.1, Risk exposure Data - State of the Art report, 30/10/2005, NTUA,CETE-SO, DRD, INRETS, KTI, LNEC, SWOV, TOI

[Ref 2] SafetyNet, Deliverable 2.2, First classification of the EU member states on RED, 31/01/2007, NTUA,CETE-SO, DRD, INRETS, KTI, LNEC, SWOV, TOI

[Ref 3] SafetyNet, Deliverable 2.3, Risk Exposure Data - A Common Framework, 31/07/2008, NTUA,CETE-SO, DRD, INRETS, KTI, LNEC, SWOV, TOI

[Ref 4] SafetyNet, Deliverable 2.4, Risk Exposure Data – Pilot Study, 31/10/2008, NTUA,CETE-SO, DRD, INRETS, KTI, LNEC, SWOV, TOI

[Ref 5] Sampling Techniques. Second edition, Wiley, Cochran, W.G, 1963.

[Ref 6] Safetynet, Deliverable 7.8, The CARE data in perspective, 31/10/2008, KfV, NTUA, IBSR,INRETS