Road Safety Performance Indicators: Manual. Deliverable D3.8 of the EU FP6 project SafetyNet

Deliverable D3.8:

Road Safety Performance Indicators Manual

Please refer to this report as follows:

Hakkert, A.S and V. Gitelman (Eds.) (2007) *Road Safety Performance Indicators: Manual.* Deliverable D3.8 of the EU FP6 project SafetyNet.

Contract No: TREN-04-FP6TR-S12.395465/506723
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:
SWOV

Editors for this deliverable:
Shalom Hakkert and Victoria Gitelman (Technion)

Report Author(s): Kerstin Auerbach **(BASt);** Roland Allenbach **(BfU);** Vojtech Eksler **(CDV);** François Riguelle **(IBSR);** Elke Moons, Geert Wets **(IMOB);** Mouloud Haddak **(INRETS);** Péter Holló **(KTI);** Elisabete Arsenio **(LNEC);** Eleonora Papadimitriou, George Yannis **(NTUA);** Kirsten Duivenvoorden, Charles Goldenbeld, Sjoerd Houwing, Robert Louwerse, René Mathijssen, Chris Schoon, Martijn Vis, Wendy Weijermars **(SWOV);** Victoria Gitelman, Shalom Hakkert, Malka Avitzour **(TECHNION);** Terje Assum **(TØI);** Lucy Rackliff **(VSRC)**

Due Date of Deliverable: 31/12/2007 **Submission Date: 31/12/2007**

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**	

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

Executive Summary

Safety performance indicators (SPIs) are measures (indicators), reflecting those operational conditions of the road traffic system, which influence the system's safety performance. Basic features of SPIs are their ability to measure unsafe operational conditions of the road traffic system and their independence from specific safety interventions. SPIs are aimed to serve as assisting tools in assessing the current safety conditions of a road traffic system, monitoring the progress, measuring impacts of various safety interventions, making comparisons, and for other purposes.

Seven problem areas in road safety were selected for the development of SPIs in Europe, they are: alcohol and drug-use; speeds; protective systems; daytime running lights; vehicles (passive safety); roads (infrastructure) and the trauma management system.

The theory behind the development of SPIs in each of the seven safety areas was presented by Hakkert et al (2007)[1]. The data obtained from the cooperating countries and the comparisons of safety performance of 27 countries[2], in terms of the estimated SPIs, were presented in two other reports[3] – Vis and van Gent (2007a), Vis and van Gent (2007b).

This report is called a *Manual* as it should assist the countries in establishing the necessary systems of data collection for producing national SPIs, in each one of the predefined safety fields, and to make them comparable on a European level. For each safety area, the report defines quantitative SPIs, demonstrates existing practices for their measurements, provides best practice examples (when available), and details the procedures which are necessary to collect and process the required data for the estimation of the SPIs' set on a national level.

Recognizing the potential for road safety improvements coming from the use of harmonized SPIs across the EU, enabling benchmarking as a proven tool in road safety policy, the Member States are encouraged to seek ways of applying a uniform methodology for producing national SPIs. The procedures and methods presented in the *Manual* should be treated as minimum quality requirements for producing national SPIs, in each one of the predefined safety fields.

In addition, the report provides a more general theoretical background concerning the sampling issues in estimating SPIs (in general and in the context of specific SPI areas). Regarding setting up an SPI survey, the main questions considered are: sampling procedure to obtain a national sample; sampling size; sampling error; stratified sampling (combination into a single SPI by weighting); representativeness of the results and estimating confidence intervals of the SPI values. These issues are discussed in Chapter 2 and in the Statistical *Appendix*.

Alcohol and drugs

In the area of alcohol and drugs, the proposed SPI is the number and proportion of severe and fatal injuries resulting from crashes involving at least one active road user who was impaired by psychoactive substance abuse (concentration above a predetermined impairment threshold).

---

[1] Hakkert, A.S., Gitelman, V., and Vis, M.A. (Eds.) (2007) *Road Safety Performance Indicators: Theory*. Deliverable D3.6 of the EU FP6 project SafetyNet.

[2] Those include the 25 Member States as the EU at the end of 2006, and Norway and Switzerland

[3] Vis, M.A., and Van Gent, A.L. (Eds.) (2007a) *Road Safety Performance Indicators: Country Comparisons*. Deliverable D3.7a of the EU FP6 project SafetyNet.

Vis, M.A. and A.L. van Gent (Eds.) (2007b) *Road Safety Performance Indicators: Country Profiles*. Deliverable D3.7b of the EU FP6 project SafetyNet.

The proposed SPI can be implemented step by step, starting with the BAC of fatally injured drivers and gradually extending to a larger set of psychoactive substances used by all active road users involved in severe injury crashes. Chapter 3 details the successive requirements for each step as well as the method, measurement tools and quality control issues.

It is emphasized that ethical and privacy problems might arise when body fluids are collected from active road users. Hence, the use of data on impairment by psychoactive substances for judicial purposes should be discussed in the country as well as other ethical and privacy problems.

Institutional, judicial and other prerequisites for establishing the national system on alcohol and drugs' SPI are also discussed.

Speed

Chapter 4 contains a manual for guiding the planning and implementation of representative vehicle speed surveys. It is aimed at providing guidance to professionals, practitioners, and policymakers in the field of transportation and road safety for the planning, collection, processing, and analysis of vehicle speed data.

When setting up the survey, two issues should be considered: the purpose of speed measurements and expected data outcomes. The indicators have to be chosen according to road safety targets and the chosen methodology must allow computing these indicators. Chapter 4 details which locations are suitable for speed measurement; which road types should be considered; how the set of measuring locations can be sampled; which time periods are valid for speed measurements; how to determine speeds for different types of vehicles on the basis of identified requirements for speed measurements, and so on.

The Manual provides detailed recommendations on the selection of measurement sites (random procedure, what "appropriate location" means, etc); periods of measurement; practical considerations for the measurements (sample size, measuring devices, additional information required, etc); data control and analysis; documentation and reporting.

The speeds are analysed for free flow hours only, which are hereafter defined. The minimum set of speed indicators, for each road type, should include:

- Average speed for light vehicles during day
- Average speed for light vehicles during night
- Standard deviation of speed for light vehicles during day
- Standard deviation of speed for light vehicles during night
- 85th percentile of speed for light vehicles during day
- 85th percentile of speed for light vehicles during night
- Percentage of light vehicles over the speed limit during day
- Percentage of light vehicles over the speed limit during night
- Percentage of light vehicles 10 km/h over the speed limit during day
- Percentage of light vehicles 10 km/h over the speed limit during night

Protective systems

The SPIs developed in this field are:

– daytime wearing rates of seat belts in front seats (passenger cars + vans /under 3.5 tons), in rear seats (passenger cars + vans /under 3.5 tons), by children under 12 years old (restraint systems use in passenger cars), and in front seats (HGV + coaches /above 3.5 tons); and

– daytime usage rates of safety helmets by cyclists, moped riders and motorcyclists.

The SPIs are estimated by means of a national observation survey, where the measurements should be preferably classified according to main road types, such as motorways, other rural roads and urban roads. The values for major road types should then be aggregated into one indicator (of each type) for the country.

The national system for producing SPIs on protective systems should fulfil the requirements as detailed in Chapter 5, which concern the sampling demands and procedures, data collection and processing, documentation requirements, evaluation and aggregation rules. Besides, calculation rules are suggested for estimating over time progress, e.g. the annual increase in the protective systems' usage rate, the conversion rate – a year-to-year decrease of non-use of the devices. Several calculation examples accompany the procedures and the requirements presented.

International or regional comparisons of protective systems' use rates are vital tools for recognizing deficiencies, setting priorities and stimulating efforts at the political level. Applying the common rules presented in this Manual allows production of reliable and accurate indicators that are comparable among the countries (regions).

Daytime Running Lights (DRL)

The DRL SPIs are defined as the percentage of vehicles using daytime running lights, where the value is estimated for different road categories and for different vehicle types.

A recommended system for producing national DRL SPIs is presented in Chapter 6. The system is based on a national observation survey of the DRL use. The Manual provides relevant details on the method of data collection (periods, procedures), definition of observation sites, data collected, sampling demands, rules for estimating SPIs (per road categories, vehicle types and in total), quality control and reporting issues. An example of calculating SPIs is provided for Hungary.

The suggested system of DRL SPIs enables to estimate the DRL usage rates at the national level. The system may serve as a background for both the purposes of countries' comparisons and along-time considerations on DRL-related issues.

It is stated that the background information on the DRL legislation is essential for a correct interpretation and comparison of the results. For example, comparing the countries' DRL usage rates it is reasonable to take into account whether the countries have a law/ regulation on obligatory use of DRL and if they do, since when.

In countries, where automatic DRL was introduced a long time ago (e.g. Sweden, Norway), according to expert estimates, current DRL usage rate is close to 100%, thus the DRL usage rate as a behavioural safety performance indicator does not have practical implications any more. In general, once the option of automatic DRL is introduced Europe-wide, the DRL indicators will lose their importance as an indicator of safety performance.

Vehicles (passive safety)

National vehicle fleet's crashworthiness and compatibility are suggested as SPIs in this field.

Monitoring these indicators is related to the accuracy of the national vehicle fleet database as a record of the vehicles that are actually on a country’s roads at a point in time. Additionally, there are minimum requirements for the level of detail contained within the database, to make calculation of the SPIs for vehicles possible, such as:

1) Provide detailed and accurate descriptions of vehicle makes and models.

2) Classify vehicles according to vehicle-types compatible with CARE definitions.

3) Distinguish between smaller (less than 3.5 tonnes) and larger goods vehicles, since these are significantly different when assessing their compatibility in collisions with passenger cars or vulnerable road users. These issues are detailed in Chapter 7.

Roads

SPIs for roads aim to assess the safety hazards by infrastructure layout and design. Two SPIs for roads were developed: the road network SPI and the road design SPI. Because there were no examples yet that used these SPIs, a case study was executed.

In the case study (Chapter 8), the SPIs were applied to a case study area in The Netherlands. The objective of this application was twofold. First, it illustrates the rather theoretical concept of the SPIs and second, it aims to provide more insight into the practical applicability of the method. The case study consisted of the following steps: collecting and processing of data; determination of a list of urban area connections that need to be assessed; comparison of the theoretically required road categories with the actual road categories, and calculation of the actual SPIs. For the calculation of the road design SPI, EuroRAP Road Protection Scores (RPS), that were already determined for the case study area, were used.

The case study proved that it is possible to apply the theoretical concept in practice. The SPIs provide insight into the safety quality of the road network as a whole and of individual roads.

However, the process of calculating the network SPI is quite complex and a large amount of data is needed (also in case of a sample). The SPI should be applied in a number of countries in order to investigate whether the SPI is generally applicable. Furthermore, although the results of the case study seem reasonable, evaluating the results of the case study in more detail is recommended, in consultation with the road authority. Besides, the study resulted in a list of issues that need further research, such as: a more detailed definition of types of urban areas; considering the areas' limitations in case of natural barriers or administrative borders; the assignment of the theoretically required connections to the actual road network, etc.

Applying the method in various countries is essential for a further evaluation and refinement of the method. Furthermore, the SPIs only assess the safety performance of rural roads and motorways. Future extension to and adaptation of the model to include urban roads is recommended.

Trauma management

Trauma management SPIs should estimate the speed and the quality of the post-crash care, both initial and further, in the country. Accounting for the limitations of data available in the countries, a minimum set of Trauma Management (TM) SPIs was developed, which refers to the availability of EMS stations; availability and composition of the EMS medical staff; availability and composition of the EMS transportation units; characteristics of the EMS response time and the availability of trauma beds in permanent medical facilities.

Recognising the importance of the trauma management indicators for road safety programs and for road safety policy-making in general, presently only a few examples of such indicators can be found in use.

In Chapter 9, two countries having more comprehensive data on trauma care are presented: Germany – on Emergency Medical Services (EMS) and Israel – on in-hospital trauma care (Trauma Registry). In each country, the data are collected and the performance indicators are estimated systematically, which could serve as a reasonable background for the application of trauma management SPIs.

The minimum set of the TM SPIs includes fourteen indicators which are estimated based on seven data items provided for the country. Data requirements and calculation rules for estimating the TM SPIs are presented in Chapter 9. Besides, general requirements for the data are detailed, including sampling rules recommended for estimating the average EMS response time.

The suggested TM SPIs are recommended for application as a minimum set for the initial characteristic of the trauma management system's performance in the country. However, it was emphasized that the suggested set of SPIs mainly enables characterization of the *potential* of trauma care system for the treatment of road crash injuries and not the *actual performance* of the system. To characterize the actual system's performance data from trauma registry and other hospital databases are required, whereas such databases are presently not accessible in the majority of countries.

In the future, with further development of combined road crash related databases, an extended set of TM SPIs with both figures on the availability of services and characteristics of quality of the treatment supplied, should be considered for application.

Contents

1 Introduction

GITELMAN[A], V., HAKKERT[A], S.

[A]TECHNION

1.1 Safety performance indicators for safety management

The 2003 European road safety action programme stated the target of halving the number of road fatalities by 2010 (EC, 2003). The programme encouraged Member States to adopt national road safety plans putting road safety at the top of their political concerns.

Many countries have developed and are currently enacting their national safety plans. The mid-term review (EC, 2006) indicated that the Member States' safety plans typically include the topics of strengthening compliance with traffic rules; improvements of passive and active vehicle safety; improvements of road infrastructure; strengthening the legislation on driver education, training and standards of fitness for driving; protecting and educating users at risk and vulnerable users. The actions' scope and priorities depend on the analysis of road safety problems in a specific country.

It is generally accepted that the safety plans and targets have to be periodically monitored, to verify the progress made and to adopt necessary changes based on recent trends observed.

Monitoring the progress, road safety is usually assessed in terms of accidents, injuries or their social costs. However, simply counting accidents or injuries is often an imperfect indicator of the level of road safety. Typically, accidents and injuries are only the tip of the iceberg, because they occur as the “worst case” of unsafe operational conditions of the road traffic system. At the same time, those managing road safety need to take into account as many factors influencing safety as possible or, at least, those factors they are able to affect or control. Hence, additional safety performance indicators (rather than accident/ injury numbers) are required to provide a means for monitoring the effectiveness of safety actions applied.

Safety performance indicators (SPIs) are seen as measurements which are causally related to accidents or injuries and are used in addition to the figures of accidents or injuries, in order to indicate safety performance or understand the processes that lead to accidents (ETSC, 2001). Besides, they provide the link between the casualties from road accidents and the measures to reduce them.

As expected, safety performance indicators can give a more complete picture of the level of road safety and can point to the emergence of developing problems at an early stage, before these problems show up in the form of accidents.

1.2 The set of SPIs developed

The place of SPIs in a safety management system was originally shown by LTSA (2000) and then by ETSC (2001). The original model defined the essential elements of a safety management system which were safety measures/ programmes, safety performance indicators (intermediate outcomes), measures of final outcomes (killed, injured) and measures of the social costs of accidents and injuries. This model allocated safety performance indicators on the level of intermediate outcomes but did not differentiate explicitly between SPIs and the deliverables of programmes/ measures.

The Road SPI Theory report (Hakkert et al, 2007) provided further methodological fundamentals for the SPI' development. A key point in the development of SPIs was that they should be able to reflect unsafe operational conditions of the road traffic system and therefore, be more general than direct outputs of specific safety interventions. In order to demonstrate a more general character of SPIs and their independence from interventions, the layer of "intermediate outcomes" was further split into "operational conditions of road traffic system" and "outputs" (from measures/ interventions). The SafetyNet concept of the place of SPIs in a safety management system is shown by Figure 1.1. Ideally, SPIs should reflect the unsafe operational conditions of the road traffic system and be sensitive to their changes.

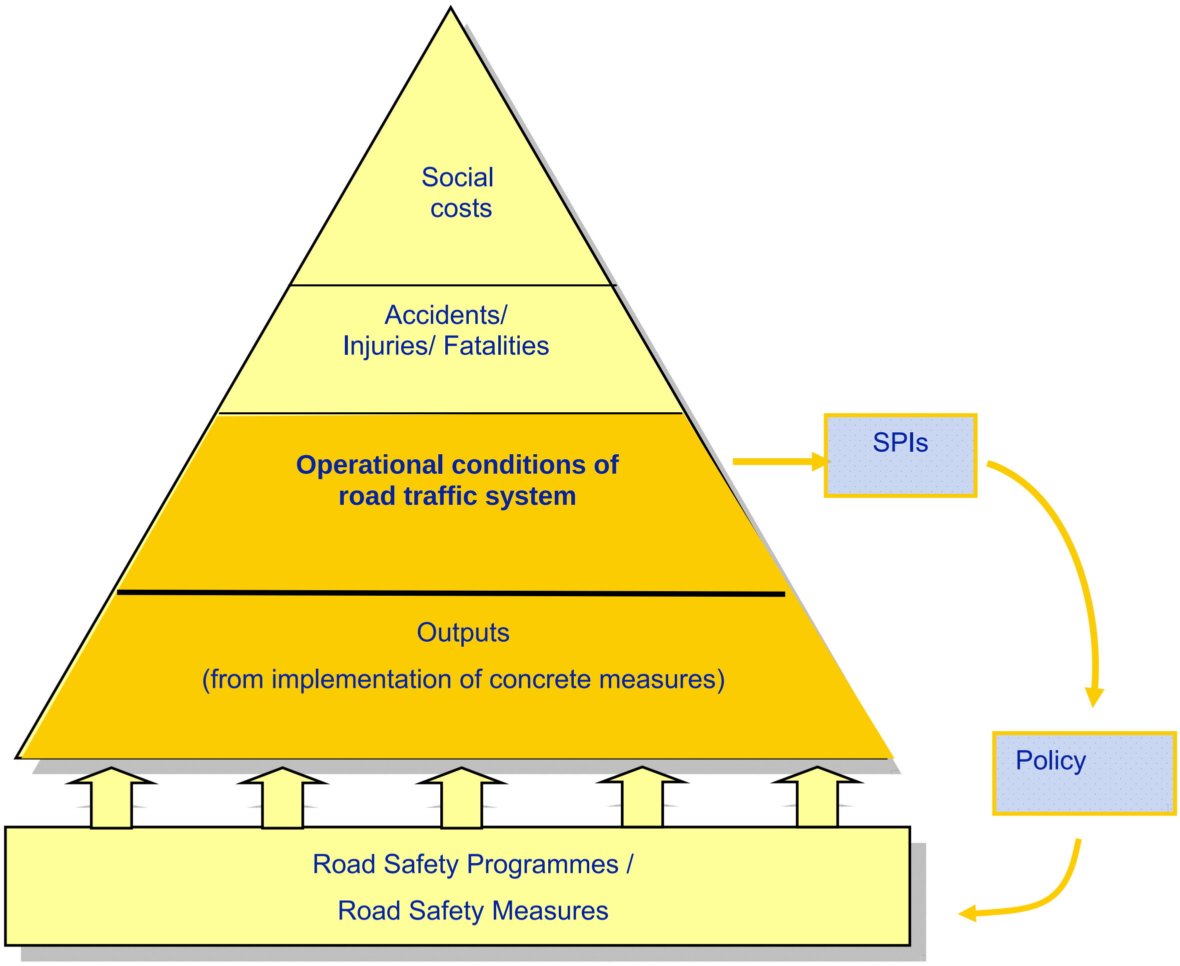


Figure 1.1: Place of SPIs in safety management system

For example, in the case of speeding, the unsafe operational conditions of the road traffic system (i.e. speeding) are affected by outputs from a road safety program or specific safety measures (e.g. speed enforcement). The outputs are the physical deliverables of the intervention (e.g. speed cameras in use), whereas the outcomes should be seen in improving the operational conditions (e.g. lower level of speeding), which can be measured by SPIs. The improved operational conditions will result in accident or injury reductions, whereas the whole process should reduce the social costs.

The SafetyNet definition of safety performance indicators is as follows (Hakkert et al, 2007):

Safety performance indicators are the measures (indicators), reflecting those operational conditions of the road traffic system, which influence the system's safety performance.

The SPIs' purpose is

- to reflect the current safety conditions of a road traffic system (i.e. they are considered not necessarily in the context of a specific safety measure, but in the context of specific safety problems or safety gaps);

- to measure the influence of various safety interventions, but not the stage or level of application of particular measures,
- to compare between different road traffic systems (e.g. countries, regions, etc).

According to Hakkert et al (2007), SPIs are developed for a certain safety domain (e.g. user behaviour, active vehicle safety, road infrastructure, etc), where they should reflect the factors contributing to road accidents/ injuries and characterize the scope of the problem identified. Developing SPIs should begin with a definition of the problem (the operational conditions of road traffic system which are unsafe and lead to accidents/ fatalities as the "worst case“) and continue with a conversion of this information into a measurable variable.

The SPI Theory report suggested a common procedure for the SPIs' elaboration, to make the process more consistent across different road safety domains. The procedure assists to verify whether a direct measurement of the recognized safety problem is possible, or indirect measurements are required, or even a further subdivision of the problem into sub-problems should be applied. In the worst case, when the measurement is possible only for outputs of certain road safety measures, the limitations of this consideration are clearly stated. This way, the procedure enables to bridge between the "ideal" and "realizable" SPIs.

Based on the potential of different road safety domains for promoting road safety as well as on the experiences and data available, seven problem areas were stated as central for road safety activity in Europe and, therefore, selected for the development of SPIs. They are:

a. Alcohol and drug-use
b. Speeds
c. Protective systems
d. Daytime running lights (DRL)
e. Vehicles
f. Roads
g. Trauma management.

The SPI Theory report (Hakkert et al, 2007) dealt with the theory behind the SPIs' development for each one of these seven areas. The report provided the rationale behind their development, the proofs for their relevance in the specific fields and the existing limitations that led to the adoption of specific SPIs. Within each field, the SPIs developed are directly related to that field of activity, can be quantitatively measured, can provide the basis for the assessment of the level of road safety in each country, can serve as an indicator to describe the level of activity in that field and can provide a yardstick for comparison.

The SPIs developed for the seven safety domains are as follows.

a. Alcohol and drug-use

Accounting for the limitations in the current state of accident data collection and data from surveys on the levels of alcohol and drugs in the driving population, three SPIs were proposed:

1. The number and percentage of severe and fatal injuries resulting from road accidents involving at least one active road user impaired by psychoactive substance (concentration above a predetermined impairment threshold);
2. The percentage of fatalities resulting from accidents involving at least one *driver* impaired by alcohol;
3. The percentage of fatalities resulting from accidents involving at least one *driver* impaired by drugs other than alcohol.

b. Speeds

The speeds that are most relevant for safety purposes are spot speeds measured at various locations on the road network during periods when traffic can be considered free flowing.

The SPIs developed are the mean speed, the standard deviation, the 85th percentile speed and the percentage of drivers exceeding the speed limit (by 0 and 10 km/h).

These indicators should be segregated by road type, vehicle type, period of day and period of the week (week-days and weekends).

c. Protective systems

The major protective systems in vehicles that are relevant for the development of SPIs are seat belts for adults and for children, in various types of vehicles and the use of safety helmets by cyclists, moped riders and motorcyclists.

The SPIs developed are:

– daytime wearing rates of seat belts in front seats (passenger cars + vans /under 3.5 tons), in rear seats (passenger cars + vans /under 3.5 tons), by children under 12 years old (restraint systems use in passenger cars), and in front seats (HGV + coaches /above 3.5 tons);

– daytime usage rates of safety helmets by cyclists, moped riders and motorcyclists.

Measurements should preferably be classified according to main road types, such as motorways, other rural roads and urban roads. The values for major road types should then be aggregated into one indicator (of each type) for the country.

d. Daytime running lights

DRL SPIs are suggested in the form of the percentage of vehicles using daytime running lights. The general indicator is estimated for the whole sample of vehicles observed in the country. Similar values are calculated for different road categories and for different vehicle types.

The road categories to be considered are: motorways, rural roads, urban roads, and DRL-roads, where the term “DRL roads” implies the road categories where the usage of DRL is obligatory. The vehicle types to be considered are: cars, heavy good vehicles (including vans), motorcycles and mopeds.

e. Vehicles (passive safety)

For each country, the entire vehicle fleet database according to vehicle type, make, model and year of first registration, as it stood in a certain year (2003), was analyzed. The EuroNCAP score was used as an indicator of passive safety for individual vehicles. Within the national fleet, a EuroNCAP score was attributed to all eligible vehicle types and models. An average figure was then calculated for each year of vehicle's registration and weighted by the number of vehicles from that year present in the 2003 fleet. An overall average EuroNCAP score was then awarded for each country and together with the median age of passenger cars in the fleet, these two figures made up the SPI for each country.

In order to validate the SPI with real-world data, car occupant fatality rates in each of the countries were considered. The number of car occupant fatalities in 2003 for each country was divided by the number of passenger cars present in each 2003 fleet, to give a figure for the number of car occupant fatalities per million cars. The average EuroNCAP score for each country was corrected by the percentage of passenger cars in a country’s 2003 fleet, which were less than 10 years old. This figure for each country was then plotted against the car occupant fatality per million cars figure for each country.

f. Roads

Two SPIs for roads were developed, namely the road network SPI and the road design SPI. The road network SPI indicates whether the actual road category is appropriate given the urban areas that it connects. Connections between these urban areas are assessed by comparing the theoretically needed road category with the actual road category. The road design SPI is the percentage of appropriate current road category length per theoretical road category.

The road design SPI determines the level of safety of the existing roads. For the road design SPI the EuroRAP Road Protection Score (RPS) is used. The safety of a road segment is expressed in one to four stars. The number of stars depends on the scores on a number of road characteristics. The road design SPI is the distribution of stars per road category.

g. Trauma management

Trauma management refers to the system, which is responsible for the medical treatment of injuries resulting from road crashes. It covers the initial medical treatment provided by Emergency Medical Services (EMS), at the scene of the crash and during the transportation to a permanent medical facility, and further medical treatment provided by permanent medical facilities (hospitals, trauma centres). Better performance of the system is associated with shorter response time by EMS; higher level of the EMS staff; standardization of the EMS vehicles; adequate hospital trauma care.

Based on the analysis of data available in the countries, a *minimum set* of trauma management SPIs was developed. It includes:

Availability of EMS stations

- the number of EMS stations per 10,000 citizens and per 100 km length of rural public roads

Availability and composition of EMS medical staff

- percentage of physicians and paramedics out of the total number of EMS staff
- the number of EMS staff per 10,000 citizens

Availability and composition of EMS transportation units

- percentage of Basic Life Support Units, Mobile Intensive Care Units and helicopters/planes out of the total number of EMS transportation units
- the number of EMS transportation units per 10,000 citizens
- the number of EMS transportation units per 100 km of total public road length

Characteristics of the EMS response time

- the demand for EMS response time (min)
- percentage of EMS responses meeting the demand
- average response time of EMS (min)

Availability of trauma beds in permanent medical facilities

- percentage of beds in trauma centres and trauma departments of hospitals out of the total trauma care beds
- the number of total trauma care beds per 10,000 citizens

Furthermore, a *combined indicator* was developed to measure a country's overall performance for trauma management relatively to other countries.

1.3 The goal of this report

This report provides details on the procedures necessary to collect the required data for the estimation of each set of SPIs in each country. It is called a *Manual* as it should assist the countries to establish the necessary systems of data collection for producing national SPIs, in each one of the predefined safety fields, and to make them comparable on a European level.

For each safety area, the report defines quantitative SPIs, demonstrates existing practices for their measurements, provides best practice examples (when available), and details the procedures which are necessary to collect and process the required data for the estimation of the SPIs' set on a national level. In addition, the report provides a more general theoretical background concerning the sampling issues in estimating SPIs.

This report supplements other reports on the development and application of SPIs in Europe: the Road SPI Theory (Hakkert et al, 2007), the Road SPI Country Comparisons (Vis and van Gent, 2007a) and the Road SPI Country Profiles (Vis and van Gent, 2007b). As mentioned above, the first document deals with the theory behind the development of SPIs in each of the seven safety areas. The two other documents provide results on the data collected so far for each of the 27 European countries (the EU at the end of 2006, Norway and Switzerland) and the SPIs estimated, based on the data submitted by each of the countries.

The structure of this report is as follows.

Chapter 2 provides general theory on sampling issues which need consideration in providing nationally representative measurements and SPI values. Regarding the setting up of an SPI survey, the main questions considered are: sampling procedure to obtain a national sample; sampling size; sampling error; stratified sampling (combination into a single SPI by weighting) and representativeness of the results. More details on statistical issues in estimating SPIs and their confidence intervals are further given in the *Statistical Appendix*.

Chapters 3-9 provide detailed manuals on collecting the data and estimating SPIs for each of the seven problem areas. For each topic, the context for using SPIs is discussed. Then, existing practices of using SPIs are summarized and a detailed description of available best practices is presented. Finally, a system recommended for producing national SPIs is introduced. The system details the data to be collected, including sampling rules, measurement methods, devices if relevant, etc; data requirements; the rules for estimating SPIs; quality control and other issues which are relevant to the problem area considered.

Chapter 10 provides a summary and conclusions of the report.

A large number of researchers from the various countries participating in the project have collaborated on this document. The authors are mentioned in each of the chapters.

1.4 References

1. EC (2003). European Road Safety Action Programme. Halving the number of road accident victims in the European Union by 2010: a shared responsibility. European Commission, COM(2003) 311 final, 2 June.

2. EC (2006). European Road Safety Action Programme. Mid-term review. European Commission, COM(2006) 74 final, 22 February.

3. ETSC (2001). Transport Safety Performance Indicators. European Transport Safety Council.

4. Hakkert A.S., Gitelman V. and M.A. Vis (Eds.) (2007). Road Safety Performance Indicators: Theory. Deliverable D3.6 of the EU FP6 project SafetyNet.

5. LTSA (2000). Road Safety Strategy 2010. National Road Safety Committee, Land Traffic Safety Authority, Wellington, New Zealand.

6. Vis, M.A. and A.L. van Gent (Eds.) (2007a) Road Safety Performance Indicators: Country Comparisons. Deliverable D3.7a of the EU FP6 project SafetyNet.

7. Vis, M.A. and A.L. van Gent (Eds.) (2007b) Road Safety Performance Indicators: Country Profiles. Deliverable D3.7a of the EU FP6 project SafetyNet.

2 Sampling issues in estimating SPIs

WETS[A] G., MOONS[A] E., GITELMAN[B], V., HAKKERT[B], S.

[A]IMOB, [B]TECHNION

2.1 General

Considering the data collection methods for estimating SPIs, the seven SPI domains can be split up in two broad categories. Four areas: speeds, roads, protective systems and daytime running lights, all require observational techniques on which results can be based. For each of these areas a sampling frame should be defined. The remaining three SPI areas: vehicles, alcohol and drugs, and trauma management, mostly need national statistics and centrally registered data and are not necessarily based on observations. Therefore, different approaches are required for sampling the data in these two groups of SPIs.

The main idea of this Chapter is to outline a sampling process that can be applied in general and to make sure that the sampling occurs on a nationally representative basis, both in time and space. The rules and methods presented below are mostly applicable for the areas where SPIs estimation is based on observational surveys.

There are five main questions that can be posed regarding the setting up of a survey for safety performance indicators:

1. Sampling procedure to obtain a national sample
2. Sampling size
3. Sampling error
4. Stratified sampling: combination into a single SPI by weighting?
5. What if we only have partial results? Can this be expanded to a nationally representative sample? (i.e. *transformation rules*)

General theory of sampling which is applicable to the majority of SPIs developed is given in Sec. 2.2-2.4. Further details on the techniques which are recommended for estimating SPIs in different safety areas are presented in Chapters 3-9. For the statistical background for estimating final SPIs and their confidence intervals refer to the Statistical *Appendix*.

2.2 Sampling procedure

The main components to be determined when setting up a survey are: survey population, sampling unit, sampling design, survey instrument.

The *survey population* are road users, vehicles depending on the SPIs definitions (e.g. cyclists, children in passenger cars under 12, etc.), the total road network, etc.

The *sampling unit* may be individual, section of road, vehicle driver, etc.

Main types of *sampling designs* are as follows (combinations are possible):

1. Simple random sample: most pure form. Each element has an equal chance of being selected. In order to carry this out, a list of all elements in the survey population is required, such that a sample can be selected at random.
2. Stratified sample: the population is first divided in non-overlapping strata, after which a simple random sample will be determined in each stratum. Advantages, if the stratification is carried out properly, are: the variance is lower, while you obtain the pureness from random sampling; costs are often lower, because of administrative reasons. Correct results are to be found in the general sample if one respects the proportions within the strata. Suppose that one stratification variable is used, e.g. province. In a hypothetical example, suppose that there are 3 provinces and that they represent 50, 30 and 20% of the total country. If 50% from the general sample will be

drawn from the first province, 30% from the second and 20% from the third, then the results reflect those of the country in total.

3. Clustered sample: this is a simple random sample of groups (clusters) of elements. The advantage is that one does not need to have a complete list of elements of the population, only a complete list of the groups. It is often cheaper to carry out if one takes clusters e.g. based on geographical location.
4. Multiple stage design: draw a simple random sample of clusters, first, and afterwards a simple random sample of elements within that cluster. This is typically used when the clusters are too large to question all elements. E.g. people within a certain community. The advantages are the same as in a clustered sample. The clustered and the multiple stage design sample also share some disadvantages. If the clusters are too intra-homogenous, if there are too few clusters or if they are of very unequal sizes the variance might increase rapidly. In summary, one could say that clustered and/or multiple stage sampling is often a pragmatic solution, but it can turn out worse than a simple random sample or a stratified sample in terms of precision of the required results.

These are the most frequently used designs. The type of sample is often determined by the method of surveying. For interviews, e.g. very often a two-stage sample is used in order to have communities close to one another that can be surveyed. If the survey is carried out by mail or phone, a clustered sample or a (stratified) random sample seems the best feasible option.

The *survey instrument* may be paper and pencil (papi), interview, observational study, etc.

2.3 Sample size, sampling error and stratified sampling

2.3.1 Simple random sample

The minimal sample size depends on different factors such as the accuracy with which one wants to draw conclusions. If one is satisfied with rather general statements on the population, i.e. if one wants to know a certain population parameter only approximately, then a rather small sample size suffices. Based on calculations as proposed by Billiet and Waege (2001), some sample sizes based on a ***simple random sample*** will now be determined. Similar sample sizes are shown in Boyle, Dienstfrey and Sothoron (1998). For a more detailed sample size determination, the standard deviation or the variance of the sample or within a stratum of the sample is necessary, and this is very often not available. From classical statistics, it is known that the confidence interval for a *population proportion* is determined by:

$$p - z\sqrt{\frac{p(1-p)}{n}} < \pi < p + z\sqrt{\frac{p(1-p)}{n}},$$

with p the survey proportion, n the sample size and z the z-value of the desired confidence interval. Based on this calculation, one can determine a maximal deviation (md), with which the survey proportion p can deviate from the population proportion π:

$$md \geq z\sqrt{\frac{p(1-p)}{n}}$$

Isolating n leaves us:

$$n \geq \frac{z^2 p(1-p)}{md^2}.$$

Based on this formula, it can be seen that the sample size n depends on the survey proportion p, the accuracy with which we want to draw conclusions via the value of z and of the accuracy itself via md. If we take the most ‘safe’ case, i.e. p=0.5, then Table 2.1 gives possible sample sizes for some classical values of confidence and maximal deviation.

p=0.5		Accuracy/Sampling error (*md*)			
Confidence level	**z-value**	**0.1**	**0.05**	**0.02**	**0.01**
0.90	**1.65**	68	271	1 691	6 764
0.95	**1.96**	96	384	2 401	9 604
0.98	**2.33**	135	541	3 382	13 530
0.99	**2.58**	166	663	4 147	16 587

Table 2.1. Minimal sample size in function of confidence level and accuracy/sampling error

Smaller values of p will lead to smaller sample sizes, but because we are often not too sure about the survey proportion, this is the most 'safe' option.

For two SPI areas in particular (protective systems and DRL), the proportions usually take more extreme values. Therefore, we provide Table 2.2 for a proportion of 0.2 or 0.8.

p=0.8 (or p=0.2)		Accuracy/Sampling error (*md*)			
Confidence level	**z-value**	**0.1**	**0.05**	**0.02**	**0.01**
0.90	**1.65**	44	175	1089	4356
0.95	**1.96**	62	246	1537	6147
0.98	**2.33**	87	348	2172	8687
0.99	**2.58**	107	427	2663	10651

Table 2.2. Minimal sample size in function of confidence level and accuracy/sampling error

As can be observed, for the same sampling error and the same confidence level, the necessary sample sizes are much lower.

For a sampling error of 5% and a confidence level of 95%, 384 sample units seem a minimum in the safest option. Note that there is a difference between the sampling error and the confidence level, if we expect a proportion of e.g. 30%, and we allow it maximally to deviate between 28 and 32% then the sampling error equals 0.02, while the confidence with which one wants to draw conclusions can still vary between 90 and 99%.

Further statistical background on the issue can be found in the *Appendix "Statistical issues in estimating SPIs and their confidence intervals"*.

For an SPI that is not determined by a proportion, but e.g. by a mean value, a similar formula can be derived (see Groves *et al.*, 2004; Levy and Lemeshow, 1999 amongst others). A confidence interval for a *population mean* can be determined by:

$$\bar{x} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z\frac{\sigma}{\sqrt{n}}.$$

The maximum error margin with which the sample mean and the true population mean may differ from one another is given by:

$$md \geq z\frac{\sigma}{\sqrt{n}}$$

Isolating *n* from this formula leaves us thus:

$$n \geq \frac{z^2\sigma^2}{md^2}.$$

This formula can be used when the population variance is known. A general rule of thumb however, is that when the sample size is larger than 30, one can replace the population variance (σ^2) by the sample variance (s^2). This population/sample variance needs to be determined based on results from a prior study on the same population or from a small pilot study.

2.3.2 Non-simple random sample

The above series of formulas are not valid in case of ***non-simple-random samples*** (such as clustered samples/ stratified samples) and there is no universal formula suitable for all situations. In medical and health related surveys (clinical trials), one often has specific issues to deal with, such as exposure, randomization, matched pairs, etc. and many of the formulas concerning ***clustered sampling*** have only been derived for specific problems (see, e.g. Hayes and Bennett, 1999; Aliaga and Ren, 2006). The correct formula depends on the sampling and weighting procedure and various formulas and methods for estimating the variance of the sample mean. Once the variance of the estimate is known, the standard error is then the square root of that variance and it is customary (see Dorofeev and Grant, 2006) to express the standard error of this sample mean in terms of the effective sample size (n_e) and/or the design effect (*DE*). This design effect is expressed as the ratio of the correct sample variance, *V* (of the sample mean) over the sample variance of a simple random sample, V_{random} (=s^2/n)

$$DE = \frac{V}{V_{random}}$$

while the effective sample size is equal to the sample size divided by the design effect:

$$n_e = \frac{n}{DE} = \frac{nV_{random}}{V}.$$

Depending on this design effect, the effective sample size can thus be smaller or larger compared to a sample size under simple random sampling. For a clustered sample, e.g. this design effect equals:

$$DE = 1 + (b-1)\rho$$

with *b* the number of elements in each cluster and ρ the intra-class correlation coefficient. When clusters have different sizes, the weighted average cluster size (*b'*)

$$b' = \frac{\sum_{i=1}^{m} b_i^2}{\sum_{j=1}^{m} b_j}$$

is used instead. Assuming that there are *m* clusters in total $x_{i,j}$ being the *j*-th element in cluster *i* and b_i being the number of elements in cluster *i*, then the intra-class correlation coefficient (ρ) equals:

$$\rho = \frac{\sum_{i=1}^{m} b_i^2 (\overline{x}_i - \widetilde{x})^2 - \sum_{i=1}^{m}\sum_{j=1}^{b_i} (x_{i,j} - \widetilde{x})^2}{\sum_{i=1}^{m} (b_i - 1) \sum_{j=1}^{b_i} (x_{i,j} - \widetilde{x})^2},$$

where $\overline{x}_i = \left(\sum_{j=1}^{b_i} x_{i,j} \right) / b_i$ is the mean value in cluster *i* and

$$\tilde{x} = \frac{\sum_{i=1}^{m}(b_i - 1)\sum_{j=1}^{b_i} x_{i,j}}{\sum_{i=1}^{m} b_i(b_i - 1)}$$

the 'intra-class' mean.

For more information on the calculation of these figures in other situations, we refer to Dorofeev and Grant (2006) or Kish (1965).

2.3.3 Stratified sample

Note that the above sample sizes are based on a simple random sample or on a clustered sample. Since ***stratified samples*** are often suggested in the description of the SPIs, a short discussion will be provided here on stratified sampling. Two major questions are raised: what should the sampling size be, when we use a stratified sample, and how do we combine the obtained results in each strata to one SPI value?

To answer the first question, it is necessary to decide on which level the conclusions need to be based and what the aim of the research is. If the aim is to have a sample that is as representative as possible, then the proportion of the different strata in the population and in the sample should be the same. On the other hand, if the aim is to have an optimal comparison between the strata, then an equal proportion of sampling units should be assigned to each stratum. If it is important to compare the different strata, then a sample size of at least 30 is required within the strata. This makes it easy to compare the strata, since starting from 30 sampling units, a normal distribution can be assumed and a two-sample t-test can be used to carry out the comparison between the different strata. If the number 30 is really not attainable, 20 is sufficient, however, then a non-parametric test should be used for comparison purposes. The total number of sampling units for a stratified sample matches the sample size that was required for a simple random sample at a certain level of sampling error and confidence interval. For example, suppose that one wants to have an optimal comparison between four different road types (1 stratification variable with 4 categories) and one wants to draw conclusions with an accuracy of 5% and a confidence level of 95%. The total required sample size requires that at least 384 sampling units are questioned. The fact that an optimal comparison between the different strata is required, demands that the requested sample exists of four equally sized strata, leading to at least 96 sampling units within each stratum. Since this number is larger than 30, a two-sample t-test can be used for comparison purposes.

If on the other hand, the focus lies on comparing countries at a national level, instead of at the level of the strata, very often one aims to have a sample that is as representative as possible. This means that the total number of sampling units are divided according the population proportions. This means that if we have four strata, and one of the four takes up about half of the total population, than half of the total number of sampling points should be located in that particular stratum. Once again, a rule of thumb that is often used as a minimum number of sampling units in a stratum is 30 to 50 (rule of thumb in stated preference research).

The results can easily be combined if a national level is aimed at. Since the sampling points are distributed according to the proportions that these strata occur in the population, we can just make one large database of the requested results in the separate strata, and this will provide us with the results at national level. If, however, an equal sample size for each stratum is used, weighting according to their population proportions is necessary. The above set of rules is merely to be seen as some easy-to-work-with guidelines. The exact sample size formula for stratified samples can be found in Levy and Lemeshow (1999). If one wants to determine the number of elements needed to be a 100(1-α)% certain of obtaining a stratified random sample (with *H* strata), an estimated mean that does not differ more than 100 x ε% (with ε being the sampling error) from the true mean $\overline{X}$, the formula equals:

$$n \approx \frac{\left(\frac{z_{1-(\alpha/2)}^2}{N^2}\right)\left(\sum_{h=1}^{H}\frac{N_h^2\sigma_h^2}{\pi_h \overline{X}^2}\right)}{\varepsilon^2 + \left(\frac{z_{1-(\alpha/2)}^2}{N^2}\right)\left(\sum_{h=1}^{H}\frac{N_h\sigma_h^2}{\overline{X}^2}\right)}$$

with σ_h the stratum variance, $\pi_h = \frac{n_h}{n}$, N the population total and N_h the stratum total in the population. As can be seen, this is not an easy-to-work-with formula since it requires more information about parameters of the distribution than is likely to be available or can be guessed with any degree of confidence. If one assumes a proportional allocation, the formula reduces to the form:

$$n \approx \frac{Nz_{1-(\alpha/2)}^2(\sigma_w^2 / \overline{X}^2)}{N\varepsilon^2 + z_{1-(\alpha/2)}^2(\sigma_w^2 / \overline{X}^2)}$$

with $\sigma_w^2 = \frac{\sum_{h=1}^{H} N_h\sigma_h^2}{N}$. This means that if the population size and its distribution over the strata is known, as well as the mean and variance of the variable under questioning for each of the strata, that we can determine the required total sample size. The allocation over the different strata is then equal to their population proportion.

An important issue, that should not be overlooked, is the sampling unit. It is not the case if one selects e.g. only 100 location sites to take a survey that the sampling error is equal to 10%. If the sampling units are the locations as such, this is true, while if vehicles are questioned at these sites, the number of vehicles questioned will determine the sampling error.

2.3.4 Non-response

A second issue that may not be overlooked is non-response. Depending on whether or not people are questioned, the response rate also needs to be accounted for. A major problem in collecting information these days is the growing number of non-responses. After all, when people are asked to participate in a survey, they are generally free to accept or reject that request. However, lately, the number of people rejecting to answer or handing incomplete information is growing larger and larger. This reflects a change in social behavior that has to be dealt with and accounted for in a proper way (Moons and Wets, 2007), meaning that either the list of people that one wants to contact needs to be at least 3 times as large as the required sample size, or one needs to deal with the missing items in an appropriate way (i.e. certainly not by ignoring them).

2.4 Transformation rules

It has to be emphasized that the use of transformation rules is not really advocated, in the sense that if it is possible to do a nationally representative survey, one should do so. The proper results then need to be determined based on a sampling procedure as will be outlined in the respective Chapters.

If one can be 100% sure that by examining only a few locations and aggregating these results (e.g. by taking the proportion of this type of location into account with regard to the whole national population), one can obtain a fully representative result for the whole country, this might be an option. However, it is unlikely that one can ever be 100% certain that this can be done. There are so many factors that might play a role (infrastructural characteristics, type of location, weather circumstances, etc.) that ensure that a particular location is not fully representative for this type of location throughout a country, that this option is usually not advisable.

When it is not feasible to carry out a national survey, a third option could be to apply the procedure as suggested in Greaves (2000). In his study Greaves uses a representative travel survey from one particular region (country, city, etc) and he updates this survey, based on important characteristics (e.g. socio-demographic information), in order to end up with a 'new', simulated survey from another region. The question then remains: which are the important characteristics that are necessary in order to transfer the results from one region to another. Some might argue that exposure and infrastructure are very important, others might think of density, degree of urbanisation, etc.

Timmermans (2006) summarizes the findings and discussions of a workshop on the topic of synthetic data, stimulated by the research paper by Greaves. It says that if one is willing to assume that the surveyed behaviour is transferable across time and space, then data collected in another area can be used to predict the behaviour of the target population in the area of interest. Given a set of variables that is assumed to influence aspects of the behaviour and (more important!) assuming that these variables are known for the target population, micro-simulation can be used to predict the behaviour for the target population. The degree of transferability depends on the stability of various aspects of the behaviour across time and space. The discussions in the workshop illustrate that although it seems feasible to develop a more data-driven approach to simulate (transfer) data from an area with lack of data, the simulation soon becomes quite complex, if increased realism is built into the process. One might save on data collection, but the application of the simulation approach (and hence the determination of the transformation rules) may still involve a considerable amount of time, and is not necessarily the same for every Member State.

2.5 References

1. Aliaga, A. and Ren, R. (2006) Cluster optimal sample size for demographic and health surveys. Paper presented at the 7th International Conference on Teaching Statistics, Salvador Bahia, Brazil.
2. Billiet, J. and Waege, H. (2001) *Een samenleving onderzocht – Methoden van social-wetenschappelijk onderzoek*. Standaard Uitgeverij.
3. Boyle, J., Dienstfrey, S. and Sothoron, A. (1998) National Survey of Speeding and Other Unsafe Driving Actions - Volume 1: Methodology. Technical Report DOT HS 808 748.
4. Dorofeev, S. and Grant, P. (2006) *Statistics for Real-Life Sample Surveys – Non-Simple-Random Samples and Weighted Data*. Cambridge University Press.
5. Greaves, S. (2000) Simulating Household Travel Survey Data in Metropolitan Areas. Ph.D. Dissertation, Louisiana State University.
6. Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2004) *Survey Methodology.* Wiley Series in Survey Methodology.
7. Hayes, R.J. And Bennett, S. (1999) Simple sample size calculation for cluster-randomized trials. *International Journal of Epidemiology*, Vol. 28, pp/ 319—326.
8. Kish, L. (1965) *Survey Sampling*. John Wiley and Sons.
9. Levy, P.S. and Lemeshow, S. (1999) *Sampling of Populations – Methods and applications – 3rd edition*. John Wiley and Sons.
10. Moons, E. and Wets, G. (2007) *Modeling Mode Choice in the Presence of Missing Values: How to Deal with an Often Overlooked Problem*. Paper submitted for the 86th Annual Meeting of the Transportation Research Board.
11. Timmermans, H. (2006) *Using Microsimulation to Generate Activity-Travel Data Under Conditions of Insufficient Data*. In: Travel Survey Methods – Quality and Future Directions, Stopher P. and Stecher, Ch. (eds.).

3 Detailed manual for SPIs on alcohol and drugs

ASSUM[A], T., HOUWING[B], S., MATHIJSSEN[B], R., HOLLO[C], P.

[A]TØI, [B]SWOV, [C]KTI

3.1 Introduction

Alcohol and drug use by road users, especially drivers of motor vehicles, increases the road accident risk considerably. Consequently, most countries ban the use of these psycho-active substances among drivers, or set low legal limits for blood alcohol and drug concentrations. Nevertheless, a high proportion of fatal accidents involve drink or drug driving in most countries. Road safety policy makers need information about the state of this problem in their countries.

The SPIs for alcohol and drugs show the state of alcohol and drug use as a risk factor in a country at a certain time. The SPIs for alcohol and drugs can be used by road safety authorities and politicians in assessing the needs for countermeasures such as legislation, enforcement, education and publicity. The SPIs for alcohol and drugs can also be used in evaluating countermeasures. When effective countermeasures against drink and drug driving are implemented, the SPIs for alcohol and drugs should demonstrate lower (improved) values. However, valid comparisons between countries and over time can only be made if definitions and methods used for estimating SPIs are the same across countries and over time.

Theoretically, the 'ideal' SPI of the alcohol and drug related road toll might be: the prevalence and concentration of impairing substances among the general road user population. In practice, however, major methodological problems are associated with this SPI, even when used within one country and when including only one psychoactive substance, namely alcohol. Problems will only increase when all EU countries will have to agree on a common sampling and testing protocol and when psychoactive substances other than alcohol will have to be included. Simultaneous random testing for alcohol and drugs is not only very expensive, but also time consuming which will result in relatively small road user samples. However, rather large samples are required to obtain reliable results, because the prevalence of psychoactive substances in the general driver population is likely to be low in statistical terms in most countries.

For all these practical reasons, focus on an SPI that is based on psychoactive substance abuse among crash-involved active road users is recommended.

Using an SPI that is based on accident data seems to be out of line with the basic idea of SPIs. However, unlike other risk factors, the road safety situation of a country with respect to the use of alcohol and drugs can be monitored very well directly from road accident data, since the size of this problem is clearly demonstrated by testing accident-involved road users for psychoactive substances.

Although such an SPI may be different from a theoretical point of view, it is more feasible in practice, especially if only severe and fatal injuries are included, because road users involved in severe and fatal injury accidents make up a much smaller number than road users in general. Moreover, this SPI is a direct indicator and can therefore be preferred above a general-population indicator. Initially, the SPI may even be limited to fatal injuries only to limit the number of road users to be tested, although it might be difficult then to evaluate annual developments, especially in countries with small numbers of road fatalities and/or a low prevalence of psychoactive substances among road users. When the number of fatal accidents is low, the random variation from one year the next increases.

The proposed SPI is:

The number and proportion of severe and fatal injuries resulting from crashes involving at least one active road user[4] who was impaired by psychoactive substance abuse (concentration above a predetermined impairment threshold).

The *number* allows assessment of the (societal) cost per country in absolute terms, but it is not suitable for comparison between years and countries, because it is very sensitive to factors, other than psychoactive substance abuse that influence accident and injury risk (like population size and traffic volumes, the state of roads and vehicles, the quality of trauma care, age distribution of road users etc.). For international comparisons proportions should be used, as is actually done in the country-comparisons report, even if proportions are also subject to variation due to other risk factors such as speed.

The proposed SPI can be implemented step by step, starting with the BAC of fatally injured drivers and gradually extending to a larger set of psychoactive substances used by all active road users involved in severe injury crashes. The successive requirements for each step would be:

1. Mandatory blood testing of all fatally injured drivers, for a fixed set of psychoactive substances.
2. Mandatory blood testing of all drivers involved in fatal accidents, for a fixed set of psychoactive substances (whether or not the drivers are killed or injured).
3. Mandatory blood testing of all active road users involved in fatal accidents, for a fixed set of psychoactive substances. (To note: breath testing can only be used for alcohol).
4. Extension of procedures mentioned under 1-3 to severe injury accidents, starting with testing severely injured drivers and resulting in testing all active road users involved in severe injury accidents.

Presently some countries have reached step 1, 2 or 3 above, whereas others have no relevant data or did not reply to the request for information. To be able to compute an SPI at this time, it is necessary to adjust the definition to the data available in at least some countries, i.e. focus on the impairment of drivers of motor vehicles, disregarding impaired pedestrians and bicycle riders, as very few countries have data for impairment of these road user categories.

3.2 Alcohol and drugs measurement methods and tools – existing practices

3.2.1 Examples of alcohol and drug measurement methods

There are no countries yet, that can provide data for calculating step four of the SPI. Some countries do not collect data on the road toll of driving under the influence of psychoactive substances, other countries collect data that can be classified in one of the four SPI steps.

Measuring drugs other than alcohol in such a way that the results will be comparable internationally and over time, requires international agreements on such issues as: which drugs to be included, which medium (blood, saliva, urine, etc) to be used, as well as which markers and analytical methods to be used.

Table 3.1 presents the classification of 25 EU member states plus Norway and Switzerland according to availability of data for alcohol and drugs' SPI.

---

[4] Active road users are all road users except for passengers.

	Alcohol	Drugs	Comments
Not known	Ireland, Luxembourg, Slovenia	Ireland, Italy, Lithuania, Luxembourg, Netherlands, Slovakia, Slovenia, Sweden, UK	--
No data	Malta	Denmark, Germany, Greece, Estonia, France, Latvia, Hungary, Malta, Austria, Poland, Portugal	Hungary has data for medicines, but not for illicit drugs.
Step 1: Fatally injured drivers	Spain, Portugal, Sweden, Norway	Norway	--
Step 2: All drivers involved in fatal accidents	Belgium, Czech Republic, Denmark, Germany, Greece, Estonia, France, Italy, Cyprus, Latvia, Lithuania, Hungary, Netherlands, Austria, Poland, Slovakia, Finland, UK, Switzerland	Belgium, Czech Republic, Spain, Cyprus, Finland, Switzerland	In Belgium only 20% of drivers involved are tested. Testing for drugs is incomplete in Switzerland and is likely so in several other countries.
Step 3: All active road users involved in fatal accidents	–	--	--
Step 4: All active road users involved in fatal or severe accidents	--	--	--

Table 3.1. Country profiles: SPIs on Alcohol and Drugs

3.2.2 Examples of alcohol and drug measurement tools

Traditionally, blood specimens are taken from drivers to measure blood alcohol concentration (BAC) and drug concentration. Taking blood specimens is cumbersome and costly, and the analyses may be time consuming.

Alcohol can be measured in the breath even for evidential purposes, a fact which makes testing for alcohol less costly and cumbersome.

However, taking a breath test may not always be possible from all active road users involved in accidents. Some road users may not be physically able to undergo a breath test. In this case passive breath testers[5] might be useful. In the case of severe or fatal injury a blood test is probably the best way to collect a biological sample.

When collecting specimens to detect impairment for psychoactive substances other than alcohol, both blood or saliva specimens can be used. Saliva samples are less invasive and less costly than blood samples, but on the other hand, they have a lack of sensitivity for some frequently used drugs like benzodiazepines and THC.

3.2.3 Possible alcohol and drugs parameters versus selected SPIs

Alternatives to the selected SPIs are discussed in Hakkert et al (2007). The most obvious alternative is the prevalence of alcohol and drug use among the general driver population or the general active road user population (all road user categories except passengers), i.e. the number of road users under the influence of alcohol or drug divided by the total number of road users (in the same BAC distribution category). However, prevalence in the general driver or road user population was rejected as alcohol and drug SPIs because of difficulties in defining a representative sample of the road user or driver population. Moreover, countries such as the UK and Germany do not allow random breath testing of drivers. Most countries do not allow random testing for drugs. Road-side surveys are carried out, but participation is then voluntary at least for drugs. The validity of the results may then be discussed, as road users under the influence of drugs are likely to refuse to participate.

3.2.4 Practical examples

For alcohol, the Czech Republic seems to be the country with the most complete data. In spite of the very low value of alcohol-SPI (the percentage of fatalities resulting from accidents involving at least one driver impaired by alcohol), the Czech authorities' representative insists that *all* drivers involved in fatal accidents are tested for alcohol. The Czech reporting practice as well as their very good results as to alcohol as a risk factor should be studied more closely as an example for other countries.

For drugs it is difficult to find a best practice country. All countries reporting statistics for drugs indicate more or less clearly that these statistics are incomplete.

3.3 Recommended system for producing national alcohol and drugs SPIs

3.3.1 Setting up a national system

Depending on the SPI step a country has reached, more or fewer improvements need to be made to reach a higher step or to reach the most feasible SPI.

The *first* step includes blood testing of all fatally injured drivers, for a fixed set of psychoactive substances. A medically trained person should take the blood specimen.

For practical reasons some drivers will not be tested, such as cases when taking the driver to a medical doctor or nurse for a blood specimen would take too long. Only when the driver is dead already at the accident scene, can the police be sure that it is a relevant fatal accident. When a person dies later on, taking a blood sample of the dead driver(s) will be both too late and impractical. Consequently, the SPI should in practice be limited to drivers dead at the accident scene in the *first* step and to drivers and road users involved in accidents where somebody is dead at the scene in the *second* and *third* steps. When severe injury accidents

[5] Passive breath testers are breath testers that can be held close to but not touching the mouth of the person and do not require active blowing. These breath testers indicate only if there is alcohol in the breath of the person or not.

are included in the *fourth* step, the problem of people dying within 30 days of the accident will be reduced, as these cases will presumably be defined as severe injury accidents anyway.

Amendments of the road traffic law may be needed in countries where alcohol and drug testing of drivers involved in fatal accidents is not mandatory.

The *second* step includes all drivers involved in fatal accidents. The requirements will be the same as on step 1, only now for all fatally injured active road users instead of drivers only.

The *third* step includes all active road users involved in fatal accidents, for a fixed set of psychoactive substances. Since testing is no longer restricted to fatally injured drivers or fatally injured active road users, saliva collection devices and breath testers could be used to collect biological samples. As will be explained in Section 3.3.2, however, there is a problem comparing drug concentration levels based on blood analysis and saliva analysis.

The *fourth* step is the extension of procedures mentioned under step 1-3 to severe injury accidents, starting with testing severely injured drivers and resulting in testing all active road users involved in severe injury accidents.

The extension of the injured drivers also makes it possible to include drivers who die in the hospital after the crash.

The third and fourth steps might be difficult to reach since significant law amendments could be needed. The costs of taking and analysing the extra blood specimens will also be considerable.

3.3.2 Method

For alcohol, blood testing may be used or evidential breath testing for surviving road users involved in fatal accidents.

The results of breath alcohol concentrations and blood alcohol concentrations are very well comparable. Fixed factors can be used for the transformation from one concentration to the other. However, in Europe different transformation rules are used: e.g. the Netherlands and the United Kingdom use a factor of 2300 between the blood and breath alcohol concentration, whereas most other European countries use a factor of 2100.

In order to avoid problems with comparability, the blood alcohol concentration (BAC) should at least be reported above or below a predetermined impairment threshold. This is in order to make international comparisons for certain levels, as legal limits vary from 0.0 to 0.8 in Europe. A level of 0.5 grams per litre blood could be used.

Another issue is the lack of sensitivity of saliva samples for benzodiazepines and THC. A country that uses saliva specimens for collection for drugs could have an under registration, compared to countries that use blood specimens. However, as the technique of analysing saliva is developed further, a transformation factor from saliva to blood is likely to be developed.

For drugs, a saliva specimen or a blood specimen may be used. All illicit drugs and relevant psychoactive medicinal drugs should be tested for. For drugs neither the types of drugs nor concentration limits are established so far. International agreements need to be achieved as to for which drugs to test and categories of test results.

Core substances may include: alcohol, cannabis, cocaine, opiates, XTC, amphetamines and benzodiazepines. In addition to these core substances the drug panel could be expanded to include other drugs that are nationally frequently used.

The biggest issue however, is the lack of uniform impairment levels for psychoactive substances in blood. For drugs other than alcohol research is conducted to try to determine impairment levels for blood.

Routines for registration of impaired active road users should be part of the existing road accident data registration. It is very important that samples should be collected as soon as possible after a crash occurs. In order to calculate the drug concentration at the time of crash, data are needed on:

- Time of crash
- Time of sample collection

Sampling issues do not exist in this context as, depending on the SPI-step, all drivers or all active road users involved in fatal or serious accidents are tested.

Calculating SPIs should be performed as described in Section 3.1.

3.3.3 Quality control

The most important quality aspect is the proportion of involved drivers tested for alcohol and drugs. In principle this should be 100%, but in practice this may be difficult to achieve. To be able to assess the quality of data, it is important to report both the number of drivers/road users who have and who have not been tested.

Routines to ensure that all motor vehicle drivers (active road users) included in all fatal (severe injury) accidents are tested need to be established. These routines will vary according to the implementation steps 1 – 4 described in Section 3.1.

Such a routine can be:

a. For each fatal accident, register the number of drivers/active road users involved.

b. Check that blood samples are collected for each driver/active road user.

c. Check that the results of the blood analysis are reported to the institution responsible for compiling the statistics.

d. Make a distinction between on-the-scene fatal accidents and accidents where a person dies later on.

e. The number of involved drivers tested and not tested as well as the reason for absence of test must be reported. It is most important to distinguish between drivers not tested and drivers negative for alcohol and drugs.

As mentioned above, for practical reasons some drivers will not be tested. Only when a person is dead at the accident scene, can the police be sure that it is a relevant fatal accident. When a person dies later on, taking a blood sample of the driver(s) involved will be both too late and impractical.

For alcohol, the BAC level should be reported, rather than just positive/negative answers (above/below legal limit). These values are important in order to make international comparisons for certain levels as legal limits vary from 0.0 to 0.8 in Europe.

For drugs, neither the number of drugs nor concentration limits are established so far. International agreements need to be achieved as to which drugs to test for, categories of test results, test methods and biological markers.

3.3.4 Analysis and reporting

Analysis of blood or saliva specimens from involved drivers/road users should be made by the national forensic laboratory or other authorized laboratory.

The statistical compilation and analysis of the data should be carried out by the national agency responsible for road accident statistics or by another competent and authorized agency. The SPIs should be produced and reported for each country for each year as a minimum. Results from each country should be reported to ERSO for international comparisons.

Preferably the SPIs should be reported for provinces within each country and for day and night time, day of the week and month. Time series/trends should be established when the SPIs have been reported for several years.

3.4 Discussion

Ethical and privacy problems will arise when body fluids are collected from active road users. The use of data on impairment by psychoactive substances for judicial purposes should be discussed as well as other ethical and privacy problems.

The police should ensure that blood, breath and/or saliva samples are taken from all drivers involved in on-the-scene fatal accidents. Police will usually collect such samples for judicial purposes. The police should report the results to the agency responsible for road accident statistics or produce and publish the SPIs themselves.

The national road transport authorities should present to the national parliaments a proposal for the legislation necessary to make the police collect the necessary samples.

The national forensic laboratory or nationally authorized laboratory should carry out the analyses of the samples. These laboratories across Europe should cooperate to establish international norms for analysis methods and routines and ensure that these norms are complied with.

The national agency responsible for road accident statistics should merge the alcohol and drug data from involved drivers into their road accident data base.

3.5 References

1. Hakkert A.S., Gitelman V. and M.A. Vis (Eds.) (2007). Road Safety Performance Indicators: Theory. Deliverable D3.6 of the EU FP6 project SafetyNet.

4 Detailed manual for SPIs on Speed

RIGUELLE[A], F., ARSENIO[B], E., EKSLER[C], V., GOLDENBELD[D], C.

[A]IBSR, [B]LNEC, [C]CDV, [D]SWOV

4.1 Introduction

4.1.1 Reasons for setting speed SPIs

Speed is one of the main causes of accident and has a direct influence on accident severity. The Transportation Research Board (1998) reports that speed has been found to be a major contributory factor in around 10% of all accidents and in around 30% of fatal accidents. Due to the massive character of speeding and inappropriate speeds, managing drivers' speed has therefore a high safety potential.

The effectiveness of the speed management policies, including enforcement or education campaigns cannot be evaluated only with accident data. The number of accidents is influenced by other factors than speed alone and must be studied over a long time period to avoid being misguided by random fluctuations. It is therefore important to produce statistics about speed itself. However, speed measurements are often used to evaluate the impact of local measures (i.e. the installation of a speed hump) but they do not give a general view of the extent of the speed problem as a whole. This is why the introduction of speed safety performance indicators is useful. These indicators aim to give straightforward and summarised information about speed for an entire territory (i.e. a country). A speed Safety Performance Indicator (speed SPI) is thus a useful tool for road design and managing purposes, and a decision-making support tool for policy purposes: it allows policy makers related to road safety to take better decisions and to evaluate the impact of those made. Speed SPIs are not only handy for the policy makers interested on road safety, but also for those interested in traffic management or environmental issues (e.g. higher speeds are usually associated with higher fuel consumption rates and increased pollutant emissions).

It is clear that before speed data can be used to support policy decisions, they should be representative, reliable, valid and precise enough. Therefore, there should be clarity on how this quality can be produced and maintained. In this respect it is helpful to consider the whole process of production of speed data and to set up standards or recommendations for the execution of specific steps in this process.

4.1.2 Contents and aim of this chapter

This chapter contains a manual for guiding the planning and implementation of representative vehicle speed surveys. Its primary objective is to provide guidance to professionals, practitioners, and policymakers in the field of transportation and road safety for the planning, collection, processing, and analysis of vehicle speed data.

This manual does not contain detailed technical standards or specifications, but it provides general guidelines and recommendations on how to set up and improve vehicle speed data collection methods. Specialists in the field of survey research or statistics are required to apply the general knowledge communicated by this manual to the circumstances and conditions of a particular country or region.

4.2 Speed measurement methods and tools

4.2.1 Existing methods and tools

Speed has been systematically monitored in many European Countries for various purposes such as:

- Tracking performance of the road network by road authorities
- Evaluation of the effectiveness of anti-speeding measures
- Estimation of traffic related environmental impacts
- Enforcement by the police
- Monitoring of driver behaviour e.g. by insurance companies
- Scientific studies e.g. on road safety measures' performance

Road safety practitioners and researchers make use of speed data in several ways, and for various purposes, such as to:

- Monitor the extent of speeding on selected roads in order to identify those with a high proportion of offenders and roads sections with extreme speed offenders
- Monitor the development of speeding over time in order to identify hours per day, months in a year or seasons in a year which show disproportionally high numbers of offenders
- Monitor the development of speeding over time in relation to the actual speeds enforced by the police (enforcement margins) and activities near the measured road type
- Monitor the development of speed distribution over time and identify time-frames exhibiting a deviant distribution with possible negative effects on road safety
- Monitor and analyse the relationship between traffic intensity and traffic speeds
- Monitor the proportion of heavy goods vehicle over times in order to study the possible connection to road safety

Speed data usually contains a large amount of different information, such as on vehicles, roads, seasons and time that can be disaggregated and analysed in various ways. Measurement systems typically produce both individual and aggregated data, but there are also systems in use, which primarily measure speed data on an individual basis.

This document will concentrate on tools that measure speed at a point or over a very short section of road, thus producing spot speed data. Speed can also be measured over a length of road (e.g. with ANPR - Automatic Number Plate Recognition - systems) or with the help of floating cars moving in the traffic flow. The speed of a restricted number of individual vehicles can also be assessed continuously with the help of on-board data-loggers. Various additional techniques of monitoring are also developing fast, including aerial photography, satellite tracking or even gathering of speed data through the mobile phones of road users (i.e. GSM tracking).

Knodler et al. (2005) distinguished between three types of devices for collecting spot traffic/speed data:

- hand-held devices
- in-road devices
- out-of-road devices

Further, a basic overview of the main devices' characteristics is provided. Only the most commonly used devices are described. For a more detailed overview and analysis of different in-road and road-side devices, one may refer to NHTSA (1996) or UTMC (2000).

Hand-held devices

Radar and laser guns are portable instruments that are handled by a human operator. They are flexible tools for the investigation of speeds since they do not require any permanent installation. However, the data can only be collected over relatively short periods of time due to the compulsory presence of a human operator.

The use of radar and laser guns is more efficient on less-trafficked roads, as it becomes impossible for the observer to monitor the speed on highly-trafficked ones. Technical problems, due to secondary reflections from other moving vehicles can also alter the measurements. Last, but not least, even if the starting cost is low, the overall cost of surveys with radar guns is high due to the manpower costs.

Example of use: Ireland

The National Roads Authority (NRA) in 1999 launched a biannual program of free-speed monitoring in order to evaluate the progress made towards national targets for speed and highlight the areas where improvement needs to be made. Radar gun devices were found as the most appropriate for this study requiring measurements of strictly selected vehicles done on different road profiles The person responsible for the measurement must judge if the speed of a vehicle is unaffected by other vehicles. (This means vehicle gaps of at least 200 meters). On urban arterial roads, speeds are measured early in the morning between 5h30 and 7h30. On other roads, it is between 9h30 a.m. and 5h30 p.m., from Monday to Friday (NRA, 2005).

In-road devices

Loop detectors are widely used traffic surveillance systems. They generally include a set of rectangular form wires embedded into the roadway. The loop is traversed by an electric current creating an electromagnetic field. The metallic content of a vehicle that passes over the loop disrupts the electromagnetic field. From the duration and nature of the disturbance, data on traffic volumes and occupancy can be derived. Speeds and classification data are not directly measurable by single-loop detectors but can only be derived from other parameters by means of algorithms. Several different methods exist but none of them allow single-loops to be as accurate as dual-loops detectors. Speed data collected with a single loop can thus have a low accuracy. A classification can only be provided if the traffic is free-flowing. If the speed is low and the headways are small, the loop is unable to differentiate between following vehicles.

Loops usually store data on traffic counts, vehicle classifications and speeds in the in-built memory. These data are downloaded periodically and stored in a large database. Often the data are recorded as grouped frequencies (e.g. the hourly number of vehicles in the speed class between 40-50 km/h, between 50-60 km/h, etc.).

Loops are cheap but their installation is expensive, requiring to stop the traffic and to cut the road surface. However, the costs are incurred only once. It is unusual for loops to be installed for a short term count. Maintenance costs average around 10% of the original installation and capital costs per year. Loops may suffer from problems of road repairs and movement of road surface.

Loops are used in many member states for the permanent monitoring of traffic and vehicle speed. But since they have been primarily employed for traffic monitoring purposes, these loops are often present in high-trafficked roads and on motorways rather than on a representative sample of the national road network.

Portable loops also exist that can be easily placed over the road for small periods. They can be installed at places where it is not possible to install loops. They have the same precision problems as classical loops.

An alternative in-road technology is the use of *axle detectors*. These detectors may be of different types: pneumatic, piezo-electric or quartz-electric. *Pneumatic tubes* are put across the roadway on its surface. Two hollowed cables are needed to measure speed, each one connected to a central box, measuring the change in air pressure inside the cable tubes. A vehicle rolling over the tubes compresses this air. The device measures the time difference between the change in pressure from the first and from the second strip. On the basis of this time difference and the known distance between both strips, the speed of the vehicle can be determined. A rough classification of the vehicle can be done as well provided that the headways between vehicles are not too small. As the tubes do not count vehicles directly but rather axles, a correction factor must be applied in order to get correct traffic information. These correction factors are based on knowledge of the typical traffic characteristics of different road types but they must be adjusted depending on the specificity of the road where measurements are carried out.

Tubes are not suitable for long periods of measurement because they are vulnerable and may be subject to vandalism. The tubes may be torn away by vehicles that speed up or brake over them. The technique is also sensitive to weather conditions (especially to cold and humidity).

Tubes must be installed each time data are collected and can break after short periods of operation. The pneumatic tubes are cheap, but it is the installation and retrieval costs which largely determine the cost of the measurements, not the length of time for which the equipment remains on site (DfT, 2001). Each installation and retrieval means the necessity to stop the traffic. The tubes can only be placed on stable roads. Furthermore, in order to get counts for single lanes in multi-lane situations, several tubes must be used.

Piezo-electric detectors may be in the form of tubes (for temporary use) or in the form of bars in the pavement fixed by epoxy (for permanent use). They work in a similar way to the pneumatic tubes. They are more accurate but also more expensive and wear out more quickly than tubes. Maintenance costs are minimal and are dependent upon the stability of the cable in the road. Piezo-electric detectors also suffer problems with bad weather conditions.

The technology of optical fibres using *quartz-electric effect* should reduce the problem of weather conditions. The quartz-electric effect from optical fibres is more stable than the piezo-electric effect. However, the optical fibre technology is still a developing technology and the quartz-electric detectors are far less spread than piezo-electric detectors.

<u>Example of use: the Netherlands</u>

Except on 100 and 120 km/h roads, speed is not monitored by a national body in the Netherlands. However, most provinces carry on their own traffic and speed monitoring by means of permanent measurement devices.

Indeed, traffic and speed in the Netherlands are mostly recorded by (visible or invisible) measurement loops attached to a data recorder which classifies data according to a pre-specified format. For example, the so-called Golden River Marksman M660 is a data recorder which is being used in the Netherlands but also in many other EU-countries (UTMC, 2000).

Data recorders are read out and set up again every 4-6 weeks by a person, or are read out by transmitting the data over a phone line. New developments are that data recorders are able to send data minute by minute to a large database by GRPS wireless communication technology. For example, in the province of Zuid-Holland, 18 measurement loops send their data every minute by wireless technology to a central database. Interestingly, besides the regular speed and intensity data, car-by-car gap data can also be delivered by this large, minute-by-minute growing database.

Table 4.1 provides an example of the structure of hourly speed data measured on rural roads in the Dutch province of Drenthe (Steyvers, 1995).

	Direction 1		Direction 2	
Speed (km per hour)	< 7,2 m	> 7,2 m	< 7,2 m	> 7,2 m
<60	Count	Count	Count	Count
60-70	Count	Count	Count	Count
70-80	Count	Count	Count	Count
80-90	Count	Count	Count	Count
90-100	Count	Count	Count	Count
100-110	Count	Count	Count	Count
>110	Count	Count	Count	Count

Table 4.1: Example of structure of speed data in a speed project on rural roads in the Netherlands (Source: Steyvers, 1995)

Out-of-road devices

Doppler-based microwave radars send a constant wave (24.5 GHz) which is rebounded by the surface of the vehicle. From the modified frequency, it is possible to calculate the speed and the type of the vehicle. They are placed along the roadway (usually fixed on existing poles). The microwave radars are relatively unaffected by the weather conditions and are thus often preferred to other types of radars. Another similar technique is *frequency modulated carrier wave radar* (FMCW) that is however more expensive than Doppler radar for similar performances. These systems are non-intrusive. It is usually not needed to interrupt the traffic to install them and they are thus easy to place on busy roads. If the traffic is not too heavy, several lanes can be measured with the same device. The radars are not sensitive to weather conditions. With these devices, one can only obtain a coarse classification of vehicles based on their dimensions.

LIDAR devices (light detection and ranging) work similarly to radars but on a different wavelength and a different type of wave. LIDARs use a laser wave and gather the reflected wave to get information on the encountered objects. Their main field of application is enforcement because of their very high accuracy. But their price is significantly higher than those of classical radars.

Passive acoustic devices measure speed on the basis of the noise of passing vehicles. They are not expensive but can be disturbed by low temperatures. Ultrasonic devices generate a beam and measure the time taken for the reflected signal to come back. The devices themselves are not very expensive but they must be mounted above the measured traffic lane what generates additional costs.

Passive infrared detectors detecting the heat emitted from an object are usually employed to identify vehicles or pedestrian waiting at crossings. They can also detect moving objects but are not accurate for moderate and high speeds.

Active infrared devices use the same principle as microwave radars but with infrared wavelengths. Smaller wavelengths made them more accurate than microwave, which is especially helpful to distinguish between vehicles types. The system is however more expensive and subject to errors in bad weather conditions.

Cameras are placed on a certain height above the roadway and film the passing vehicles. For a correct image setting two points are introduced with a known mutual distance. The device measures the time which is necessary for the vehicle to drive from the first to the second point. The speed of the vehicle is calculated by means of this time and the known distance. The vehicle length can be deduced as well.

Example of use: Belgium

Speed is monitored since 2003 on an annual basis by the IBSR (Belgian Institute for Road Safety) in order to have a nationally representative estimate of speed. Regional authorities also perform their own systematic traffic and speed monitoring on motorways and regional roads. Speed data by individual vehicles are collected by automatic roadside Doppler-based microwave radars at 150 measurement locations during one week in November.

From this system, individual vehicle data are available. That allows a fine analysis of the data, taking into account vehicle types, headways and traffic counts.

The measurement locations are chosen randomly from a network database. However, some minimal requirements are required for the location, such as being on a straight segment of road, far from junctions, from speed humps or speed enforcement cameras.

4.2.2 Possible speed parameters versus selected SPIs

Speed is a complex road safety phenomenon. Both excessive speed (exceeding the posted speed limit) and inappropriate speed (faster than allowed by the prevailing conditions) are important accident causation factors. While it is possible to study excess speed on a large scale (e.g. country based), it is impossible to do so for inappropriate speed. Indeed, inappropriate speed must be studied case by case, and it requires the knowledge of the specific road, weather and traffic condition. The speed SPIs aim at giving a synthetic view for a certain aggregation level (country, region) level. It is thus inevitable that they will only provide information on the excess speed problem but not on the more complex problem of inappropriate speed.

The way speed is measured will influence the type of indicators that can be produced. Speed can be measured at a point in the traffic stream which will produce spot speed data (i.e. data on instantaneous speed of vehicles at a point). Measurements that take place on a very short section of road will also give approximate measures of the instantaneous speed of vehicles. On the other hand, measures that take place over sections of road or with floating vehicle will give information on the journey time of vehicles over longer section of roads, also named travel speeds.

The link between spot speed indicators and the risk or the severity of accidents is widely studied in the road safety literature. Travel speed indicators are less relevant from a road safety perspective. It does not prevent travel speed indicators from being better for other purposes, like studying the performance of a road network. As Walters (2001) confirms, spot measurements are especially usable, from a statistical point of view, if the aim is, such as ours, to estimate parameters related to all traffic on a specific road network.

Individual measurements are not of interest to evaluate the speeding problem. Values for individual vehicles must be aggregated. An example of an aggregated indicator is “time mean speed” which is "the arithmetic mean of individual spot speeds that are recorded for vehicles passing an observation point over a selected time period”. “Space mean speed” is another widely used indicator but as Hall (2002) mentions, there are several definitions of it in the literature. One of these says that space mean speed is “the average time taken to cross a given distance” and is thus based on measures of travel speeds. But it is acknowledged that space mean speed can also be evaluated by computing the harmonic mean of spot speeds data. The more theoretically correct way to measure average travel speed on a network is to calculate space-mean speed. However, our purpose is slightly different as the SPI speed aims to assess the behaviour of the road users rather than the efficiency of the network. Moreover, for relatively uniform flow and speeds, the two mean speeds are likely to be equivalent for practical purposes (Hall, 2002). We thus chose “time mean speed” as an indicator rather than “space mean speed”. It is easier to understand and calculate and is already used in most EU countries. It is possible, through calculation, to move from time mean speed to space mean speed and vice versa.

Time mean speed is an indicator that gives information about the central tendency of the data. The distribution, range and dispersion of speeds are other valuable characteristics of speed data that may be computed. It is thus important to select the indicators that are relevant as safety performance indicators:

- Time mean speed (i.e. arithmetic mean of speeds)
- Standard deviation of speed
- 85th percentile of speed (commonly written as “V85”)
- Percentage of vehicles travelling over the speed limit (by 0 and 10 km/h)

The choice of four indicators does not mean that each country should only produce four SPIs. Indeed, the indicators have to be computed separately for different road classes and vehicles types. Section 4.4.2 describes more specifically what should be done. Also, it is preferable to produce separate indicators depending on whether speed is measured at day or at night (see Sec. 4.4.2 for details).

Figure 4.1 shows a speed distribution (frequency of vehicles as a function of speed) with the different indicators depicted on it.

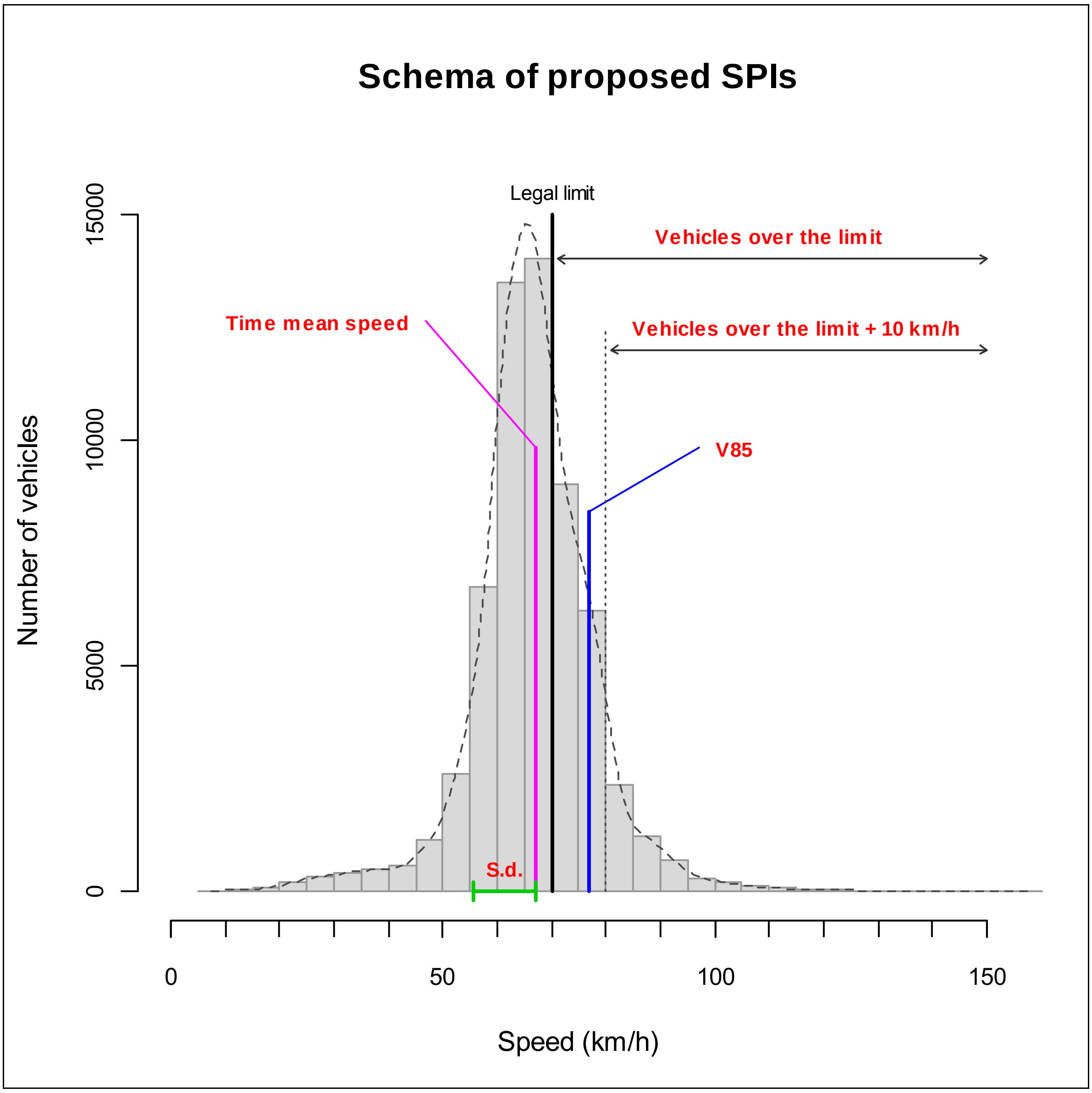


Figure 4.1: A speed distribution and the suggested safety performance indicators on Speed

4.3 Best practice example

Speed is monitored in most EU countries. However, various methodologies are being used due to the different purposes of the speed surveys. The French methodology is presented here as it is a good example of a systematic representative speed survey aiming at producing nationally representative indicators.

The Inter-ministerial Committee for Road Safety (CISR) is the decision making body for road safety in France. It defines the government's road safety policy and ensures its implementation. It was created in 1972 and comprises the ministers concerned with road safety.

The National Inter-ministerial Observatory for Road Safety (ONISR) is responsible for collecting and analysing road safety related data, including speed. It organizes the speed measurements in France for many years in order to collect information on average speeds and the proportions of infractions of several road user categories on several different road types. The survey is reproduced every four months. Since 2002, the speed measurements are coupled with measures of time between successive vehicles (i.e. vehicle headways). The execution of survey is entrusted to a specialised company.

Selection of measuring locations

Measurements are carried out on eight different road types that are defined according to the terrain morphology, the speed limit and the administrative status of the roads. Motorways, rural and urban roads are surveyed.

On the whole, speed is measured at 362 locations. There is no aim to produce regional estimates in France. Locations thus only constitute a representative sample of different road types at the national level.

All the locations selected had to conform to strict conditions in order to ensure free-flowing speeds at the place of measurement, such as:

- Straight and flat roads;
- No traffic perturbation;
- At least 1000 m from the nearest intersection or traffic signal;
- At least 5000 m from any enforcement radar.

Measurement procedure

About 50 surveyors are involved in the survey. They monitor speeds by means of mobile radars (type MESTA 208) that are either placed in a car or on a bridge over motorways.

Speed is measured only a couple of hours at each location but the sample is constructed so that all types of day and all periods of day between 9h30-16h30 and 22h00-02h00 are covered by the measurements. At each measuring site, speed is measured every four months at the same time of day and during a similar day of the week. 280 locations serve for day time measurements and 77 for night time measurements. Altogether, it represents around 2000 observation sessions and 200,000 individual observations per year.

Data analysis

Data are systematically split into three vehicle categories:

- Cars
- Motorcycles
- Heavy goods vehicles (with additional disaggregation)

The determination of vehicle types is done by human surveyors which are present at the location of each measurement.

For each of these vehicle types, day and night time data are analysed separately and different indicators are produced.

Bad weather conditions (rain, snow, hail, strong wing) are also analysed separately. In this way, the percentage of vehicles over the speed limit under bad weather conditions can also be computed despite the fact that speed limits vary depending on weather conditions in France.

Reporting

Reports are published by the ONISR at the same frequency as data are assessed: every four months. The publication of a periodic report occurs only about 2 months after the related four month period has finished. Moreover, the annual report on road safety in France contains an extended chapter concerning the speed surveys (ONISR, 2006).

In each report the previous results are presented and comments are made on the temporal evolution of speed indicators.

The average speeds and the proportions of vehicles over the limits, by different margins, are the indicators on which the reports are focussed.

Figure 4.2 shows an example of graphical reporting of the indicators. Average speed of cars for five different road types is plotted against time for the period 2002-2006. Since the end of 2003, the number of fixed speed camera has greatly increased, which might partially explain the big drop in average speeds on national roads even if the speed measurements are made far from any speed enforcement radar.

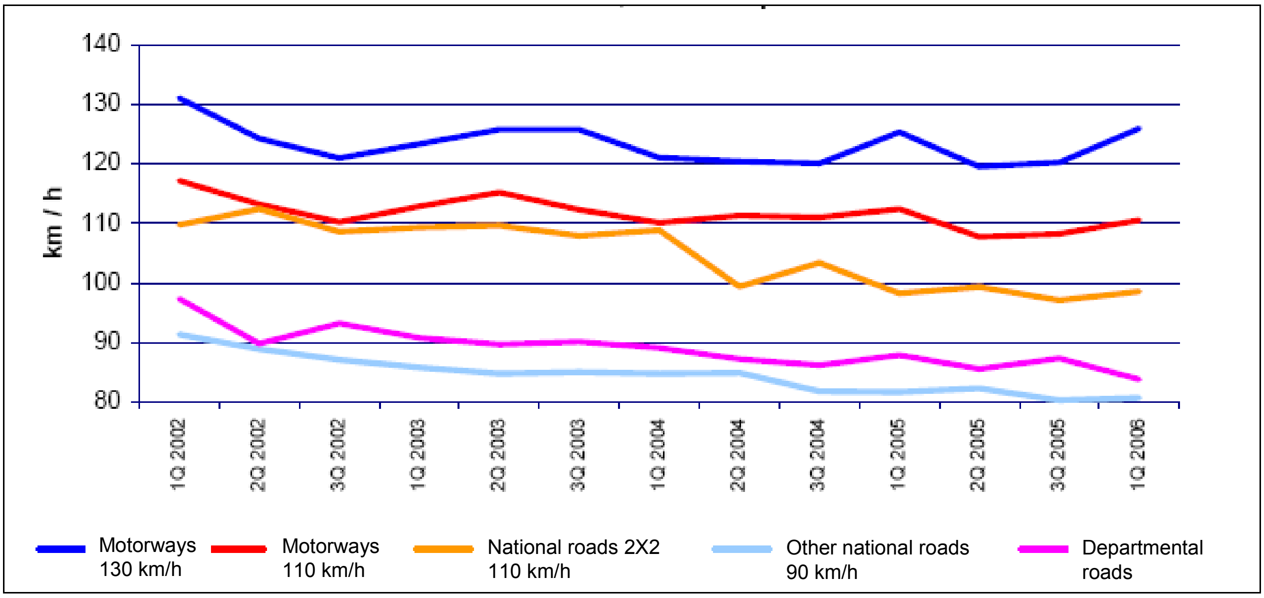


Figure 4.2: Evolution of the average daytime speeds of cars in France, in 2002-2006; bad weather conditions are excluded. (Source: ONISR, 2006)

4.4 Recommended system for producing national speed SPIs

4.4.1 Setting up a national system

In order to start a good and representative national speed survey program, it is necessary for this program to be backed by the national authorities and to take part in the national (and European) road safety programmes. The objectives and methodology of the program should be the object of a strong agreement for several years in order to avoid the existence and the methodology of the speed survey being frequently questioned and changed.

When setting up the survey, two issues should be considered: the purpose of speed measurements and expected data outcomes. The indicators have to be chosen according to road safety targets and the chosen methodology must allow computing of these indicators.

A good cost-benefit analysis should be carried out in order to assess what is the best equilibrium between the spending of resources and the amount of information that will be available through the survey. An important aspect is to ensure that the speed survey can be reproduced over several years on a regular basis (at least a yearly based assessment) so that temporal evolutions of speeding phenomenon can be analysed.

4.4.2 Methodology for data collection

The process of speed data collection must be carefully planned in order to produce reliable and meaningful indicators that could account for the high variability of speeds, both in time and space.

Section 4.4.2.1 explains which locations are suitable for speed measurement, which road types should be considered and how the set of measuring locations can be sampled.

All time periods are not equally valid for speed measurements. Aggregating data over periods with very different characteristics is certainly not meaningful. This is explained in Section 4.4.2.2.

Section 4.4.2.3 provides guidelines on how to determine speeds for different types of vehicles on the basis of identified requirements for speed measurements.

4.4.2.1. Where to measure speed?

The number of locations where one can imagine measuring speed is theoretically infinite. Indeed, the speed of a vehicle varies constantly along its journey. It is thus necessary to define a sampling procedure to help select a restricted set of locations that can produce speed data that are representative for bigger parts of the network. However, road characteristics are so variable that it is not meaningful to set as an objective computing indicators that are representative of the whole road network. Roads should display some specific characteristics in terms of road design to be suitable for speed measurements. Furthermore, different speed indicators should be computed for different classes of roads.

4.4.2.1.1 Road types

The speed at which a vehicle is driven on a road (operating speed) is highly dependent on the characteristics of this road, including its posted speed limit and its design speed. The relation between road design and operating speeds is especially complex to define. TRB (2003) lists several design variables that influence the operating speed, including the access density, median type, parking along the street, pedestrian activity level, curvature, grade, lane width and roadside developments. But even for a given design speed, TRB reports that there is a big variability in driven speeds.

Aggregating measures of speed taken on different roads with very different designs is obviously not meaningful. One must define separate road types based on their design characteristics and compute separate indicators for separate road types. Within the SafetyNet project the functional road classification has been proposed as presented in Table 4.2. It is recommended that countries arrange the analysis and reporting of national speed data according to this functional classification.

If available resources do not allow for developing representative samples of measuring locations for all nine road types, it is better to concentrate on producing good samples for selected road types rather than on insufficient samples for all road types. **It is recommended to measure speed at least on motorways (road type AAA), single**

carriageway rural roads (road types A and B) and single carriageway urban distributor roads (road type D). These roads carry a significant proportion of road traffic and exhibit uniformity in terms of design allowing more meaningful comparisons.

The definition of SPIs is further complicated by the fact that there is no one-to-one relation between speed limits and road category. For countries where there is more than one speed limit for one category of road, it is recommended to **compute the SPI for the dominant speed limit** (it is not convenient to aggregate data from roads with different speed limits). Along with the values of the SPI, information on speed limits should be provided. It is better to have both the SPI and the information on the speed limit rather than one single piece of information (SPI normalised by speed limit) which is much influenced by how the speed limits are set.

	SafetyNet road classes		**Functional road category**	**Separation of opposing directions**	**Lane configuration**	**Obstacle-free zone**	**Intersections**
Rural areas (outside built-up areas)	**AAA:**	Motorway	Through-road (road with a flow function)	Dual carriageway	2x2 or more	Very wide or safety barrier	Grade-separated
	AA:	A-level road 1		Dual carriageway	2x1, 2x2	Wide or safety barrier	Preferable grade-separated
	A:	A-level road 2		Single carriageway, preferable with lane separation	1x2, 1x3, (1x4)	Wide or safety barrier	Preferable grade-separated
	BB:	Rural distributor road 1	Distributor road	Dual carriageway	2x1, 2x2	medium	Preferable roundabout
	B:	Rural distributor road 2		Single carriageway, preferable with lane separation	1x2, 1x3, (1x4)	medium	Preferable roundabout
	C:	Rural access road	Access road	Single carriageway	1x2, 1x1	small	
Urban areas (inside built-up areas)	**DD:**	Urban distributor road 1	Distributor road	Dual carriageway	2x1, 2x2		
	D:	Urban distributor road 2		Single carriageway	1x2, 1x3, (1x4)		
	E:	Urban access road	Access road	Single carriageway	1x2, 1x1		

Table 4.2: Functional road classification

The SafetyNet classification may be difficult to apply in each country due to a specificity of roads. If one considers that two types of roads may be very different in terms of speed profile, the speed indicators for these road types should be computed separately even if the distinction is not made in the SafetyNet classification. When communicating about the speed indicators for people outside the country, some details should be given on road design of the roads considered and not only a general name of the road type.

4.4.2.1.2 Road design and environment

Road design characteristics and the surrounding environment influence speeds at which drivers operate their vehicles. In order to ensure reliability and comparability of speed data, the locations at which speed measurements are carried must be chosen carefully. All places where vehicles are likely to stop, accelerate or brake, should be avoided. Therefore, the locations should meet the following criteria:

- Straight and uniform section of road
- Section where it is possible to drive at a higher speed than the speed limit
- Section with a small gradient (<5% on at least 500 meters preceding)
- Away from junctions (>500 meters)
- Away from any speed calming device (> 500 meters)
- Away from road works (> 500 meters)
- Away from pedestrian crossings (> 500 meters)
- Away from any speed limit change or sign (> 1000 meters)
- Away from work zones, parking zones, important roadside developments
- Pavement surface in good condition
- Away from sections where speed used to be enforced by Police

Of course, speeding problems occur everywhere on the road network, not only on straight road sections far from any traffic perturbation. For example, insufficient slowing down on curves may dramatically increase the accident risk. However, it is not reasonable to collect speed data on curves in the framework of a nationally aggregated SPI due to specificity of each curve in terms of appropriate speed. Indeed, the maximum safe speed that may be driven on curves is different from the maximum safe speed on straight sections of the same roads. Driving at a speed equal to the speed limit is very dangerous in many curves! Therefore, no universal reference speed, including the speed limit, can be used to determine whether the driven speeds are dangerous in the case of a sample composed of different types of road profiles (curves, straight sections, vicinity of speed humps, etc), which prevents producing a meaningful aggregated speed SPI. On the contrary, only using generic straight roads away from traffic perturbations, we may assume that the speed limit can be used as a universally-relevant reference speed, allowing for the production of meaningful SPIs.

4.4.2.1.3 Selection of measurement locations: sampling procedure

The purpose of a speed survey defines how measuring locations should be selected. In several countries, traffic counters have been placed on major roads with the general purpose of monitoring traffic flows on major roads of the road network. Since these counters can also produce speed data, the speed measurements in several countries are based on these traffic counters. In this case, speed measurements are not based on a random sampling technique. Also, in some countries, the speed measurements are done to evaluate the effects of intensified enforcement activities. In these cases, speed measurements will be done on roads that were selected for enforcement purposes and that may be above average in terms of accidents.

If the purpose of speed study is different, such as **achieving a representative estimate of speeds driven on national roads**, the sets of speed measurements which were originally set up for monitoring traffic flows or monitoring enforcement success may not serve this purpose adequately. On the contrary, **the selection of measurement sites should be based on a random sample of roads.** The fact that measurements are made at randomly chosen spots of the road network will allow generalisation of the results to all traffic in that particular road network (Walters, 2001).

For countries that already have permanent counters installed, it may be not feasible to change the system completely. If counters are only on main roads, an option would be to randomly sample fewer sites but to take all of them on “non-main” roads. But then, a specific weighting procedure would be needed in the process of SPI estimation in order to take into account the respective share of main and “non-main” roads. More generally, it is not possible to infer an SPI that is representative for a road type if measurements are taken only on a sub-category of roads belonging to this road type. In this case, the SPI would only be representative for this sub-category of roads.

Units of analysis

Vehicles are the basic units of analysis. It is necessary in order to make comparisons between different vehicle types or different traffic conditions possible (e.g. vehicles in free-flow versus all vehicles).

Sampling units

We consider road segments as primary sampling units. The reasons for this choice are explained in Annex 1 to Chapter 4. **The sampling procedure consists of a random selection of road segments across the whole road network (first stage) and identifying appropriate locations on them (second stage).**

However, as we mentioned in section 4.4.2.1.1, separate indicators must be computed for different road classes. It means that **for each road class, distinct independent samples should be drawn**.

Sampling Frame

Ideally, **the sampling procedure should comprise a selection from a database** consisting of a list of uniform road segments, with their geographic coordinates and their characteristics such as:

- Road type (e.g. Table 4.2: motorway, rural road, urban road...)
- Actual speed limit
- AADT (Annual Average Daily Traffic)
- Number of lanes (not including additional lanes at intersections)

Additional useful information is:

- Type of median provisions (median divided, flush median, no median)
- Surrounding environment (inner city, outer suburbs, extended shopping area)
- Road design characteristics (slope, curvature, etc)

Good databases on the road networks exist in the majority of member states. Indeed, road databases are required by a wide range of applications (starting with vehicle navigation systems), which encouraged several private companies to start in the business of digital data gathering and commercialising. Even if those digital network data can be quite expensive, **it is recommended that each body involved in speed surveys works towards the establishment of such a database** in order to have a good knowledge of the road network and be able to select measuring locations in a rigorous way.

Sampling procedure

Usually, road network databases are already divided in segments of relatively small length (typically one segment per portion of road between two intersections). In this case, **the sample of measuring locations should be selected via a simple random sampling with probabilities of inclusions that are proportional to the length of the segments** (e.g. the likelihood of a segment of 2 km to be included in the sample is twice higher as the likelihood of a 1 km segment). If the initial database is not constituted of sufficiently small segments of roads, one should split the initial segments into smaller ones and repeat the operation described above. A simple random sample with unequal probabilities and without replacement is a common type of sampling that can be applied in common statistical software.

Once the road segments have been sampled, it is necessary to conduct field work before the measurements take place. **On each segment a profile must be found that corresponds to the suitability criteria listed in section 4.4.2.1.2, i.e. allows an unobtrusive measurement of speeds and does not compromise safety of the measuring device operators**. If such a profile is not available within the sampled road segment, one should either sample another road segment or (more pragmatically) search for a suitable location on segments that are contiguous to the segment that was initially sampled. Contrary to the sampling of the road segments, the precise selection of the measurement spots is thus not random. It is indeed prevented by the impossibility to have a complete list of the suitable locations in advance. Even with the development of road inventories and geographical information systems that allow having more and more information on the characteristics of roads, it is still preferable to conduct the field work. This small investment usually turns out to be profitable by avoiding getting bad and unusable data from some measuring locations.

Ideally, measurements should be made in both directions on each sampled road segment. But it is better to concentrate on accurate measurements for one single direction rather that producing coarser data for both directions.

It is recommended to define a sample of locations and use these over a period of several years, in order to get more reliable and accurate speed SPIs for time-series analyses.

As an aside: Some countries are interested in the production of regionally representative estimates of speed in addition to national scale indicators. In this case a stratified sample (with the regions being the strata) should be drawn and a minimum number of measuring locations per region fixed. In each region, a simple random sample (still with unequal probabilities of inclusion) would be drawn. This stratification of the sample has an impact on the process of computation of the national SPI (section 4.4.3)

Sample size

The sample size is determined by the maximum margin of error that one is willing to accept and by the desired confidence interval on the estimates (typically 95%). Details can be found in Annex 2 to Chapter 4.

The sample size is directly dependent on the variance of the **population**. Since separate samples are taken for each road type and only straight segments of roads are considered, the variance of the population should be quite small. But it is possible that large variances are found in the **sample** due to the fact that the measuring sites are not chosen in accordance with guidelines on road design listed at Section 4.4.2.1.2. So, before increasing the size of the sample, it is better to check if the current sites fulfil all the requirements to be good measuring locations. **As a rule of thumb, 30 locations per surveyed road type should constitute an accurate sample if the locations are carefully chosen.**

In practice, as the measurement and analysis of speed data is likely to be the largest cost of a speed survey, the available resources will be very important in determining the number of sites where speed can be measured. The number of measurement points is directly linked to the overall cost of the survey. If it is not possible to measure speed on a satisfactory number of locations to obtain a reasonable degree of confidence, then it is recommended to concentrate the speed measurements on only the more important road classes rather than having too small sample sizes for more road classes.

4.4.2.2. Time of measurement

It is known that speed, such as the frequency of violations, is time dependent. Changing traffic and weather condition are two well-identified sources of temporal variations of speed. If the influence of these two factors is not taken into account carefully when scheduling a speed survey, different traffic or weather conditions would explain most of the variation between two speed indicators, hiding variations that are really due to voluntary change of drivers' behaviour. Furthermore, daily, weekly, and seasonal changes should also be taken into account while choosing the time when speed is measured.

The number of periods of measurement and the length of time during which it is possible to measure will vary much depending on the measuring technique that is used and on the available resources. The following section outlines recommendations on the choice of best measurement periods, namely within a survey context where only techniques to measure speed for small periods can be used.

The more data that can be gathered, the better, but it does not exempt one from taking temporal variations of speed into account in the analysis of data. Indeed, having a lot of data does not mean that all of them should necessarily be used for the production of speed indicators. Some data may have to be discarded and some other considered separately and used for the production of separate indicators. The following three sections come back to that in more details.

4.4.2.2.1. Traffic conditions

Traffic conditions have a significant impact on the speeds at which drivers operate their vehicles. Similarly, a vehicle's speed depends on the traffic density.

It is not possible to compare speeds for the same road types if these are related to different traffic conditions.

This is why, **when producing indicators aiming at evaluating speeds, only the vehicles in reasonably free flowing traffic conditions should be considered**.

In order to obtain measures of speed in reasonably “fluid” conditions, the first obvious thing to do is to **avoid measuring speed where and when congestion is most likely to occur**, meaning typically:

- During morning and evening peak hours (e.g. 7h30->9h30 and 16h00->19h00)
- Near local events (market days, sports events, etc)
- Near roadworks

If one wants to make very stringent selection of free-flowing vehicles, the presence of a human observer is needed on the spot of speed measurement. Indeed, there is an element of subjectivity needed to judge if headways are large enough to ensure that drivers can drive at the speed of their own choice. The presence of a human observer is only practical for short periods of measurement, typically associated with a survey using radar guns (e.g. the surveys carried out in Austria or Ireland).

A suitable method would be to use data on vehicle headways and only select those vehicles that have a significant headway to the vehicle in front. In practice, vehicles with headways of more than 5 to 10 second are typically selected.

However, the selection of vehicles on the basis of their headways implies heavy computations and needs the use of equipment that can generate headway data. A more pragmatic way to proceed is simply to discard from the data all periods when the traffic flow is too heavy. We recommend **excluding all hour periods with more than 600 vehicles per hour and per lane.**

Extremely congested traffic may also result in small traffic counts. Although it is unlikely that extreme conditions of congestion would cause fewer than 600 vehicles to pass by the measuring site during an hour, these extreme periods can be removed from the data if an error analysis is made, by identifying periods with abnormally low speeds.

4.4.2.2.2. Weather conditions

Weather conditions have a significant impact on speeds. To be able to compare speeds from different places or countries, speed data should only be considered if they have been recorded under weather conditions that have a small speed reduction effect.

The following weather conditions should be avoided:

- Moderate and heavy rain
- Snow
- Freezing
- Fog and other circumstances of bad visibility
- Strong wind

So, if speed can only be measured for short periods, it should not be planned in winter because weather conditions are more likely to prevent valid speed measurements. Periods of poor weather conditions occurring during the measurement should be written down and relevant data removed in the process of data analysis.

With permanent measurements, it is inconvenient to take weather condition into account if there is no automatic weather monitoring system implemented which produces data that are easy to couple with the speed data. However, it is not a well-spread practice at the moment in Europe. If it is not possible to take the above into account, at least extreme weather events should be removed from the data.

Some weather conditions (e.g. rain, snow) are very common in a number of Member states and these countries may want to have speed data for these periods. For methodological research, it might also be important to compare speeds under different weather conditions. It is not a problem that speed is also monitored under bad weather conditions as long as resulting data are not mixed up with data of speed under different weather conditions. In this manner, speed data collected under good weather conditions only can still be employed for applications, such as international comparisons, that require that sort of data.

4.4.2.2.3 Other temporal variations

Traffic and weather conditions do not account for all temporal variations. Speeds also have more systematic variations e.g. between years, months, days and within the day. However, even if it is better to produce indicators more regularly, indicators are often produced for a whole year on the basis of only short periods of measurement within the year. The continuous temporal variations in speeds implies that, in order to get data that are representative for the whole year, those short periods of measurement have to be chosen very carefully.

On a yearly basis, measurements should be taken in a month that is "neutral" as far as a seasonal variation in traffic is concerned, meaning avoiding both school and bank holiday periods and the winter period (due to a risk of bad weather). **Late spring and early autumn** are thus recommended.

Furthermore, speed varies according to the day of week, especially between the week-end and weekdays. It is recommended to concentrate **measurements on typical working days (Tuesday, Wednesday and Thursday)** and to avoid market days and early-closing days.

Even more important is the variation between **day and night**. Comparisons between day and night usually reveal significant differences in speeds not only because of the variation in lightness, but also due to the difference in traffic conditions and in the composition of the population of drivers between the two periods.

For a sole country, indicators that do not distinguish between day and night measurements may be compared for similar periods of different years if the share of measurements between day and night has remained constant. Still, these indicators are less meaningful than separate day and night indicators.

On the other hand, in the framework of international comparisons, it is very unlikely to find two single countries with the same share of measurements between day and night. Thus, **specific indicators for day and night are required**.

If the resources do not allow making measurements both during the night and during the day, one should concentrate on day measurements. Taking into account constraints due to traffic conditions, day measurements should be made between 9h30 and 15h30. Night measurements should be done between 22h00 and 6h00. However, practically, in order to obtain enough measurements in a decent interval of time, it is not recommended to start measurements in the second half of night.

Inside the recommended periods for speed measurements, it is better to take readings at different times of day and on different days of week (to vary between mornings and afternoons and between Tuesday, Wednesday and Thursday). Each set of measurements should last long enough to allow **at least 200 vehicles to be monitored**. A sample of 200 vehicles would normally give an estimate of the 85 percentile speed for that period with a 3% accuracy of estimate (e.g. 65 km/h ± 2.0 km/h) at the 95% confidence level (DfT, 2001). Among these 200 vehicles, it is very likely that there is a huge majority of light vehicles (cars and light vans). If one wants to make speed estimates for heavy goods vehicles or motorcycles, the measurements should be lengthened even if the target of 200 vehicles (of this type) is not reachable. During night, it can be impossible to reach the 200 vehicle target on some less trafficked roads.

4.4.2.3. Vehicle types

Different types of vehicles have different speed characteristics and show diverse shares in traffic between roads and countries. Some heavy duty vehicles may also have different speed limits than the rest of the vehicles. Thus, an aggregation of all vehicles is not meaningful. The **SPI should be computed separately for different types of vehicles** in order to be comparable.

The **priority is to produce indicators for light vehicles (passenger cars and vans) only**. Additionally, indicators about the speed of heavy goods vehicles and of motorcycles may be computed.

The way to distinguish between vehicle types depends on the measuring technique. With radar guns, a human observer is present, allowing an easy and accurate categorisation of vehicles. Most widespread automatic speed monitoring techniques (loops, tubes, radar classifiers) are less precise. A classification of vehicles can only be obtained by indirect measurements:

- Pneumatic tubes give information on vehicle lengths, number of axles and sometimes axle loads (based on the pressure on the strips).
- Inductive loops use algorithms based on the attended vehicle distribution, the computed speeds and the occupancy rate of the loops to classify the vehicles. The determination of vehicles types becomes coarse when the traffic flow is heavy, usually resulting in an overestimation of the proportion of long vehicles.
- Roadside radars determine the lengths of vehicles on the basis of time they stay in the beam of radar. The following relation is proposed between vehicle length and vehicle type (Table 4.3):

<2.00 m	Motorcycles and mopeds
2.00-6.00 m	Light vehicles
6.1-12.0 m	Small trucks
> 12.00 m	Heavy goods vehicles

Table 4.3: Relation between vehicle length and vehicle type

Fortunately, even the coarser classifications (by inductive loops or roadside radars) are satisfactory to distinguish light vehicles from others vehicles, at least when the traffic flow is not too heavy. (As explained in section 4.4.2.1.1, we advise against measuring speed under heavy traffic conditions anyways). In the case of devices that can only produce aggregated data (for periods of 1 hour for example), it must be ensured that they are programmed to split data by vehicle type before the aggregation of data. Use of correction rules afterwards on aggregated data, such as determining indicators for cars on the basis of indicators for all vehicles type is not recommended. It would not be possible to determine a correction rule that can adapt itself to the specificity of each road.

A specific problem with heavy vehicles is that these often have different speed limits compared with the situation for cars or light duty vehicles. Furthermore, different types of vehicles have similar length (buses, coaches, trucks) and may also have different speed limits. Devices that determine the vehicle type on the basis of vehicle length may thus classify vehicles with different speed limits into the same category. Based on the national situation, computation of indicators for 'long vehicles' on the basis of this kind of equipment may thus be not meaningful.

4.4.3 Calculating SPIs

Several steps are needed to compute synthesized speed SPIs from the raw data. We describe here a typical procedure assuming that the speed data have been collected in a coherent way across the country.

Step 1: Error control

Speed data, as is the case for all data, should be controlled for errors. Data that are obviously wrong should be discarded. It may happen that all the data coming from a measuring location should be left out of the analysis. More details on error control are given in Section 4.4.4.

Step 2: Selection of relevant data

As already stated, not all data are suitable for computing an SPI. Data that were collected during congested traffic conditions or during bad weather should be removed from the dataset. Then data that correspond to the desired SPI should be selected. If the purpose is to make an SPI for light vehicles during weekday day-hours, only those data that correspond to that period of measurement should be selected.

To facilitate the computation of the final SPI, it is better to measure speed for the same duration at each measuring location (e.g. if speeds have been monitored for 3 hours at most locations and for 4 hours at few locations, it is better to select only 3 hours of data also for these few locations). In this way, using weights to take into account the duration of measurements is avoided (see step 4).

Step 3: Computation of indicators for each location

For each measuring location the following indicators should be computed, for day and night separately:

- Average speed for light vehicles
- Standard deviation of speed for light vehicles
- V85 of speed for light vehicles
- Percentage of light vehicles over the speed limit
- Percentage of light vehicles 10 km/h over the speed limit

If the measuring technique allows for a good determination of vehicle types other than light vehicles, similar indicators should be computed for these additional vehicle types. The indicators which are required to produce the desired speed SPI are presented in Section 4.2.2. Any other indicator may be added based on specific needs of each country.

It is important that beside the indicators, information is kept on the number of observations that were used to compute them (i.e. the vehicle count). This information will be used to determine the influence that the indicator for each location should have on the final SPI.

Step 4: Computation of the SPIs

In order to produce SPIs, data have to be aggregated from the individual location level to the country level. What is evocated as ‘SPI’ can be a result of the aggregation of any of the indicators described in step 3. The procedure of aggregation is indeed the same for each of them. Separate SPIs should be computed for different road classes. The procedure below should thus be applied separately for each road type.

The aggregation must be done in a way that preserves the representativeness of each of the inferior aggregation levels. Consider that SPI_i is the value of the SPI for the location *i,* which is part of the *n* sampled road locations and that W_i is the number of vehicles that have been measured at location *i*. The national *SPI* can be obtained as follows (Equation 1):

$$SPI = \frac{\sum_{i=1}^{n}\left(SPI_i * W_i\right)}{\sum_{i=1}^{n} W_i} \qquad \text{Equation 1}$$

This is simply a weighted average of the value of the indicator for each location weighted by the vehicle counts (number of observations per location).

Special cases:

Equation 1 is only valid if the traffic has been measured for the same duration in each location, thus allowing a comparison between traffic counts. If speed has been measured for different lengths of time T_i at different locations, one must correct for this difference. The relation becomes (Equation 2):

$$SPI = \frac{\sum_{i=1}^{n}\left(SPI_i * \frac{W_i}{T_i}\right)}{\sum_{i=1}^{n} \frac{W_i}{T_i}} \qquad \text{Equation 2}$$

If a stratified sampling was implemented (e.g. in order to have speed estimates at the regional level), the aggregation to the national level is somewhat more complicated. Equation 1 can be use to calculate the regional SPIs but a different weighting procedure should be used to figure out the national SPIs. It must be done in a way that avoids regions with a road network carrying relatively less traffic influencing the final SPI value too much.

If the region where location *i* is located is noted *r* and the traffic performance in terms of million vehicle-kilometres (MVKms) for the road type of interest in this region *r* is noted L_r, equation 1 becomes:

$$SPI = \frac{\sum_{r=1}^{R}\left[\left[\frac{\sum_{i=1}^{n_r} SPI_{ir} * W_{ir}}{\sum_{i=1}^{n} W_{ir}}\right] * L_r\right]}{\sum_{r=1}^{R} L_r} \qquad \text{Equation 3}$$

The aggregation process thus consists of the computation of SPIs for each region and then of a weighted average of these regional SPIs using respective MVKms as weighting values. MVKms are used as weighting values instead of the length of road network in each region in order to ensure that regions that are carrying more traffic have more influence on the final value of SPI than regions where roads are less trafficked.

However, since the estimates on the number of MVKms driven by particular vehicles on considered road types per region are often not available, the length of roads ($Long_r$), multiplied by the regional average number of observations per site may be used instead as a weight, resulting in equation 4:

$$SPI = \frac{\sum_{r=1}^{R}\left[\left[\frac{\sum_{i=1}^{n_r}\left(SPI_{ir} * W_{ir}\right)}{\sum_{i=1}^{n} W_{ir}}\right] * Long_r * \frac{\sum_{i=1}^{n_r} W_{ir}}{n_r}\right]}{\sum_{r=i}^{R}\left[Long_r * \frac{\sum_{i=1}^{n_r} W_{ir}}{n_r}\right]}$$

Equation 4

4.4.4 Documentation requirements

All sample design, data collection, and estimation procedures used in country surveys must be well documented. At a minimum, the documentation must:

- For sample design:
    - o List the road classes that are surveyed
    - o Describe how the roads segments were assigned to road classes
    - o Describe the general procedure used for sampling (is there some stratification included?)
    - o If strata were created, define what methods were used for allocation of the sample units to the strata
    - o Explain how the sample size was determined
    - o Describe the tools (software, road databases) and the practical procedure used to sample the measuring sites

- For data collection:
    - o List all the measuring sites and the period of measurement for each of them
    - o For each measuring site, a document with all the related information should be produced. This document should include the exact locations and characteristics of each measuring site so that the site can be easily found back if one wants to reproduce the survey. Problems occurring during the measurements or elements that may have had influenced the measures should also be written down. An example of that kind of document is included in Annex 3.
    - o Describe what vehicles were observed.
    - o Describe the equipment used and its technical characteristics
    - o Describe the data recording procedures

- For SPIs' estimation:
  - Document the whole statistical methodology used for aggregation process
  - Describe the tools used for estimation (software, algorithms)
  - Ideally, also compute and document the margins of errors of indicators for a confidence interval of 95 percent.

4.4.5 Quality control

All the stages in the process from the implementation of a speed survey to the reporting of the final results can lead to errors or inadequacies. Main issues to be aware of are mentioned in this section.

Preparation of measurements

It must be ensured that complete requirement specifications are delivered to the body, which will be responsible for the measurements.

Back-up plans should be in place in case of problems during the measurements (e.g. vandalism, theft, bad weather). The monitoring budget should take that into account.

Having selected the possible measurement road sections on ‘paper’, it is essential that these be driven over prior to performing measurement to ensure that the road sections are indeed representative and practical for measurement and to allow the survey team to correctly document the characteristics of the road.

Before carrying out the measurement, one must ensure that all the people that may be concerned by the measurements are aware of them. It means that each municipality where a speed measurement is carried out should be informed. The same recommendation is valid for higher administrative levels. Similarly, the police should also be informed. It is important to ensure that the degree of enforcement level at the measurement locations is not changed to avoid upsetting the validity of measurements being made.

Measurements errors

Beside measurement errors due to bad choice of locations or time of measurement, technical mistakes can occur, e.g. when a physical mistake happens at one of the detectors (for example, a disruption when the temperatures are too low). They can also happen due to communication problems, e.g. when the system does not upload anymore to the server, or some problems in the measurement technique can lead to inaccuracy (stop and go traffic, extreme low speeds, etc). Furthermore, it must be ensured that the requested lane(s) of traffic is measured by the device (e.g. not measuring both directions of travel without identifying each measurement, when only one is required or both required separately).

To avoid most of these mistakes, the devices that will be used for a survey should be evaluated by comparing their measurements with those of other devices (including various techniques). Even if the companies that produce the devices always make those types of tests, it is better that an independent test would also be carried out. All the elements to which the measurements prove to be very sensible should be written down and particular attention should be paid to these elements when installing the devices.

The Highways Agency (1981) reports the following requirements in the case of roadside radar:

- The operating instructions for the device should be carefully followed, especially in relation to interference and calibration.
- Some meters can be adjusted to ignore vehicles travelling in the "wrong" direction. With most meters, however, the response to those vehicles should be minimised by careful aiming of the antenna or by reducing the range setting.

- A precise angle should be respected between the measuring device and the vehicle's travel direction in order to obtain accurate values. For Doppler Microwave radars, this angle is typically equal to 45°. With radar guns, the angle between must be kept to a minimum. The measured speed is only accurate when the angle is zero, and becomes progressively less than the actual speed if the angle increases. The magnitude of the error, based on the angle of observation, is shown in Table 4.4.

Angle	Correction factor
0	0
1	0,999
10	0,985
20	0,940
30	0.866
40	0.766

Table 4.4: Correction factors for angle of observation with a radar gun (Source: PIARC, 2003)

- Situations where the radar beam may be obstructed by parked cars should also be avoided.
- Except at very light flows, it is not advisable to measure the speeds of vehicles on the far side of a single carriageway road (and especially not on the far carriageway of a dual carriageway road).

Automatic equipment should be checked more than once a week to ensure it is continuing to operate correctly and has not been vandalised (DfT, 2001). To avoid vandalism, the roadside monitoring equipments should be strongly fixed (with chains) to a post and locked. One may advice (e.g. by a sticker) that the material is used for the monitoring of traffic but not for enforcement.

Speed measurement should occur unobtrusively. In a case study in the Netherlands, Oei and Goldenbeld (1996) found that speed measurements with radar were giving lower estimates than speed measurements with inductive loops because some drivers lower their speed when seeing a car with radar equipment.

Data collection

Problems may also occur in the process of transfer of the data from the measuring devices to the server of agency responsible for the analysis. Bugs may occur in the internal software of the device or it may be badly configured (e.g. bad rounding off or aggregation of data).

The data format should be compatible with the software that will be used to analyse the data. Many companies develop their own analysis software along with their measuring equipment. These software packages are not necessarily suitable for the analysis of type that is wanted. So, one may want to get the data in a specific format from the body responsible for measurements. Quality standards of delivery should thus be contracted with this body, including delivery of additional information for each measurement site (coordinates, picture, etc).

Statistical software that can handle a great bulk of data must be chosen for the analysis of data (e.g. individual vehicle data for one week of monitoring at one site may already greatly exceed the 65536 lines limitation of Excel).

In the process of data collection and analysis, each measurement site should be identified by an unambiguous indicator in order to avoid mixing up of the data.

Error check

An error check of the data should be carried out as quickly as possible after the measurements so that it will be less difficult to track back the cause of strange data. The check for errors should be systematic with error correction decisions/procedures.

One example of error detection is given by Goldenbeld and Schagen (2005). During a field experiment on the safety effects of mobile radar enforcement in the Dutch province of Friesland, speed data were obtained from speed measurement induction loops. Speed and length class of every passing vehicle was registered electronically per hour (24 hours a day, 7 days a week). Every month, the data was downloaded from the roadside data box, checked on minimal quality criteria, and forwarded for further analyses.

Special checks on possible errors in the speed measurement data were performed before the actual analyses took place. Per road section and per day, specially developed software checked first whether the 24-hour traffic flow deviated over six times the standard deviation from the average 24 hour traffic flow on that road section. If this was the case the data for that particular day was marked as possible error. A further automatic check verified for each day and for each road section whether the speed data approached a normal distribution. For days with more than 2000 observations, the Kolmogorov-Smirnov one sample test was applied and for days with less than 2000 observations the Shapiro-Wilk test (both tests described in Stevens, 1996). If the speed data did not have a normal distribution, again the data for that particular day was marked as a possible error. In consultation with the supplier of the data, the province of Friesland, it was concluded for most of these days whether the deviations were caused by a measurement error or not. If it was the case, the days were left out of the analysis. All together, for less than 5% of the days data were either missing or removed because of measurement errors.

Data aggregation and SPI computation

This stage is very prone to human errors such as loss of data, corrupted backup and errors in calculations. The original datasets should always be kept and backed up on separate server and supports.

When aggregating data, careful consideration must be given to the procedure and formula to be employed in order that each measurement has a correct influence on the final indicators' estimates.

4.4.6 Summary of recommendations

Selection of measurement sites

- Should be based on a random procedure
- Procedure: random selection of road segments across the whole road network (first stage) and identification of appropriate locations on them (second stage)
- Appropriate location means:
    - Straight and uniform section of road
    - Section where it is possible to drive at a higher speed than the speed limit
    - Section with a small gradient (<5% on at least 500 meters preceding)
    - Away from junctions (>500 meters)
    - Away from any speed calming device (> 500 meters)
    - Away from road works (> 500 meters)
    - Away from pedestrian crossings (> 500 meters)
    - Away from any speed limit change or sign (> 1000 meters)
    - Away from work zones, parking zones, important roadside developments
    - Pavement surface in good condition
    - Away from the sections where the speed used to be enforced by the Police
- Separate samples should be drawn for different road types
- Speed should be at least monitored on motorways (road type AAA), single carriageway rural roads (road type A or B) and single carriageway urban distributor roads (road type D).

Period of measurement

- Avoid measuring speed during congestion-prone periods: peak hours, local events
- Avoid measuring speed under bad weather conditions:
    - Moderate and heavy rain
    - Snow
    - Freezing
    - Fog and other circumstances of bad visibility
    - Strong wind
- Preferably measure speed during late spring or early autumn
- Concentrate measurements on typical working days
- For day measurements, measure between 9h30 and 15h30
- For night measurements, measure between 22h00 and 6h00

Practical considerations for the measurements

- Measure at least 200 vehicles
- Measure traffic count during the period of measurement
- Carefully follow the instruction on use of the device
- Document the maximum information on the measuring sites (including speed limit)

Data analysis

- Start with an error control of the data
- Exclude from the analysis all hour periods with more than 600 vehicles per hour and per lane
- Split the data for day and night periods
- Split the data per vehicle type
- Compute at least the following indicators:
    - Average speed for light vehicles during day
    - Average speed for light vehicles during night
    - Standard deviation of speed for light vehicles during day
    - Standard deviation of speed for light vehicles during night
    - V85 of speed for light vehicles during day
    - V85 of speed for light vehicles during night
    - Percentage of light vehicles over the speed limit during day
    - Percentage of light vehicles over the speed limit during night
    - Percentage of light vehicles 10 km/h over the speed limit during day
    - Percentage of light vehicles 10 km/h over the speed limit during night

Documentation and reporting

- Document all steps of the survey carefully
- Publish results at least annually

4.5 References

1. DfT [Department for Transport] (2001). Road safety good practice guide, Local Government and the Regions DTLR, London.

2. Goldenbeld, C. and Schagen, I. van. (2005). The effects of speed enforcement with mobile radar on speed and accidents. An evaluation study on rural roads in the Dutch province Friesland. *Accident Analysis and Prevention*. Vol 37, Issue 6, pp. 1135-1144.

3. Hall, F. L. (2002). Traffic stream characteristics. In Traffic flow theory: A state of the art report - revised monograph on traffic flow theory, N. H. Gartner, C. Messer and A. K. Rathi, Eds., Oak Ridge, Tennessee, Oak Ridge National Laboratory.

4. Highways Agency. (1981). Vehicle speed measurement on all purpose roads. DMRB Vol 5, Section 1, Part 4 - TA 22/81. UK.

5. Isaksson, A. (2003) Survey models for a vehicle speed survey. Linköping studies in statistics n°2, Linköping University, Sweden.

6. Knodler, M.A., Rothenberg, H. and Benavente, M. (2005). Spot speed study pilot training. Manual and presentation. Report 04E-G019-001. Amherst, Massachusetts Traffic Safety Research Program (MassSAFE), University of Massachusetts.

7. Oei, H-L, and Goldenbeld, C. (1996). Snelheidscampagne in Zuid-Oost Friesland in 1995. Resultaten van het evaluatie onderzoek (Speed campaign in South-East Friesland in 1995; Results of evaluation research (Report in Dutch with summary in English). R-96-16. Institute for Road safety Research SWOV.

8. ONISR [Observatoire National Interministériel de Sécurité Routière] (2006). La sécurité routière en France. Paris, La Documentation Française.

9. NHTSA - Federal Highway Administration. (1996). Detection Technology for IVHS – Task L Final Report, Contract DTFH61-91-C-00076, U.S. Department of Transportation, Wash., D.C.

10. NRA [National Roads Authority] (2005). 2005 Survey of Free Speeds (Urban and Rural). Dublin, National Roads Authority.

11. PIARC [World Road Association]. (2003). PIARC Road Safety Manual. Kent, United Kingdom

12. Stevens, J. (1996). Applied multivariate statistics for the social sciences (Third ed.). Lawrence Erlbaum Associates Publishers, Mahwah New Jersey.

13. Steyvers, F.J.J.M. (1995). Gedetailleerde data van de snelheidsmetingen in het project "Snelheidsbeperkende maatregelen op 80 km-wegen in Drenthe". (Detailed data of speed measurements in the project "Speed reducing measures on 80 km/hr roads in Drenthe"). Report WR-95-03. Traffic Research centre, Haren, the Netherlands.

14. TRB [Transportation Research Board] (1998). Managing speed; review of current practice for setting and enforcing speed limits. Special report 254. National Academy Press, Washington DC.

15. TRB [Transportation Research Board] (2003). Design speed, operating speed and posted speed practices. National Cooperative Highway Research Program (NCHRP): Report 504. Washington DC.

16. UTMC [Urban Traffic Management and Control project] (2000). Review of Current and Future Data Requirements and Detector Technologies and the Implications for UTMC.

17. Walters, L. (2001). Estimating the speed and acceleration in actual road traffic by spot measurements. Thesis FiF-a 48, Linköping University, Sweden.

5 Detailed manual for SPIs on Protective Systems

EKSLER[A], V., ALLENBACH[B], R., HOLLÓ[C], P., SCHOON[D], C.

[A]CDV, [B]BfU, [C]KTI, [D]SWOV

5.1 Introduction

Many European countries have been assessing the use of protective systems at the national level for several decades. Some others have recently made the first steps towards such assessment, whilst there are still others seeking to do so in the near future.

Assessing protective systems use in traffic has several applications:

1. It allows identification of the problem at the national (regional) level and depicts its magnitude, so the estimation of social costs and avoidable fatalities (injuries) can be calculated.
2. It provides evidence for arguments about why their use is essential and why it should be supported by different policy measures.
3. It provides a base-line indicator that can be used for monitoring and evaluating progress towards programmes' targets in time.
4. It allows international comparisons of a country's performance and allows the identification of ineffective national policies.

To assess the situation in a consistent and reliable manner, the quality of indicators and of underlying data is essential. The basic requirements on SPIs for protective systems were summarized in the Road SPI Theory report (Hakkert et al, 2007). These are further developed in this report which is meant as a guideline on SPI data collection and production.

5.2 Methods and tools for assessing protective systems' use – existing practices

5.2.1 Examples of existing methods

Various methods exist for assessing protective system use in traffic. All of them may produce estimates necessary for later application in road safety management, but some of them are not necessarily appropriate for international comparisons. It has been shown in the Road SPI Theory report that there is a need to rely exclusively on independent observational surveys of protective systems' use in traffic. This is despite the fact that a survey done through questionnaires may produce very similar results as has been experienced in Sweden, where the Swedish Road Administration (VV) together with the Statistics Sweden (SCB) produce each year estimates of seat belt wearing rates. The latter are disaggregated per sex, age group, road type and seat, and are fully comparable with the rates observed during independent roadside surveys.

Observation of protective systems' use can be done either manually by trained observers, or by automatic devices. While the former method is applicable either on road profiles, or road segments, the use of the latter one is presently limited to road profiles.

Visual observations performed by trained observers are the most common method used to assess the rates of protective system use. Observers can be either placed along the roadside (or other convenient place), or they can be in a moving car. The latter approach makes it possible to cover the use on entire roads, but creates difficulties when it comes to appropriate statistical evaluation. Moreover, it is more costly and hence less preferred by road administrations or other bodies performing the survey.

The use of automatic video devices for assessing protective systems' use is a newly developed technique used for police enforcement. For example, the Dutch National Police Agency (KLPD) possesses cameras detecting seat belts that are used. The system has developed ways to avoid reflection of the car screen, namely the use of polarisation, and applies front edge technology showing the place where the seat belt – if worn – should be. Moreover, photogrammetry is used; a technique that changes the perspective in a way which also makes it easier to detect the seat belt. A major obstacle for the use of these techniques nowadays is their very high purchase and operational costs (annually 500,000 euro). The use of the technique is also limited only to the front seats of passenger cars.

For the sake of precise evaluation in respect to different road types, the observations done from moving vehicles cannot be recommended.

Estimating the protective system use rates, which are representative for the country, is not an easy task, as the rates vary in time and space and depend on various factors. To produce accurate national estimates, the choice of road profiles, their numbers, and spatial distribution plays a crucial role. Moreover, the proper size of observed samples at particular locations contributes much to obtaining statistically correct estimates. That is the reason why a sampling procedure must be employed to determine the optimal set of road types across a country.

According to the IRTAD study on the availability of seat belt wearing data in the OECD member countries (IRTAD, 1995): "The supplementary survey indicates substantial variation in sample size, national representation, and periodicity. These are real limitations to international comparability; nevertheless, wearing data may provide a useful background for comparing road accident fatalities between countries and also a platform for the improved collection of data on the wearing of seat belts." For example, in Ireland, the National Road Authority (NRA) assesses passenger car seat belt wearing rates on more than 100 road profiles, on each of them assessing 100 vehicles, giving a total sample of minimum 10000 vehicles, while in Hungary, the same number of vehicles is observed on a considerable lower number of road profiles.

5.2.2 Possible parameters versus selected SPIs

Figure 5.1 summarizes all data sources for the assessment of the use of protective systems, which are: police data, self-reported data, an observational survey and accidents. It further illustrates that SPI values have two dimensions, time and spatial. Firstly, there are considerable differences in rates recorded in daytime and night-time. Generally, the rates are lower during the night, as the road user spectrum is different. Secondly, the rates vary across road network and country. This can be related to the different nature of journeys and socio-economic characteristics of road users (NHTSA, 1996).

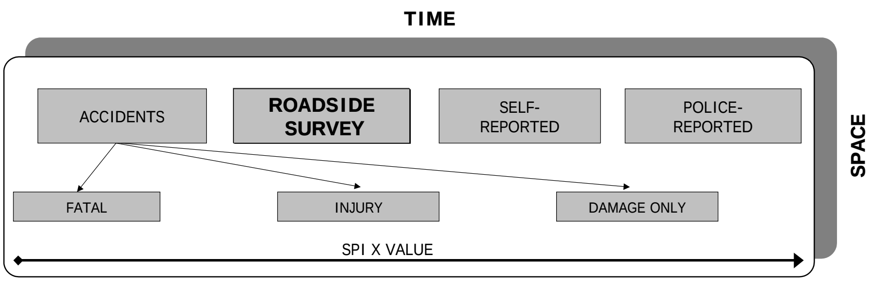


Figure 5.1: SPI values according to their origin

The daytime wearing rate of protective systems in traffic has been identified as a direct indicator, as it can be relatively easily assessed whereas the reliability and other requirements can be more conveniently met. Besides, the use of protective systems recorded by the police for fatal accident participants, is a recommended indirect indicator allowing inference about the effectiveness of the systems at national (European, regional, local) level to be made.

5.3 Detailed description of available best practices

5.3.1 Good practice country example

Several member states assess protective systems use in ways resulting in reliable and valid national estimates and allowing conclusions to be made about subgroups of observed road users. In this Section, the **Swiss** methodologies on estimating seat belt use in passenger cars and helmet use by cyclists are presented, since they are far more representative than for many other countries and have a long tradition of use.

The methodologies currently used in Switzerland are the results of permanent evolution and improvements realised over several decades. The Swiss Council for Accident Prevention (bfu) has conducted surveys on safety performance indicators for protective systems since 1962. At the very beginning, only data on seat belt wearing in passenger cars were gathered and estimations produced. Later, the surveys on helmet use, daytime running lights were also launched. The bfu's observation and proceeding methods can be understood as a compromise between the required methodology and the financial and personal efforts in fieldwork.

Survey on seat belt use (SPI-A)

The first step in Swiss methodology for assessing seat belt use by light vehicle occupants is to consider geographical regions: Switzerland consists of three linguistic regions and it is common opinion that populations in these regions differ in respect to their behaviour in road traffic. Thus, surveys on behaviour like seat belt wearing have to consider these regional differences.

The second step is to consider different road types with known differences in seat belt wearing rates. Both characteristics, regions and road type, constitute 9 stratification cells. Altogether 57 observation sites are then attributed to each particular stratum (Table 5.1).

Nr. of observation sites		Linguistic region			
		German	French	Italian	Total
Road type	Urban	16	6	3	25
	Rural	10	6	2	18
	Motorway	8	4	2	14
	Total	34	16	7	57

Table 5.1. Stratification characteristics of Swiss survey on seatbelt wearing

The Swiss regions vary according to territory and population size. The proportions of population size are about 0.75 for the German speaking region, 0.20 for French and 0.05 for the Italian-speaking region. For a proportional stratification with a total of 57 sites, the distribution for regions would result in 43/11/3 observation sites. (The proportion of population in the three regions corresponds to the distribution of traffic exposure in terms of person-km travelled on front seats of light vehicles.)

The same kind of considerations is made for different road types. The distribution of traffic on different road types is 0.336 for motorways, 0.345 for rural and 0.319 for urban roads (Table

5.2). With a proportional stratification for both characteristics we would only have one observation site per road type in the Italian speaking part of Switzerland. Disproportional stratification used here ensures a minimum of two sites for each stratum.

Stratification has two different advantages: firstly, it can reduce sampling errors, if stratification characteristics vary in the targeted estimate and, secondly, it ensures that in sampling all necessary characteristics are covered. In Switzerland, both advantages are applied: It is known, that there are regional differences and for prevention measures it is important to have estimates for each region. It would also be interesting to have estimates for every single Swiss Canton. But to realize such a detailed observation system, the stratification table shown above would be a table with 3x26 = 78 cells resulting in a minimum of 156 observation sites which goes beyond the scope of available resources.

Weights of aggregation		Linguistic region			
		German	French	Italian	Total
Road type	Urban	0.239	0.064	0.016	0.319
	Rural	0.259	0.069	0.017	0.345
	Motorway	0.252	0.067	0.017	0.336
	Total	0.750	0.200	0.050	1.000

Table 5.2. Weights of disproportional stratification characteristics, standing for the proportion of travel in terms of person vehicle km on different road types and within different regions

The guidelines for the observers state that at each observation site a total of 500 vehicles have to be observed and the observation time should not exceed one hour. So in most cases there is a total of 500 observations, but it also happens that less than 500 vehicles are observed at the location. In 2006, there was one site with only 250 drivers observed, which is still fitting to statistical requirements. As a consequence of the disproportional stratification design, estimates have to be weighted back to original proportions of the stratification characteristics.

On-site observations

The survey on seat belt wearing is conducted annually between April and June. It starts with an instruction and training day, where all observers are briefed on methodological issues and more importantly, on the issues relevant to the performance of road observations. Typically, there is one observer for one location. To keep down travelling costs, the country is divided into four regions to which the observers are assigned (Figure 5.2).

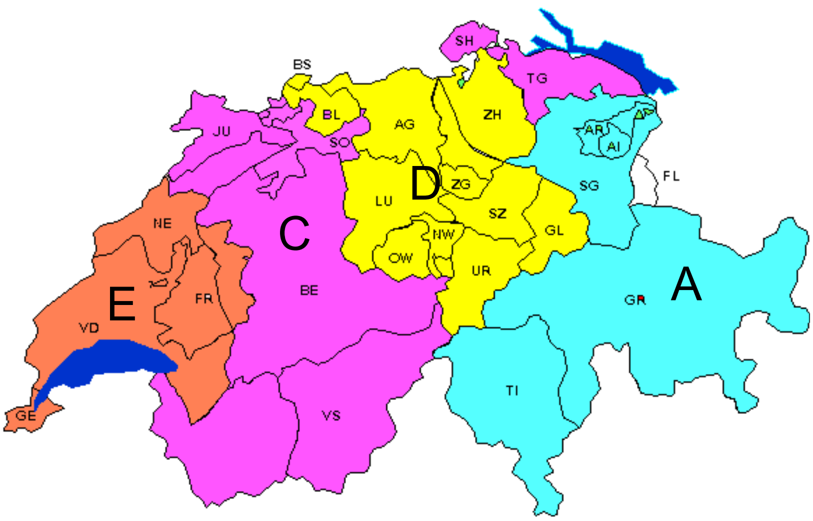


Figure 5.2. Assignment of observation sites to different artificial regions

The assignment of the observation sites to the four artificial regions is independent of the stratification considered (Table 5.3). Actually, artificial regions do not correspond to the linguistic regions, for which conclusions are to be drawn.

Weights of aggregation		Region for assignment of observation sites				
		A	C	D	E	total
road type	urban	3	4	3	4	25
	rural	5	4	5	4	18
	motorway	8	8	5	4	14
	total	16	16	13	12	57

Table 5.3. Assignment of observation sites to different artificial regions

Each observer is equipped with paper observation forms on a clipboard, mechanical hand counter, safety vest and accreditation to work as an observer signed by the head of the bfu research department. The observations on seat belt use by driver and front seat passenger are combined. So the observer first counts drivers with the hand counter and records results on the paper form. On workdays there are only a few cars with front seat passengers. To count their seat belt use the observer uses either a second hand counter or uses a paper observation form. The project leader continuously supervises fieldwork, so that every observer is supervised at least once. In addition to seat belt wearing counting, the characteristics of each observation site are registered and pictures are taken by observers.

After the fieldwork is completed the paper forms are coded twice into an Excel-sheet and the weighted estimates are calculated with the equations shown in this manual. The calculation of standard errors and confidence intervals is more complex. The sampling design is not based on a simple random sample and as a consequence, the calculation of standard errors has to take into account the design effect and the weighting of disproportional stratification. The design effect is a factor, which has to be multiplied with the standard error under assumption of a simple random design. In the design of the seat belt survey, a clustering effect has to be assumed. This means that observations within an observation site are more similar than observation between sites. Generally, clustering effects result in a design factor larger than one and increase the standard errors and confidence intervals.

There are several methods to calculate standard errors for complex sample design. They are implemented in some statistical software packages or specific software. Because these methods use linearization or replication of design effects, they cannot be calculated using a simple Excel sheet. Table 5.4 shows that the design effect varies between 0.01 and 45.37.

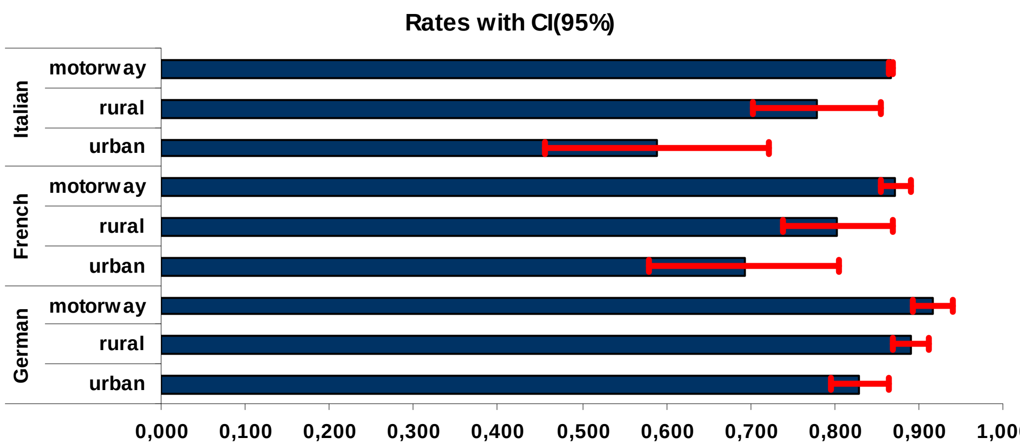
Region	Road type	Estimate	CI 95%		Std. Error	Design Effect
German	urban	0.830	0.796	0.864	0.017	16.18
	rural	0.891	0.869	0.913	0.011	5.78
	motorway	0.916	0.892	0.940	0.012	7.53
French	urban	0.693	0.579	0.806	0.056	45.37
	rural	0.804	0.737	0.870	0.033	20.74
	motorway	0.872	0.853	0.890	0.009	1.48
Italian	urban	0.590	0.456	0.723	0.066	27.39
	rural	0.780	0.703	0.856	0.038	8.53
	motorway	0.867	0.865	0.869	0.001	0.01

Table 5.4. Seat belt wearing rates and 95% confidence intervals (CI95%)

The large effects show that the sample design is not very efficient in some of the stratification cells. Reducing the number of observed cars and increasing the number of observation sites would decrease the design effect. The very small effect on Italian motorways (only 2 sites) is the result of exactly the same calculated estimates in both locations and so is a random effect.

The estimates of seat belt wearing rates can be produced for each strata considered. Figure 5.3 shows the respective rates with their 95% confidence intervals. While the CI is relatively large for some of the strata, it is relatively small for the overall value of seat belt wearing rate (see Figure 5.3).

The results were produced with the help of STATA statistical software. This programme allows accounting for stratification weight and primary sampling units simultaneously.



	German			French			Italian		
	urban	rural	motorway	urban	rural	motorway	urban	rural	motorway
Rate	0,830	0,891	0,916	0,693	0,804	0,872	0,590	0,780	0,867

Figure 5.3. Seat belt wearing rates per road types within regions and CI95%

Survey on helmet use by pedal cyclists (SPI-F)

The survey on the use of helmets by pedal cyclists has been performed in Switzerland since 1998. The three linguistic regions are considered as in the case of the seat belt survey. Altogether 53 locations are attributed to different roads across the country, while taking into consideration four different journey purposes of cyclists; commuting, shopping, school and free time/sport riding (Table 5.5). Hence, the stratification concerns regions and ride purpose.

At each carefully chosen location, at least 120 cyclists are observed during a maximum of 3 hours. The data are recorded manually onto the observation forms by the observers. Not only helmet wearing is recorded, but also other attributes, such as bicycle type, sex, age group (<14, 15-29, 30-44, 45-59, 59+), use of protective accessories, such as gloves, etc. The observation form used is presented in *Annex B.2 to Chapter 5*, giving clear evidence on the variable types recorded by observers.

Nr. of observation sites		Linguistic region			
		German	French	Italian	Total
Ride purpose	Commuting	7	1	1	9
	School	7	1	1	9
	Shopping	7	1	1	9
	Leisure	14	7	5	26
	Total	35	10	8	53

Table 5.5. Stratification characteristics of the Swiss survey on helmet use by pedal cyclists

The weights of disproportional stratification characteristics standing for the proportions of travel in terms of person vehicle km within different regions and travel purposes are further presented (Table 5.6).

Weights of aggregation		Linguistic region			
		German	French	Italian	Total
Ride purpose	Commuting	0.179	0.022	0.002	0.203
	School	0.139	0.017	0.002	0.158
	Shopping	0.083	0.010	0.001	0.094
	Leisure	0.480	0.060	0.005	0.546
	Total	0.880	0.110	0.010	1.000

Table 5.6. Weights of disproportional stratification characteristics

Since six strata have just one observation site, and because of a maximum of 120 observed individuals at these sites, calculation of reliable estimates for combining riding purpose and region is not reasonable. Thus, the results are only presented for both characteristics separately (Table 5.7,a,b). The calculation of estimates and confidence intervals follows the same considerations as in the case of the seat belt use survey.

One important factor influencing fieldwork in this survey is the weather. On days with heavy rain or other bad weather conditions nearly nobody uses bicycles and the observers leave the observation site without reaching the targeted number of observations. In the survey realized in 2006 bad weather conditions were met on two locations in the Italian speaking region, where on one site only 26 and on the second 37 bicyclists were observed during the three hour assessment period. As a result the reliability of estimates decreased.

Again, confidence intervals were calculated considering stratification weights and the primary sampling units. As a consequence, the confidence intervals are wider and more conservative than in a calculation assuming a random sampling.

Purpose of ride	Estimate	CI 95%		Std. Error	Design Effect
Commuting	0.232	0.113	0.351	0.059	16.208
School	0.288	0.145	0.431	0.071	22.257
Shopping	0.163	0.128	0.198	0.017	1.469
Leisure	0.515	0.456	0.573	0.029	10.352

Region	Estimate	CI 95%		Std. Error	Design Effect
German	0.395	0.324	0.466	0.035	18.712
French	0.331	0.216	0.447	0.057	15.729
Italian	0.476	0.297	0.654	0.089	24.815
Total	0.389	0.324	0.453	0.032	23.422

Table 5.7a,b Rates of helmet use by pedal cyclists (SPI-F) and CI95%

The same results can be further presented graphically making the estimated confidence intervals more readable (Figure 5.4).

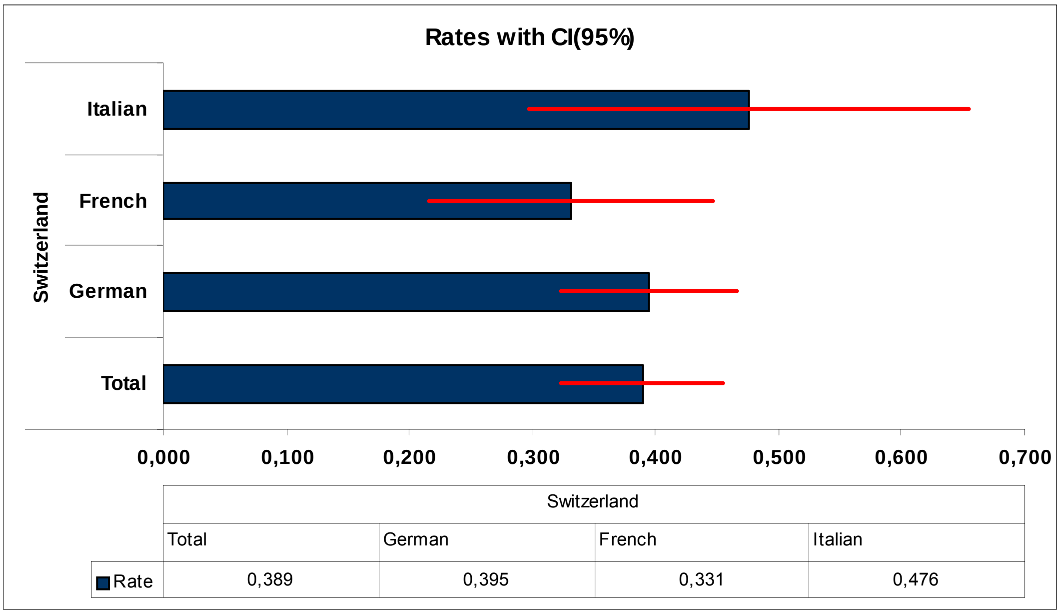


Figure 5.4 Rates of helmet use by pedal cyclists (SPI-F) together with CI95%

5.4 Recommended system for producing protective systems' SPIs

5.4.1 Setting up a national system

A set of indicators for assessing protective system use rates that has been proposed by SafetyNet (Hakkert et al, 2007), covers different road users in traffic. The very first step must, therefore, be a choice of indicators to be assessed at the national level. This should be done against the background of national conditions, such as fatalities' distribution, road safety programmes and targets, and available budget. Given the sorting of indicators corresponding to their effectiveness in preventing fatal injuries in traffic at the European level, the choice should include at least the first 3 indicators, but preferably all of them.

Significant regional disparities in protective systems' use have been observed in numerous EU countries, where they have been reported separately for relevant regions. This is particularly a case with seat belt use. In Germany, figures are reported for former Western and Eastern parts of the country, while in Switzerland, they have been reported separately for the three linguistic regions. Rates are determined for three different regions of Italy as well (Taggi, 2005). There are, however, other countries, which do not necessarily exhibit significant regional disparities, but which need regional values to be figured out as well. (For example, recently established regional road safety observatories in France (since 2005) are supposed to use regional values of some SPIs for enhancing the effectiveness of regional road safety policies. The same applies for the Czech Republic, where regional values of SPIs have been produced recently.)

Nationwide surveys can be very expensive and unnecessary if it is known that the values of the SPI are distributed more or less equally (which is a priori the case in small countries). This is the case in many Member states, but not in all, as, for example, Italy and Switzerland show great heterogeneity in e.g. seat belt wearing rates.

When starting from zero, it is important to assure regular (annual) assessment, since monitoring of the development in time is one of the most important applications for the country. If a sound system of observational surveys is established at the country level and has been used for several years, then eventual alteration of the system must be made with care to allow for a time-series analysis in the future.

The national system should further fulfil the requirements as described on the following pages.

Requirements for national systems producing SPIs in the area of protective systems

A Design requirements

Surveys should incorporate the following minimum design requirements:

A.A Probability-based requirement

The sample identified for the survey should have a probability-based design such that estimates of safety belt (helmets) use will be representative for the population of interest in the country and confidence intervals may be calculated for each estimate produced.

A.B Other sampling requirements (population, demographic, and time/day requirements)

Surveys conducted in accordance with this part should comply with the following minimum population, demographic, and time/day requirements:

A.B.A Population of interest

a) Drivers, front seat passengers, children in passenger motor vehicles (passenger cars, vans, and sport utility vehicles) and two-wheelers must be observed in the survey. (Only overall protective systems' use for the population of interest is required. However, in order to assist in the evaluation of trends and group differences, it is recommended that data be collected in such a way that estimates can be produced for population sub-groups, e.g. seat belt use estimates reported separately for drivers and front seat passengers and different road types which are present in the country.) The following road user categories are subject to observations:

A: Passenger car and vans occupants over 12 years old in front seats

B: Passenger car and vans occupants over 12 years old in rear seats

C: Children in passenger cars under 12 years old in front and rear seats

D: Occupants of coaches and heavy-duty vehicles over 12 years old in front seats

E: Occupants of coaches over 12 years old in rear seats

F: Pedal cyclists

G: Moped riders

H: Motorcyclists

b) The SPI values should be assessed on all major road types which are present in the country and relevant for each road user category (4.2.1.a).

Road types – IRTAD	Road types - SafetyNet	A	B	C	D	E	F	G	H
Motorways	AAA	X	X	X	X	X			X
Roads outside urban areas	AA to C	X	X	X	X	X	X	X	X
Roads inside urban areas	DD to E	X	X	X			X	X	X

Table A.1: Road types for assessment of SPIs

c) The target population for each of SPI category encompasses all national (regional) citizens as well as all foreigners in traffic.

d) All persons travelling in the vehicles for which the law allows exception for protective systems use should be *a priori* excluded from the observation sample (taxi, police, emergency service, armed forces' vehicles, etc).

A.C Demographics

NUTS-2 regions (official Eurostat regional classification of regions corresponding to nationally defined area units; typically of 0,8-3 millions of inhabitants) or other regional units are eligible for inclusion in the sample. If once proven (during a nation-wide survey) that no major regional differences exist [$(SPI_{i,max}-SPI_{i,min})<5\%$], then regions can be chosen randomly unless covering all regions.

A.D Time of day and day of week.

All daylight hours for all days of the week are eligible for inclusion in the sample. By "daylight", the period starting one hour after the sunrise and ending one hour before sunset is understood. Observation sites must be randomly assigned to the selected day-of week/ time-of-day time slots. Only in case of SPI_C (use of restraints by children) the observations should be carried out on weekend days. *Note: weekend days include also public (national) holidays.*

Eligible time frames	A	B	C	D	E	F	G	H
Apr.-June and Sept.-Oct.	X	X	X	X	X	X	X	X
Daytime (typically 8-12 and 13-18)	X	X	X	X	X	X	X	X
Week days	X	X		X	X	X	X	X
Weekend days			X					

Table A.2: Time of day and day of week requirements

A.E Observational requirement

Minimum requirements include the following:

(a) The sample data shall be collected through direct observations of protective systems' use by road users on roadways within the country, conducted completely within the calendar year for which the seat belt, Child Restraint Systems (CRS), or helmet use rate are being reported.

(b) Protective systems' use shall be determined by observation of the use or non-use of a seat belt, CRS and safety helmets on a predefined set of locations chosen with the help of appropriate sampling methods. In the case of CRS, the correct use of devices should be assessed separately. Under the "correct use", the use meeting requirements specified by the manufacturer is meant.

(c) Observation (registration) is performed on road profiles, intersections, petrol stations or other eligible locations such as in the vicinity of shopping centres. (Automatic) video devices can be used as well. The observation of seat belt use in coaches should be performed by an observer who is present on-board during a journey.

(d) The observations should be performed by independent observers (not uniformed police or other officers) working under the responsibility of a survey coordinator, assuring that wearing rates would be not influenced.

(e) Observers shall be required to follow a predetermined, clear policy in the event that observations cannot be made at the assigned site at the specified time (due to heavy rain, construction, safety problems, etc.).

(f) Instructions to observers shall specify which road section and which direction of traffic on that road are to be observed (observers must not be free to choose between roads at an intersection).

(g) Observers shall follow clear instructions on how to start and end an observation period and how to stop and start observations if traffic flow is too heavy to observe all concerned individuals or if they begin moving too quickly for observation (to remove any possible bias, such as starting with the next person using protective system).

The minimum and recommended number of observation sites is specified in Table A.3. For example, the survey on the use of seat belts in the front seats of light vehicles should be made on at least 10 road profiles across the country (e.g. 4 sites on urban roads, 3 sites on rural roads and 3 sites on highways), but it can be recommended to use more than 30 sites (see *Annex A to Chapter 5 and Statistical Appendix* to the report for more details on statistical sampling).

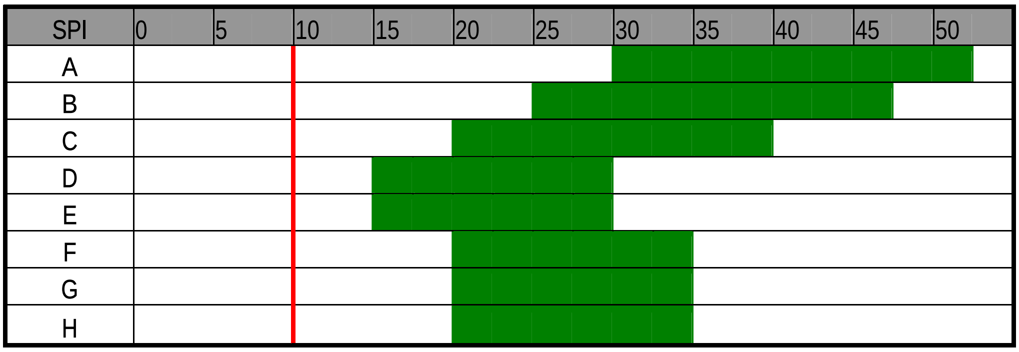


Table A.3: Minimum and recommended number of observation sites

Similarly, the minimum and recommended sample size (number of observed individuals concerned) for each observation site is specified in Table A.4

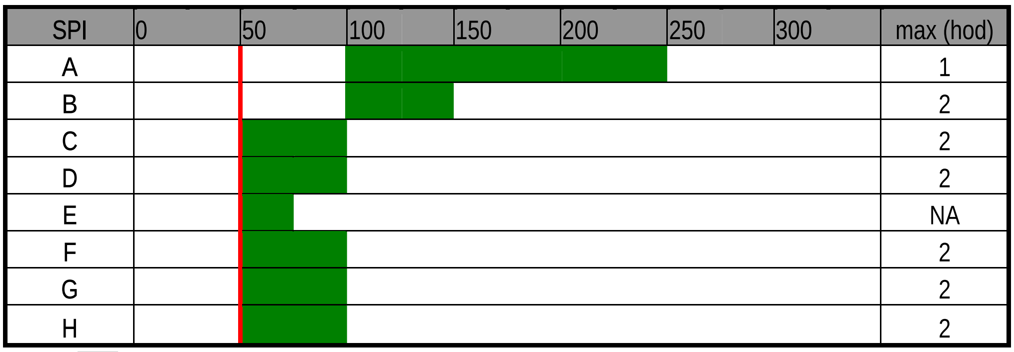


Table A.4: Minimal and recommended sample size per observation site

A.F Precision requirements

Sampling design should allow the production of SPI estimates with the following accuracy:

SPI	Accuracy (precision) in %
A	2
B	3
C	5
D	5
E	5
F	5
G	5
H	5

Table A.5 : Recommended accuracy of SPI values

Accuracy (precision) means here the confidence interval width. Consider an example, when assuming a simple random sampling, with six observation sites and a minimum of 100 observations per site, a total N of 600 is observed. Then, if the observed protective systems' use rate is 0.5, then the 95% confidence interval will be from 0.46 to 0.54, meaning a 4% precision of the estimate. In the same case, if the protective systems' using rate is 0.8, then the 95% confidence interval will be from 0.77 to 0.83, meaning a 3.2% precision of the estimate.

Sampling design corresponding to the requirements specified in Section A.E above meets the precision requirements specified in Table A.5. For more details, please refer to Annex A to Chapter 5.

B Documentation requirements

All sample design, data collection, and estimation procedures used in country surveys, which are conducted in accordance with this Manual, must be well documented. As a minimum, the documentation must:

B.A For sample design:

(a) Define all sampling units, with their measures of size;

(b) Define what stratification was used at each stage of sampling and what methods were used for the allocation of sample units to the strata;

(c) Explain how the sample size at each stage was determined;

(d) List all sample units and their probabilities of selection; and

(e) Describe how observation sites were assigned to observation time periods.

B.B For data collection:

(a) Define the observation period;

(b) Define the observation site and what procedures were implemented when the observation site was not accessible on the date assigned;

(c) Describe what vehicles were observed and what procedures were implemented when traffic was too heavy to observe all vehicles; and

(d) Describe the data recording procedures.

B.C For estimation:

(a) Display raw data and the weighted estimates;

(b) Ideally, for each estimate, provide a standard error and a 95 percent confidence interval; and

(c) Ideally, describe how the estimates and their variances were calculated.

C Requirements for the presentation of results

The following figures will be presented as results of the survey:

(a) The values of SPI_A to SPI_H
(b) The conversion rates for the time series
(c) Ideally, the confidence intervals of estimates

C.A The formulae for data evaluation and aggregation

The value of each indicator *SPI*_X (where *X* = A, B, …H) is calculated as the proportion of the relevant road users using protective systems among all road traffic participants in the roadside observation sample:

$$SPI\_X = \frac{N_{use}}{N_{total}} \cdot 100$$ Equation A.1

where:

***N**$_{use}$* is the number of road users using a protective system in road traffic, as observed during the roadside survey in order to find SPI_X value,

***N**$_{total}$* is the sample size, meaning the total number of road users observed during the roadside survey in order to find SPI_X value.

In the case of SPI_C, the indicator ideally distinguishes between proper and improper use of child restraint systems. Minimal requirements, however, apply only for the use or non-use of these systems among children in traffic.

If, for any reason, it is not possible to design the survey in such a way that it produces directly the value of the indicator, then the following aggregation and transformation rules shall be applied:

C.A.B Aggregation for different road types

Since the SPI values vary for different road types, there is a need to weigh the SPI values in respect to the traffic performance related to each road type existing in a country. For each SPI indicated (SPI_A through SPI_G), the value of the indicator is calculated by taking the weighted sum of wearing rates of the protective system under consideration, for all road types studied, in a representative sample:

$$SPI\_X = \sum_{i=1}^{N} WR_i \times TS_i \quad \text{while: } \sum_{i=1}^{N} TS_i = 1$$ Equation A.2

where:

***WR**$_i$* is the 'wearing rate', the number of persons using the protective system divided by the total amount of users of the particular vehicle type, observed in a representative sample during the independent roadside survey;

***TS**$_i$* is the 'traffic share' for road type *i*, i.e., the amount of kilometres travelled on road type *i* by the road users under consideration, divided by the total amount of kilometres travelled by the road users under consideration for all road types assessed;

i stands for road type and

N is the total number of road types assessed.

If the data on the exposure of particular road users using relevant vehicles are not available, then the exposure in terms of million vehicle-kilometres (MVKMS) of relevant vehicles can be considered.

If the exposure data in terms of MVKMS are not available at all, an expert estimation may be applied, based on data from other countries.

C.A.C Aggregation for different regions

If the value of an indicator is available for different regions only, than the Equation A.3 is to be used:

$$SPI\_X = \sum_{i=1}^{N} WR_i \times TS_i \qquad \text{while: } \sum_{i=1}^{N} TS_i = 1$$

Equation A.3

where:

WR$_i$ is the 'wearing rate', the number of persons using the protective system divided by the total amount of users of the particular vehicle type, observed in a representative sample during the independent roadside survey in region *i*;

TS$_i$ is the 'traffic share' for road type *i*, i.e., the amount of kilometres travelled in region *i* by road users under consideration, divided by the total amount of kilometres travelled by road users under consideration in the whole country;

i stands for region and

N is the total number of regions under consideration.

If the exposure data in terms of MVKMS are not available for regions, population counts can be used instead to determine the weights.

C.A.D Aggregation for driver and front seat passenger

Formula A.4 provides the way to aggregate wearing rates for front seats in passenger cars (and vans) in the country (SPI_A) for those countries, which register separately wearing rates for driver and front seat passenger:

$$SPI\_A = \frac{N_{Wd} + N_{Wf}}{N_d + N_f}$$

Equation A.4

where:

N$_{Wd}$ is the number of drivers counted as wearing their seat belt;

N$_{Wf}$ is the number of front seat passengers counted as wearing their seat belt;

N$_d$ is the number of drivers counted during the survey;

N$_f$ is the number of front seat passenger counted during the survey.

C.A.E Aggregation for different age/weight groups for SPI_C

If the data are available for different age groups, the SPI_C value is obtained through weighting all use rates per the number of children in the relevant age group, as observed during the roadside survey. The weighting per population size in the age group is not allowed.

$$SPI\_C = \frac{\sum_{i}^{n} UR_i \cdot N_i}{N_{total}}$$

Equation A.5

where:

UR$_i$ is the 'usage rate' of child restraint system by a child of certain age group *i* in passenger car as observed in a representative sample during the independent roadside survey;

N$_i$ is the number of children assessed for the use of a child restraint system in the age group i as observed in a representative sample during the independent roadside survey;

N$_{total}$ is the total number of children assessed for the use of a child restraint system in all age groups.

Example:

Hereby, the way of calculating the value of the SPI-A indicator for the UK is presented for illustrative purposes. Two separate regular survey systems exist, one for Great Britain, and one for Northern Ireland. Besides, Scotland occasionally performs its own survey on seat belt use in light vehicles showing similar results as for Great Britain. So for aggregating nationally representative value, two geographical regions are considered: Great Britain and Northern Ireland.

As the rates of seat belt wearing are assessed separately for driver and front passenger seat occupants, the two values must be aggregated. The weights corresponding to the observed occupancy rate are applied. In the next step, the resulting rates are aggregated for passenger cars and vans, using their exposure in traffic (here in terms of MVKMS) as weights. Lastly, the two values of the SPI-A indicator are aggregated for the U.K. again using exposure in traffic weights. While the calculations are done with the rates considered as real numbers (one decimal point showed here), the final value of the indicator is rounded to a whole number. Here, the SPI-A value is then 90%.

Region	Vehicle	Seat	Rate	w	Rate	w	Rate	w	SPI_A
Great Britain	Passenger cars	Driver seat	92,6	0,65	93,1	0,90	90,2	0,957	90,3
		Front passenger seat	94,0	0,35					
	Vans	Driver seat	71,0	0,65	64,5	0,10			
		Front passenger seat	58,0	0,35					
Northern Ireland	Passenger cars	Driver seat	93,0	0,65	92,7	0,93	91,3	0,043	
		Front passenger seat	92,0	0,35					
	Vans	Driver seat	68,0	0,65	64,5	0,08			
		Front passenger seat	58,0	0,35					

Table C.1: Aggregation of SPI-A values for the UK in 2005

C.B Calculating Conversion rates

Development of indicators over time is evaluated by means of two indicators:

1. **Year to year increase (annual increase)** defined as a difference of SPI value (use-rate) from last year (*t-1*) to present year (t).

$$AI_t = -\left((SPI_{t-1}) - (SPI_t)\right) = (SPI_t) - (SPI_{t-1}) \qquad \text{Equation C.1}$$

2. **Conversion rate** defined as a rate of decrease of non-use from last year (*t-1*) to next one (*t*)

$$CR_t = \frac{(100 - SPI_{t-1}) - (100 - SPI_t)}{(100 - SPI_{t-1})}$$ Equation C.2

The conversion rate provides a better measure of improvement than percentage point or percentage increase in use as it shows the percentage of belt non-users converted into users, each year. The improvement is thereby assessed in a way that does not penalize regions or other categories that already exhibit high using rates. The conversion rate is positive when the use increases, but it can also be negative when the use declines.

Example:

An example for the Czech Republic is presented below, showing changes in terms of conversion rates between the years 2000-2006.

The negative rates visible in the top-left part of matrix depict a decline in seat belt use in the period 2000-2002. The most relevant conversion rates for two subsequent years are in bold in the diagonal.

Conversion rate	2000 (60)	2001 (58)	2002 (57)	2003 (62)	2004 (64)	2005 (72)	2006 (87)
2000 (60)	0	-5	-8	5	10	30	68
2001 (58)		0	-2	10	14	33	70
2002 (57)			0	12	16	35	70
2003 (62)				0	5	26	66
2004 (64)					0	22	63
2005 (72)						0	54
2006 (87)							0

Table C.2: Conversion rates for SPI-A values for the Czech Republic (2000-2006)

C.C Presenting SPI values

Both SPI values and conversion rates should be presented as results of the survey. It is recommended to present the values both graphically and numerically.

Example:

As an example, the evolution of the SPI_A indicator over time in Switzerland is shown. It reflects a temporal introduction of a seat belt law in the period 1976-1977 and its definitive establishment in July 1981. Since 1981 the rate is increasing slowly but constantly.

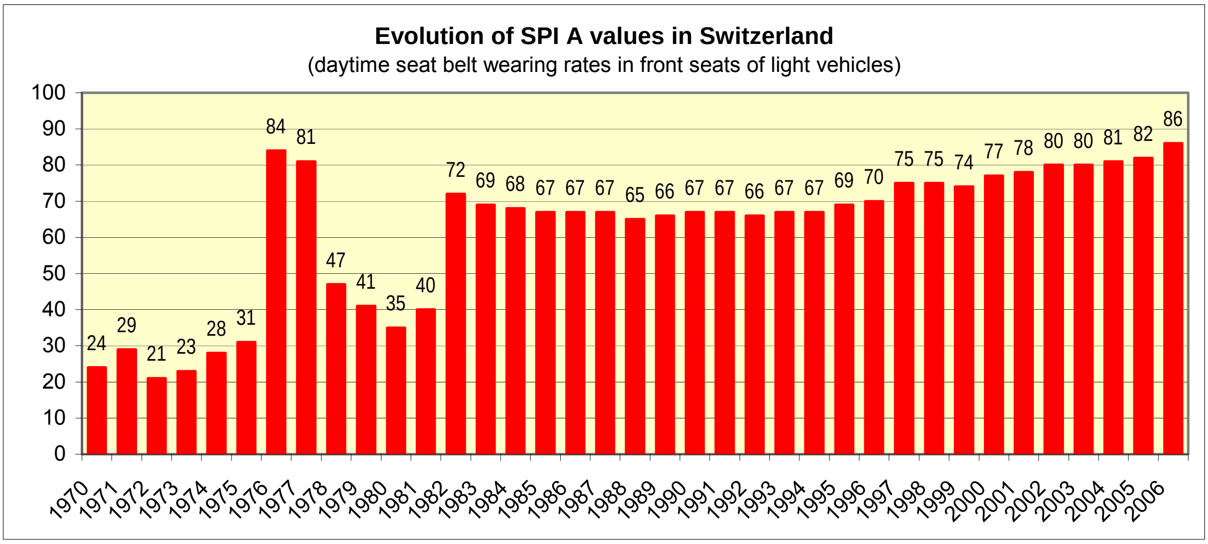


Figure C.3: The evolution of SPI_A indicator over time in Switzerland (1980-2006)

5.5 Discussion

Assessing protective systems' use is a prerequisite for recognising and evaluating weaknesses of road safety systems related to poor road user behaviour. Identifying the lack of protective systems' use among road users allows depiction of the scale of the problem. This in turn allows the setting of priorities and the launch of programmes to increase use.

International or regional comparisons of protective systems' use rates are vital tools for recognising deficiencies, setting priorities and stimulating efforts at the political level. Applying the common rules presented in this Manual allows the production of reliable and accurate indicators that are comparable among the countries (regions).

Periodical assessment of the protective systems' use is the only way to assess how to effectively employ the measures, which are meant to increase the protective systems' use by road users and thereby lower the road toll of the country. It is straightforward to think about the annual assessment as a standard, as all Member states basically evaluate their road safety level on an annual basis. More frequent evaluation can, however, be appropriate, if it aims to provide short-term evidence on the effectiveness of the measures applied (e.g. new legislation, increased penalties, penalty point system's introduction, etc).

Road user deaths in respect of use/non-use of protective systems can be used to describe the health and socio-economic impacts to society, and therefore can be useful for making persuasive arguments in support of related programmes.

5.6 Application for national road safety policies

The application of SPIs proposed for protective systems allows one to:

1) Monitor spatial and time variations and development;
2) Measure the effects of interventions;
3) Calculate social costs of prevented accidents as well as a reduction potential behind the increased rates.

The development over time is evaluated by means of two indices: the annual increase and conversion rates. Measuring protective systems' use by road users allows estimation of not only fatalities' and serious injuries' reduction due to their use, but also estimation of reduction potential by their increased use. This in turns allows one to perform cost benefit analyses of relevant road safety measures.

For this purpose, the value of an indirect indicator (a rate of protective systems' use by fatalities) is necessary. This value is much lower than the value of a direct indicator for several reasons:

(a) Protective systems' use is lower at night-time than in the day-time, when it is measured.
(b) People not using protective systems have a higher risk of sustaining injury/ fatal accidents.

It is not unusual that accident statistics do not allow identification of the value of an indirect indicator. In such cases, the rate of protective systems' use by fatally injured road users can be estimated from the known value of the indicator. Two methods can be applied here:

(a) One can assume that the risk of sustaining fatal injury is higher for those road users who do not use the systems. Then the following formula will be applied to estimate the rate of protective systems' use by fatalities:

$$F(x) = \frac{x(1-e)}{x(1-e) + R(1-x)}$$ Equation 1

where x stands for daytime usage rate of the device use, e stands for the effectiveness and R is a risk factor showing how many times the risk for those users not using the system is higher than the risk for those using the system.

For front seat occupants of passenger cars, this was empirically estimated as 1.66 (Salzberg et al., 2002). For other indicators, a risk factor of 1.5 to 2.0 can be assumed.

(b) An empirically established formula, describing the relationship between the values of direct and indirect indicators can be applied. Hereby, an example of a relationship valid for the seat belt use rates on front seats of passenger cars, established by the NHTSA in the USA is presented:

$UPF(x)=0.47249x^2 + 0.43751x,$ Equation 2

where x denotes the rate of seat belt use during the daytime and *UPF(x)* denotes the rate of seat belt use among potential fatalities when the daytime seat belt use is x (Wang and Blincoe, 2003).

The weaknesses of the two estimation formulas presented above are in somewhat inconsistent estimates and in the fact that they both were calibrated in late 1990s in the USA and not in Europe. No calibration has been made so far for the European light vehicle fleet.

Besides knowing a value of an indirect indicator, there is a need to determine the effectiveness of a particular device for the calculation. This has been discussed in the SPI Theory report (Hakkert et al, 2007), so a summary of suggested effectiveness of particular protective systems is presented:

Effectiveness[#] in % / SPI	A	B	C	D	E	F	G	H
Fatal injury	52	48	60	48	45	50	44 (8*)	44 (8*)
Serious injury	52						49 (8*)	49 (8*)
Slight injury	30	25					33 (8*)	33 (8*)

[#] ability to prevent injury of a road user involved in a road accident
* head injuries against other injuries

Table 5.8: Effectiveness of protective systems (Elvik & Vaa, 2004)

The effectiveness was estimated on the basis of in-depth accident and medical studies done throughout the world, whereas Table 5.8 provides only rough estimates of valid coefficients for further calculations. Since the effectiveness estimates are best established for fatal accidents, estimating the lives saved will be more accurate than the estimation of the number of injuries saved.

The lives saved by a protection system can be estimated as follows: if D_s people die using a safety device that has an effectiveness of e (i.e. that reduces fatalities by $e \cdot 100\%$), then one can infer that a total of $D_s /(1-e)$ people used the device in a setting in which they would otherwise die (the potential fatalities), of which $D_s(e/(1-e))$ were saved by the device. The latter is the number of lives saved as a result of the current level of usage of safety device. The number of cases in which seat belts failed to save lives provides a key to how many lives they actually saved, for a given e (Glassbrenner, 2003).

In addition to calculating the number of lives saved, one might also be interested in the additional number of lives that would be saved if occupants who currently do not use the protective devices were using them. This value is simply given by D_0e, where D_0 is the number of persons fatally injured and not using the system.

Country example:

The Czech accident database maintained by the Police Presidium contains information on the use of seat belts in accidents by all passenger car occupants. We assume that the data are reliable for occupant fatalities, as there is no obvious reason for suspecting that police officers do not report truthfully when making accident investigation. Then, one can simply calculate the number of lives saved by seat belts in passenger cars in 2004, by using the formulae presented above. The number of lives saved by seat belts in 2004 is estimated as 439, while an additional 149 could have been saved if all unbelted fatally injured passenger car occupants were using their seat belts.

Indicator 2004	Seat	Daytime rate	Effectiveness	Killed belted	Killed unbelted	Rate by fatality	Lives saved	Potential
		R_t	e (x100%)	D_s	D_0	R_{fa}	$N_s=(e/(1-e))*D_s$	$N_p=e*D_o$
A	Driver	71	0,52	269	175	61%	291	91
	Front	71	0,52	105	52	67%	114	27
B	Rear	39	0,48	37	64	37%	34	31
Altogether							439	149

Table 5.9: Estimation of lives saved and live saving potential for seat belts in passenger cars for Czech Republic in 2004

Seat belt wearing rates by occupant fatalities (indirect indicator) can be additionally estimated from the known daytime rate as observed during the survey (direct indicator) and after a repartition of killed front seat occupants into belted and unbelted. Applying Equation 1, and assuming the risk factor equals to 1.5, the estimated rate of using seat belts by fatally injured front seat occupants is:

$F(x)=(1-e)*x/(x(1-e)+R(1-x))=(1-0.52)*0.71/(0.71(1-0.52)+1.5(1-0.71))=0.539 \approx 54\%$

Applying Equation 2, the estimated rate of using seat belts by fatally injured front seat occupants is:

$UPF(x)=0.47249x^2 + 0.43751x=0.47249*0.71^2 + 0.43751*0.71 = 0.549 \approx 55\%$

Assuming the new values of indirect indicators are as follows: A=55%, B=30%, the number of lives saved due to the use of safety belts will slightly decrease to 385 (264+93+28=385), while the number of lives potentially saved if the use of safety belts by fatalities were 100%, will rise to 175 (104+37+34= 175).

The costs of one road fatal accident was 295,107€ in 2004, while the costs of one accident with serious injury was 97,355€. Assuming that using belts mitigates the consequences of a crash in such a way that the fatal injury is transformed into a serious injury and that there were 1.25 fatalities per one fatal accident, one can roughly estimate benefits from the use of belts in crashes as approximately 160,000€ per potentially saved live. For 175 lives potentially saved by their safety belts, this makes total savings of about 27,680,000€.

5.7 References

1. Broughton J. (2003) Seat belt wearing rates in cars in England 1998-2002, TRL 580, ISSN 0968-4107.

2. Cedersund, H.A. (2003). Bilbältesanvändingen i Sverige 2002, VTI 945, ISSN 0347-6049, VTI.
3. Eby, D. W., T. A. Fordyce and Vivoda, J.M. (2002) A Comparison of Safety Belt Use between Commercial and Non-commercial Light-Vehicle Occupants. Accident Analysis and Prevention, Vol. 34, pp. 285-291.
4. Elvik, R. And Vaa, T. (2004) The handbook of road safety measures. Elsevier.
5. Ewert, U. and Fitz, B. (2005). Sicherheitsgurt: Gründe für das Nichttragen und Massnahmen zur Erhöhung der Tragquote. BfU, ISBN: 3-908192-19-6.
6. Glassbrenner, D. (2003). Estimating the lives saved by Seat belts and airbags. NHTSA, Washington.
7. Hakkert, A.S., Gitelman, V., and Vis, M.A. (Eds) (2007) Road Safety Performance Indicators: Theory. Deliverable D3.6 of the EU FP6 project SafetyNet.
8. IRTAD (1995). The availability of seat belts wearing data in OECD countries, International Road Traffic and Accident Database (IRTAD) special report. OECD, Paris.
9. NHTSA (1996). Effectiveness of occupant protection systems and their use. Department of Transportation, Washington, D.C.
10. NHTSA (2000). Seat belt survey regulation, 23 CRF Part 1340.
11. NRA (2003). Survey on free speeds and seat belt wearing rates 2003, National Roads Authority (NRA), RS 467, ISBN 1900293 90 0, Dublin.
12. Ralph de Jong en Karin Bodewes (tns nipo), 2004. Gebruik van beveiligingsmiddelnen in auto's, report for AVV.
13. Resultat från 2005 års trafiksäkerhetsenkät, 2000:99, Vägverket, ISSN 1401-9612, Borlange, 2005.
14. Salzberg, P. et al. (2002) Predicting seat belt use in fatal motor vehicle crashes from observation surveys of belt use, Accident analysis and prevention, Vol.34, issue 2, pp. 139-148.
15. Sicherung durch Gurte, Helme und andere Schutzsysteme, Kompendium, BAST, Bergish Gladbach 2000.
16. Taggi, F. et al. (2005) Il sistema Ulisse per il monitoraggio dello'usso, delle cinture di sicurezza e del casco in Itali (2000-2005), ISS, OS/AMPPP/RT/217.
17. Tanner, J.C. and Scott, J.R. (1962). 50-Point Traffic Census – The First 5 Years. Road Research Laboratory.
18. Wang, J. and Blincoe, L. (2003) Belt Regression Model - 2003 Update, NHTSA Research Note, DOT HS 809.

6 Detailed manual for SPIs on DRL

HOLLÓ[A], P., GITELMAN[B], V.

[A]KTI, [B]TECHNION

6.1 Introduction

The Directorate-General for Energy and Transport of the European Commission issued on 1st August 2006 a consultation paper titled “Saving lives with Daytime Running Lights” (DRL, 2006). The European Commission is seeking the views of interested parties on its intention to propose measures to foster the rapid introduction of daytime running lights (DRL) for motor vehicles.

In the Executive Summary of the paper the following facts are presented:

- According to the research available, DRL[6] has a high potential to increase road safety: using DRL helps road users to better and earlier detect, recognise and identify vehicles. (Besides, the paper DRL (2006) does not mention it, but DRL helps to estimate the distance, speed and moving direction of vehicles, as well – see Koornstra et al, 1997.)

- Research studies estimate the life-saving potential of DRL to be in the order of 3 to 5% of the yearly number of road fatalities. If measures are undertaken to require the use of DRL throughout the European Union (EU), it could help *saving between 1,200 and 2,000 road fatalities per year* and, thus, make an important contribution to the European target of saving 25,000 lives per year on European roads.

The paper (DRL, 2006) emphasises the following research results:

- Road users not having lighting devices, i.e. pedestrians, cyclists, mopeds *do not become less conspicuous* if all vehicles feature DRL.
- A negative effect of DRL on the visibility of motorcyclists *cannot* be ascertained.
- Dedicated DRL and dipped headlines *do not* cause glare.
- It is true that DRL increases *fuel consumption and CO_2 – emissions*. However, the increase is by up to 1.5% if dipped headlamps are used, and by up to 0.3% *only* in the case of dedicated DRL.

(It should be mentioned at this point that in the case of, e.g., vehicle air-conditioning or the performance of hi-fi systems, it is not customary to investigate the environmental effects.) Nevertheless, even taking into account the effect on fuel consumption and CO_2 emissions, the safety benefits of a legal obligation to use dipped headlights on existing vehicles and to equip new vehicles with automatic dedicated DRL outweigh the costs by the factor 2 to 1, i.e. for one Euro invested into DRL, there is a benefit to society of 2 Euros.

14 Member States have mandatory rules on the use of DRL in force so far, with different requirements. Table 6.1 gives an overview of the existing situation (DRL, 2006).

---

[6] In the paper DRL (2006), the term DRL is used to describe the concept of using any lighting devices during daytime, in particular existing headlamps or dedicated lights.

Country	DRL use: where?	DRL use: when?
Denmark	All roads	All year
Estonia	All roads	All year
Finland	All roads	All year
Italy	Motorways and out-of urban roads	All year
Latvia	All roads	All year
Lithuania	All roads	November-March
Austria	All roads	All year
Poland	All roads	October-February
Portugal	Indicated roads	All year
Sweden	All roads	All year
Slovakia	All roads	October-March
Slovenia	All roads	All year
Czech Republic	All roads	All year
Hungary	Out-of urban roads	All year

Table 6.1. Fourteen European countries with mandatory use of DRL. (Source: DRL, 2006).

Furthermore, some Member States recommend the use of lights during daytime without mandating them while waiting for a harmonised European legislation. In order to avoid confusion and to improve road safety, a harmonised EU-wide rule could be necessary for the benefit of travelling citizens.

The consultation paper DRL (2006) offers five options for implementation. Option 3, which seems to be the optimal and most realistic, is the following:

"The use of DRL is required by all motor vehicles from a certain date". In addition, "new cars sold after the same date will be required to have dedicated DRL that are switched on automatically."

This option is referred to as the "behavioural plus automatic dedicated DRL option". Being implemented, it could provide the necessary flexibility and efficiency as well.

In those countries, where automatic DRL was introduced a long time ago (e.g. Sweden, Norway), the DRL usage rate is and remains close to 100%, so the DRL usage rate as a behavioural safety performance indicator cannot be interpreted any more. This means that the DRL use as a safety performance indicator will lose its importance, when the option 3 is introduced Europe-wide.

6.2 DRL usage measurements - existing practices

A number of EU countries perform observational surveys of the DRL usage rates. Recent estimates are available for Austria, the Czech Republic, Estonia, Finland, France, Hungary and Switzerland (see *Vis and van Gent, 2007*). The Netherlands has performed this kind of survey in 1990-1993.

Based on the information provided by national experts, a regular DRL survey is obligatory in Latvia and is recommended in Estonia, whereas in all other countries, including the countries having a DRL law, it is not even recommended.

All the countries, which have recent estimates of DRL usage rates (see above), calculate them according to road categories. Some countries (e.g. the Czech Republic and Switzerland) estimate the DRL usage rates according to vehicle types, as well. Each country has its specific DRL legislation for types of vehicles and types of areas.

Austria has recently introduced the mandatory use of DRL: from 15th November 2005, the mandatory use of DRL is in force on all roads, and during the whole year. Between 15th November 2005 and 15th April 2006 no fine was charged when someone forgot to turn on the lights. After that period a fine of about 15 € was proposed by the Ministry of Transport. This fine can vary among the federal states of Austria. According to information collected from Austrian colleagues, presently, all kinds of lamps can be used, e.g. special daytime running lamps, integrated fog lamps, dipped headlights, etc. According to results of a recent survey, the DRL usage rate in the country is relatively high: in 2006, it was 94% for motorways, 90% for rural roads, 87% for urban roads and 91% as a combined value for all roads. This is partly due to the fact that the lamps are switched on automatically in most cars when starting the engine.

The DRL usage rates are one of the ten groups of indicators produced by the Traffic behaviour monitoring system in Finland (Luukkanen, 2003). The system was launched in 1992, and presently is maintained by Liikenneturva – the Central Organisation for Traffic Safety in Finland.

Traffic behaviour data are annually collected using the same methods and the same measuring points. Annual DRL observations cover more than 21,000 vehicles over the country. A DRL indicator applied is the percentage usage rate, inside and outside built-up areas, where the rates are estimated by proportioning the number of observations to the population of the provinces to which they refer to (Luukkanen, 2003). Figure 6.1 presents the monitoring results for the years 1993-2002.

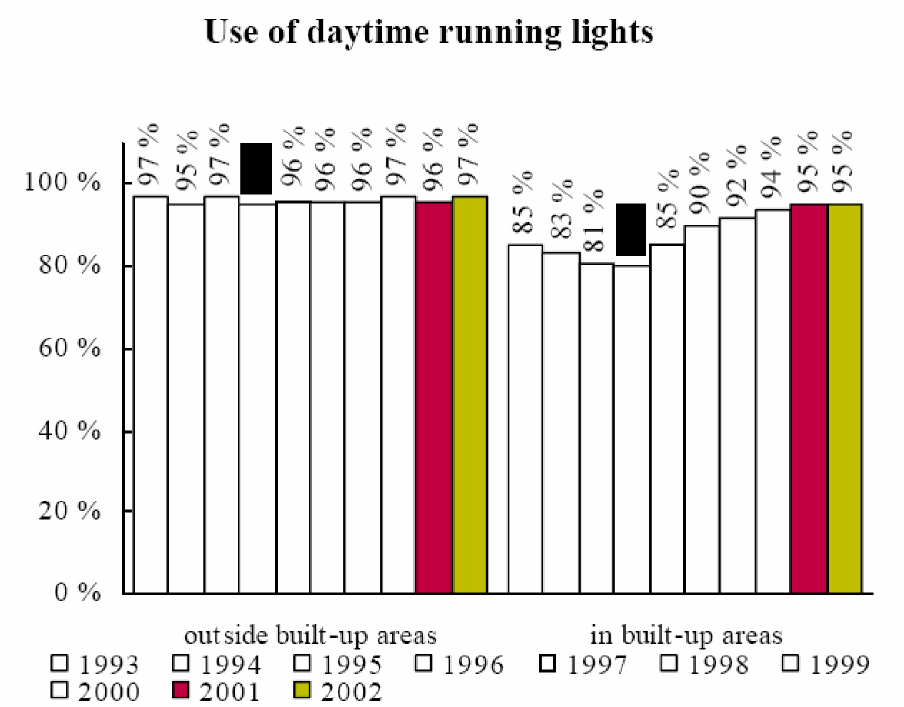


Figure 6.1 Use of daytime running lights in Finland: results of systematic surveys in 1993-2002. (Luukkanen 2003.)

In Hungary the DRL usage rates have been observed since 1993, the year of introduction of this measure. Each year more than 10,000 vehicles are being observed during the survey.

In Figure 6.2 the DRL usage rates can be seen according to road categories. (In Hungary DRL is obligatory outside built-up areas only.) Section 6.2.1 provides further details on the surveys' performance.

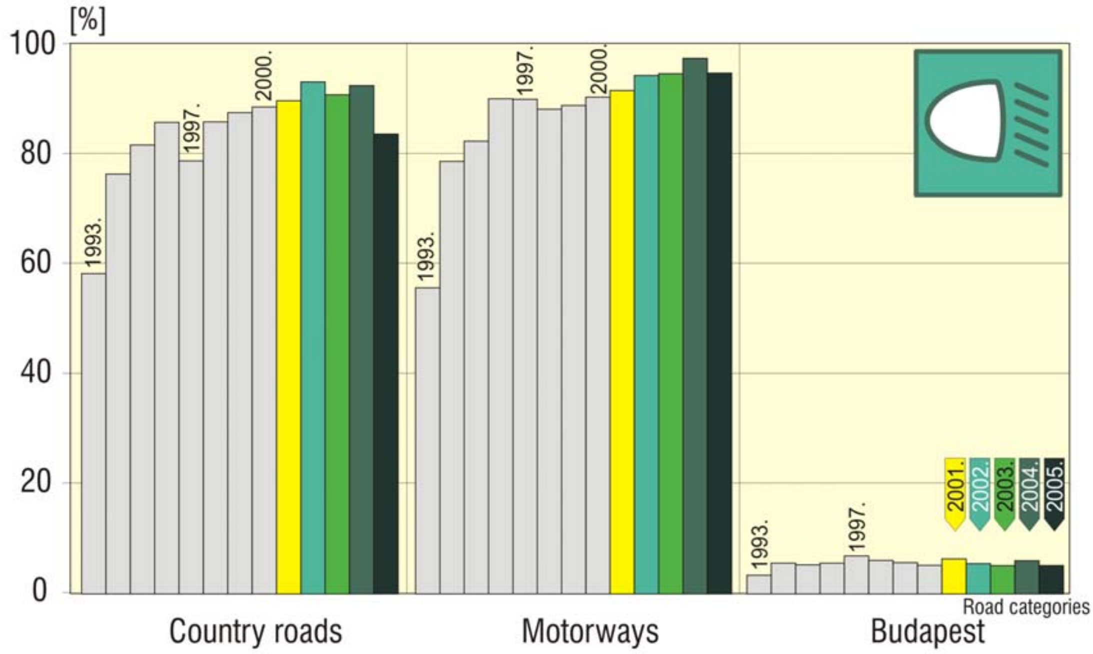


Figure 6.2 Use of daytime running lights in Hungary: results of systematic surveys from 1993 to 2005.

6.2.1 Detailed description of available best practice - Hungary

In Hungary, systematic observations of DRL usage rates have taken place since 1993. The DRL use related data refer to vehicles running with prescribed low beams or special daytime running lamps only; those using the position (parking) lights are not taken into consideration.

Surveys have always been made under clear weather conditions, in order to prevent the effect of the weather and perception circumstances influencing the usage rate of the running lights. In the case of unfavourable weather conditions the usage rate is usually higher, which has also been proven by previous measurements.

Every year since 1992, the data have been recorded on three different types of locations: in Budapest (urban area), on a highway (rural area) and on a motorway. On the highway, the observation sites have always been the same; on the motorway, the data were collected while the vehicles were in motion, and permanent locations were also used.

The total number of observation sites is 5 to 6. This means that there are one to two sites on each road category. The sample size per road type is at least 3300, so that the total number of observed motor vehicles is higher than 10,000.

Every year, almost without exception, the data were collected in the second quarter of the year.

The observation data are registered by tape recorder on the site, to be processed later. Estimated results of the long-term monitoring are presented above in Figure 6.2.

6.3 Recommended system for producing national DRL SPIs

6.3.1 General

The purpose of the survey's performance is to estimate the national level of the DRL usage rates.

The DRL SPIs are suggested in the form of percentage of vehicles using daytime running lights. The value is estimated for different road categories and for different vehicle types. Daytime is defined from dawn plus 1 hour until sundown minus 1 hour.

The road categories considered are:

A motorways,

B rural roads,

C urban roads,

D DRL-roads,

where the term "DRL roads" implies the road categories where the DRL use is obligatory. For example, in Hungary, DRL roads are those which are outside built-up areas; for Austria – all roads (see Table 6.1).

The vehicle types considered are:

1 cars,

2 heavy good vehicles,

3 motorcycles and

4 mopeds.

Additionally, a general value for the whole road network is estimated based on the whole sample.

In total, nine DRL SPIs are estimated.

6.3.2 Method of data collection

To estimate the above SPIs, a country should perform an annual survey of DRL use.

The demands from the annual survey are as follows:

a. Every year, similar observation conditions should be ensured for the survey performance. This concerns: the season of survey, days of survey, types and the number of sites, counting procedure.

b. The survey should be carried out in the period of obligatory use of DRL.

c. Working days with good weather conditions are recommended.

d. The counting procedure can be automatic or manual.

> *Manual procedure* means that observers are involved in recording the data. In the case of manual registration, using a small tape recorder (dictaphon) could be advantageous.
>
> <u>Note</u>: As known, on the site, the time for the data registration is not always enough, therefore using a tape recorder makes it possible to process the data later. This methodology is especially useful when the DRL observations are combined with observations of another behaviour, e.g. safety belt use (like in Hungary). In this case, the amount of information items to be recorded is so wide, that sometimes it is almost impossible to register exact data on the site.

Automatic counting procedure implies the use of a video recorder. This method gives the most comfortable and reliable way of data processing and evaluation. (For example, the picture can be stopped in order to study the details).

e. The observations should be performed *in motion*: the observation vehicle (with observers and/ or a video-recorder) travels in one direction of the road, when the observers or the video-recorder register the details of on-coming vehicles (i.e. from the opposite travel direction). Later, the observation vehicle can go in the opposite direction of the same road.

f. The observation sites are *road sections* of several km in length.

g. For each vehicle observed and included in the sample, the following data should be collected:

- Day and hour of observation
- Road category
- No of site (according to road number and road section)
- Lane of travel
- Direction of travel
- Vehicle type
- Using DRL: yes/ no

Foreign vehicles should be excluded from consideration.

6.3.3 Sampling

The sampling units are the vehicles on the roads.

The sample should be stratified according to three road categories: A – motorways, B – rural roads, C – urban roads.

In each stratum (road category), at least 10 (recommended - 30) observation sites should be selected at random, accounting for the total amount of roads in the national road network. Each site is a road section of certain category of roads.

The general strategy is that for each road category and for each vehicles type a sub-sample should be over 400 vehicles. This demand on the sample size is dictated by two considerations:

(a) the desired accuracy of the estimate and

(b) the actual level of DRL use that is observed in the country.

The accuracy of estimation is stated in terms of sampling error or the confidence interval width (e.g. less than 5% deviation from the population proportion) and the confidence interval level (e.g. 95%) – see Chapter 2.

As demonstrated in Chapter 2, for a sampling error of 5% and a confidence level of 95%, a sample size of 400 vehicles is enough for the safest option, i.e. when no preliminary information about the population proportion (i.e. the DRL usage rate) is available and therefore the usage rate of 0.5 is assumed. At the same time, in the cases of higher (e.g. 0.8-1.0) or lower (e.g. 0-0.2) levels of the DRL use sub-samples of less than 400 vehicles might be sufficient.

The number of observations per site should be 30 at least.

The above demands on sample sizes are suitable when the aim is estimation of the DRL usage rate per road type or for the total road network. In the case, when one wants to compare selected sites or strata, a more detailed consideration of demands on the number of sites and the number of observations per site is required – see *Statistical Appendix* to this report. Besides, one should remember that for very small proportions of presence of the feature the sample sizes should be large enough, otherwise the normality assumptions do not hold and it will be difficult to make comparisons (see details in Sections 3-5 of the *Appendix*).

The numbers of observation sites (with a sample size per site), which are the minimum and recommended figures for the DRL survey performance, are presented in Table 6.2.

a) Minimum number of sites

Strata: road category	Number of sites	Minimal sample size: number of observation units per site (vehicles)	Sub-sample according to stratum (vehicles)
A	10	50	500
B	10	50	500
C	10	50	500
Total sample	30	--	1500

b) Recommended number of sites

Strata: Road category	Number of sites	Recommended sample size: number of observation units per site (vehicles)	Sub-sample according to stratum (vehicles)
A	10-30	100-200	3000
B	10-30	100-200	3000
C	10-30	100-200	3000
Total sample	30-90	--	9000

Table 6.2. Number of observation sites and sample size per site for the DRL survey.

Stratifying the sample according to road categories, a total sample of at least 1500 vehicles is expected. However, in order to obtain reasonably large samples also according to vehicle types, the recommended total sample size should be of about 9000 vehicles.

<u>Note</u>: in the case when for a certain vehicle type (e.g. motorcycles, mopeds), a sufficient sub-sample was not attained, a separate/ complementary survey for this particular group may be arranged.

6.3.4 Calculating SPIs

As explained in Sec. 6.3.1, nine DRL SPIs should be estimated, including four values according to road categories (three strata plus "DRL-roads"), a general value for the whole road network and four values according to vehicle types.

(1) The DRL use rates according to road categories (strata of the sample) are estimated as follows:

$$SPI\_X = \frac{N_{useX}}{N_{totalX}} \cdot 100 \qquad \text{Formula 1}$$

where:

$X \in \{A, B, C\}$ – road categories,

N_{useX} - the number of vehicles using DRL in road traffic, as observed during the survey on road category *X*,

N_{totalX} - the total number of vehicles observed during the survey on road category *X*.

The 95%-confidence interval of the above estimate will be:

$$[p - z\sqrt{\frac{p(1-p)}{n}}; p + z\sqrt{\frac{p(1-p)}{n}}] \qquad \text{Formula 2}$$

where:

p = SPI_X/100,

$Var(p) = \frac{p(1-p)}{n}$,

$n = N_{totalX}$; $z=1.96$.

<u>Note</u>: The above formulae 1-2 are suitable when we apply **equal weights** for different sites in the same stratum (and the number of observations for each site is similar). When the weights are **not equal** (and/or the number of observations varies between the sites), the calculation should be performed as follows.

The population proportion is estimated as

$$p = \sum_{h=1}^{h=H} w_h p_h \qquad \text{Formula 3}$$

where:

p_h - the proportion of vehicles using DRL in road traffic at site *h*;

w_h - the weight of site *h*;

H – the total number of sites for this stratum.

As mentioned above, the observation sites are road sections of certain category of roads. Hence, the site weights may be assigned according to traffic volumes or vehicle-kilometres travelled on these road sections, where: $\sum_{h=1}^{H} w_h = 1$.

The 95%-confidence interval of the above estimate will be:

$$[p - 1.96\sqrt{Var(p)}; p + 1.96\sqrt{Var(p)}]$$

where:

$$Var(p) = \sum_{h=1}^{H} w_h^2 \frac{p_h(1-p_h)}{n_h} \qquad \text{Formula 4}$$

n_h – the number of vehicles observed during the survey at site *h*;

p_h, w_h, *H* as described above.

More details on estimating confidence intervals may be found in *Appendix*.

(2) The total DRL use rate – for the whole road network, is estimated as

$$SPI = \sum_{i=A}^{C} \lambda_i SPI\_i \qquad \text{Formula 5}$$

where:

SPI_i are separate SPIs, one for each stratum (road category), i.e. SPI_A, SPI_B and SPI_C;

λ_i designates the portion of total roads that is attributable to road category A (λ_A), B (λ_B) and C (λ_C), where $\lambda_A+\lambda_B+\lambda_C=1$. These values should (preferably) be the portions of vehicle-kilometres travelled on certain road categories, out of the total kilometre-age. If the kilometre-age values are not available, then as a substitute, the lengths of certain categories of roads out of the total road length, in the country, can be applied.

The 95%-confidence interval of the above estimate will be:

$$[SPI - 1.96\sqrt{Var(SPI)}; SPI + 1.96\sqrt{Var(SPI)}]$$ Formula 6

where:

$$Var(SPI) = \sum_{i=A}^{C} \lambda_i^2 Var(SPI\_i) \ .$$

(3) The DRL use rate for the DRL roads.

In the case when the DRL roads belong to one road category, the estimate obtained for this road category (see clause (1) above) will represent "the DRL roads".

In some countries, the DRL use is obligatory for all roads. In this case, the total DRL use rate (see clause (2) above) will represent the rate for "the DRL roads".

In the case when the DRL roads combine two of three road categories, the estimation method described in clause (2) may be applied, where the portions of total roads attributable to the road categories considered should be normalized.

For example, if "the DRL roads" include A and B road categories, then the normalized values of λ_i will be:

$\lambda_{Anorm} = \lambda_A / (\lambda_A+\lambda_B)$ Formula 7

$\lambda_{Bnorm} = \lambda_B / (\lambda_A+\lambda_B)$.

(4) The DRL use rates according to vehicle types are estimated as follows:

$$SPI\_j = \sum_{i=1}^{N} R_{ij} \times TS_{ij} \quad \text{while: } \sum_{i=1}^{N} TS_{ij} = 1$$

where:

$j \in \{1,2,3,4\}$ – vehicle types,

$i \in \{A, B, C\}$ – road categories,

N – the total number of road categories (N=3),

R_{ij} - the share of vehicles of type *j* using DRL on *i* road category, i.e. the number of vehicles of type *j* using DRL divided by the total amount of vehicles of type *j*, where the vehicles are observed on *i* road category;

TS_{ij} - the "traffic share" of vehicle type *j* for road category *i*, i.e. the amount of kilometres travelled on road category *i* by vehicle type *j*, divided by the total amount of kilometres travelled by the vehicle type *j* on all road categories considered.

Note: if the "traffic share" is not available for each vehicle type, then as a substitute the shares of total roads attributable to certain road categories can be used for all vehicle types. This means that: $TS_{ij} = TS_i = \lambda_i$ (see clause (2) above).

The estimation of confidence interval is similar to that presented in clause (2) above.

6.3.5 Example of calculating DRL SPIs

In 2005, for Hungary, in the DRL survey the data were collected as presented in Table 6.3.

We should note that the data per observation sites were not available in this case. Hence, estimating the DRL use rates per stratum (road category) we applied formulae 1-2 and not formulae 3-4 (see Sec.6.3.4).

Road category	Total number of vehicles observed	Of them: number of vehicles using DRL
Motorways	3270	3097
Rural roads	3420	2856
Urban roads	3360	176
Total	10050	6129

Table 6.3. DRL survey data for Hungary, in 2005.

The data according to vehicle types were not available.

For the same year (2005), the shares of vehicle-kilometres travelled on certain road categories were:

-on motorways: 18.3%,

-on rural roads (outside built-up areas): 56.7%,

-on main roads inside built-up areas: 25%.

Using the above data, the DRL use rates according to road categories are estimated as presented in Table 6.4.

Road category	Number of observed vehicles	Number of vehicles with DRL	DRL use rate (SPI_X*)	p#	Var(p)#	The 95% confidence interval#	
Motorways	3270	3097	94.7%	0.947	1.53E-05	0.939	0.955
Rural roads	3420	2856	83.5%	0.835	4.03E-05	0.823	0.848
Urban roads	3360	176	5.2%	0.052	1.48E-05	0.045	0.060

* see formula 1
# see formula 2

Table 6.4. Estimated DRL usage rates for different road categories, in Hungary, in 2005

The total DRL use rate, for the whole road network, is estimated using formulae 5-6. The calculation is presented in Table 6.5. We obtained: SPI= 66.0% with the 95% confidence interval of [65.2%; 66.7%].

Road category	p*	Var(p)*	λi #	[p* λi] **	[Var(p)* λi^2] ##		
Motorways	0.947	1.53E-05	0.183	0.173	5.13E-07		
Rural roads	0.835	4.03E-05	0.567	0.473	1.29E-05		
Urban roads	0.052	1.48E-05	0.250	0.013	9.23E-07	The 95% confidence interval ##	
Total (sum)			1.0	**SPI=0.660**	**Var(SPI)=** 1.438E-05	**0.652**	**0.667**

* see Table 6.4
\# see shares of vehicle-kilometres travelled on certain road categories
** see formula 5
\## see formula 6

Table 6.5. Estimated DRL usage rate for the whole road network – Hungary, 2005.

The DRL use rate for "the DRL roads" (outside built-up area) is estimated using formula 7 for normalized weights and then, formulae 5-6. The calculation is presented in Table 6.6.

We obtained: SPI= 86.2% with the 95% confidence interval of [85.3%; 87.2%].

Road categories	p*	Var(p)*	λ_{inorm}**	[p* λ_{inorm}] ##	[Var(p)* λi^2] ##		
Motorways	0.947	1.53E-05	0.244	0.231	9.12E-07		
Rural roads	0.835	4.03E-05	0.756	0.631	2.30E-05	The 95% confidence interval##	
Total (sum)			1.0	**SPI=0.862**	**Var(SPI)=** 2.39E-05	**0.853**	**0.872**

* see Table 6.4
** see formula 7
\## see formulae 5-6

Table 6.6. Estimated DRL usage rate for the "DRL roads" – Hungary, 2005.

In summary, the DRL SPIs estimated for Hungary (in 2005) are as follows:

- the DRL use rates according to road categories
  - for motorways: 94.7% (CI: 93.9%; 95.5%),
  - for rural roads: 83.5% (CI: 82.3%; 84.8%),
  - for urban roads: 5.2% (CI: 4.5%; 6.0%);
- the total DRL use rate, for all roads: 66.0% (CI: 65.2%; 66.7%);
- the DRL use rate for "the DRL roads": 86.2% (CI: 85.3%; 87.2%).

(The DRL use rates according to vehicle types cannot be estimated due to lack of data.)

6.3.6 Other related issues

Quality control of the DRL survey's results should refer to:

(a) sampling rules applied to the data collection;

(b) values of the DRL SPIs and their confidence intervals.

Reliability of the estimates obtained is assessed by comparing the results to former studies on the same issue.

It is recommended that each DRL SPI is presented by a separate time-series. Such a presentation enables:

- consideration through time of changes of the DRL usage rates;
- comparison of DRL usage rates for different road categories/ vehicle types.

For example, the above Fig. 6.2 shows the DRL usage rate in Hungary according to road categories. On the so-called "DRL-roads" (roads outside built-up areas), the usage rate is relatively high, in spite of the fact that the DRL use has no priority in road safety campaigns and police enforcement. In general, the DRL usage rates show a long-term increasing trend on country roads and on motorways; at the same time, on country roads there was a decrease in 2005. It is understandable that the DRL usage rate is low in Budapest, because urban roads do not belong to the group of “DRL-roads”, at present.

6.4 Discussion

The suggested system of DRL SPIs enables estimation of the DRL usage rates at the national level. The system may serve as a background for both the countries' comparisons and along-time considerations on the DRL-related issues.

The recommended system for producing DRL SPIs may also be applicable on a regional/ local level, if necessary adjustments are performed in the sampling procedures.

One should remember that the background information on the DRL legislation is essential for a correct interpretation and comparison of the results. For example, comparing the countries' DRL usage rates it is reasonable to take into account whether the countries have a law/ regulation on obligatory use of DRL and if they do, how long.

In countries, where automatic DRL was introduced a long time ago (e.g. Sweden, Norway), according to expert estimates, current DRL usage rate is close to 100%, thus the DRL usage rate as a behavioural safety performance indicator does not have practical implications any more. In general, once the option of automatic DRL is introduced Europe-wide, the DRL indicators will lose their importance.

6.5 References

1. DRL (2006). Saving Lives with Daytime Running Lights (DRL). A Consultation Paper, DG TREN E3, Brussels, 1 August 2006.
2. Luukkanen L. (2003). Safety management system and transport safety performance indicators in Finland. Liikenneturva – Central Organisation for Traffic Safety in Finland.
3. Koornstra, M.; Bijleveld, F.; Hagenzieker, M. (1997). The Safety Effects of Daytime Running Lights, SWOV report R-97-36, Leidschendam.
4. Vis, M.A. and A.L. van Gent (Eds.) (2007) Road Safety Performance Indicators: Country Comparisons. Deliverable D3.7a of the EU FP6 project SafetyNet.

7 Detailed manual for SPIs on vehicles

RACKLIFF[A], L., HADDAK[B], M.

[A]VSRC, [B]INRETS

7.1 Introduction

The SPI that this task is concerned with relates to the level of protection afforded by the vehicles which constitute the fleet in each EU Member State. When crashes occur, the potential of the vehicle itself to prevent (or indeed cause) injuries can determine whether the outcome is a fatality or something less serious. This SPI area differs from, for example, Speed, Alcohol and Drugs, Roads, since it is fatalities not crashes that it seeks to address. Improvements in passive safety do not affect the occurrence of accidents, but help to minimise the consequences when crashes happen. The unsafe operational conditions could be defined as:

1) the presence within the fleet of a number of vehicles that will not protect the occupant well in a collision (crashworthiness);
2) the presence within the fleet of a number of vehicles with an increased capacity to inflict injury (compatibility).

The aim of these SPIs is to construct a measurable variable that will tell us what these numbers are and their significance in each Member State.

There have been significant developments in passive safety technology over the years; for occupants, enhanced performance from restraint systems, supplementary airbags, better structural integrity and better compatibility between cars and HGVs, leading to a far better crash outcome for occupants of newer cars compared to occupants of older cars.

The contribution to fatalities of factors such as inappropriate speed, impairment and failure to use protective systems (e.g. ERSO, 2007; WHO, 2004) is well documented, and many governments have implemented policies aimed at reducing the incidence of this type of risky behaviour. In the UK, for example, the Department for Transport's "Think" campaign has incorporated measures aimed at restraint use, impairment and speed (UK, 2007). However, whilst vehicle manufacturers may have invested heavily in the development of passive safety technologies, and initiatives like EuroNCAP have raised awareness of the importance of it amongst consumers, crashworthiness and compatibility have historically not enjoyed the same profile in national road safety campaigns.

This may be because the mix of vehicles which constitutes a national fleet is the result of so many factors, most of them out of the control of the responsible authority for road safety. They include, for example, geographical factors (countries on a central European crossroads may have more large goods vehicles than small territories on the periphery), economic factors (countries with lower incomes will, on average have older vehicles), socio-demographic and human factors (is the population ageing, what are people's mobility needs? – e.g., Langey (2001).

However, there are ways in which transport policies can influence fleet composition:

- By providing fiscal incentives on new car purchases, to encourage vehicle replacement. There is evidence from Greece (Vis and van Gent, 2007) that this type of policy can lead to a measurable effect on the fleet composition.
- By favouring other modes, and providing incentives for switching freight from roads to rail, water or air, the percentage of heavy goods vehicles on the road could be reduced.

- By providing incentives for the use of public transport (especially non-road based public transport) and to reduce the numbers of motorcycles or other vulnerable road user groups.

Monitoring of the fleet, in terms of its composition and age, can assist countries in designing policies to ensure that road users are not placed at unnecessary risk of injury as a result of poor vehicles or an unfavourable fleet mix.

7.2 Vehicle fleet measurement methods and tools – existing practices

7.2.1 Examples of vehicle fleet measurement methods

Again, the Vehicles (Passive Safety) area differs from the other SPI areas, as it deals with population not sample data. In other words, whilst protective systems use, for example, must be estimated using methods that probably differ from country to country, total numbers of vehicles are taken from national databases, which are intended to be a complete record.

Preliminary results for SafetyNet work package 2 (Risk-Exposure Data) indicate that all of the countries in Europe maintain a database of vehicle registrations. The minimum information which is required to produce some calculations of vehicle age (as a proxy for vehicle crashworthiness) and fleet composition (as a measure of compatibility), are total number of vehicles listed by:

- Year of manufacture (or year of first registration)
- Vehicle type (using definitions compatible with CARE)

For those countries that are able to provide it, listing passenger cars by make, model and year of manufacture also enables calculation of the EuroNCAP score of vehicles manufactured after 1994.

7.2.2 Detailed description of available best practice

The “best practice” for vehicle fleet databases is to ensure that systems exist for maintaining the accuracy of the database. This means that procedures should be in place to:

1) Remove scrapped vehicles from the database.

2) Ensure that vehicles that are not taxed and/or licensed still appear on the database if they are still being used on the roads.

One way of ensuring that data meets these requirements is that vehicle fleet data is the responsibility of a national governmental body.

These issues are related to the accuracy of the vehicle fleet database as a record of the vehicles that are actually on a country’s roads at a point in time. Additionally, there are minimum requirements for the level of detail contained within the database, to make calculation of the SPI for vehicles (passive safety) possible. These additional requirements include:

1) Provide detailed and accurate descriptions of vehicle makes and models.

2) Classify vehicles according to vehicle-types compatible with CARE definitions.

3) Distinguish between smaller (less than 3.5 tonnes) and larger goods vehicles, since these are significantly different when assessing their compatibility in collisions with passenger cars or vulnerable road users.

4) Register all motorised vehicles, including public service vehicles and mopeds.

7.3 Discussion

Assessment of the age and composition of the national vehicle fleet has not, historically, been considered to be a high-priority element in determining national road safety policies. However, according to the Cars 21 project,

"*the combination of EU legislation for crash test standards and improved consumer information through the EuroNCAP programme has substantially raised the survivability of vehicle occupants in a crash. The issue of fleet renewal should also be given consideration by policymakers as it can have important environmental and safety implications. A vehicle fleet with a high average age of vehicles tends to have a negative effect on road safety and the environment and if vehicle owners retain their old vehicles for longer periods the market penetration of new better performing vehicles is slowed down.*"

This would suggest that the monitoring of an SPI for vehicle fleet crashworthiness and compatibility could be a useful and important element of a full road safety program.

7.4 References

1. ERSO (2007). European Road Safety Observatory. Internet site: http://www.erso.eu/knowledge/content/05_alcohol/alcohol.htm; http://www.erso.eu/knowledge/content/20_speed/speeding.htm
2. Langey, X. (2001) International comparisons: we need to know a lot more, Injury Prevention, 2001 (7): pp 267-269.
3. UK (2007). Internet site: http://www.thinkroadsafety.gov.uk
4. Vis, M.A. and A.L. van Gent (Eds.) (2007) Road Safety Performance Indicators: Country Comparisons. Deliverable D3.7a of the EU FP6 project SafetyNet.
5. WHO (2004). Report on Road Traffic Injury Prevention, World Health Organisation (WHO), Geneva.

8 Detailed manual for SPIs on roads

DUIVENVOORDEN[A], K., WEIJERMARS[A], W., LOUWERSE[A], R., PAPADIMITRIOU[B], E., YANNIS[B], G.

[A]SWOV, [B] NTUA

8.1 Introduction

SPIs for roads aim to assess the safety hazards by infrastructure layout and design. Two SPIs for roads were developed, namely the road network SPI and the road design SPI (Hakkert et al., 2007). As these are new SPIs developed for the SafetyNet project, no examples were available. Therefore, an example case study was executed. In the case study, the process is automated and the SPIs are calculated for an area in the Netherlands. The purpose of this case study is twofold. First, it illustrates the rather theoretical concept of the SPIs. Second, it provides more insight into the practical applicability and strengths and weaknesses of the method and provides guidelines for future use.

In section 8.2 the methods for calculating the SPIs are summarized. Section 8.3 discusses which data should be collected and provides some case studies on the collection of data on roads. Section 8.4 describes the calculation of the SPIs for the case study area and in section 8.5 the lessons learned from this example are discussed.

8.2 Summary of theory

Two SPIs for roads were developed, namely the road network SPI and the road design SPI. The road network SPI indicates whether the actual road category is appropriate given the urban areas that it connects. The road design SPI determines the level of safety of the existing roads. The SPIs are shortly described in Sections 8.2.1-8.2.2. For a more detailed description of the method, the reader is referred to Hakkert et al. (2007).

8.2.1 The road network SPI

The road network SPI is based on a rather quantitative method for the assessment of network and design quality aspects of a safe road infrastructure at the regional level presented in a Dutch study of Dijkstra (2003a). In the Dutch study the roads in the investigated network were compared to the theoretical required roads, which meet the requirements with respect to road safety.

The classification method of urban areas in the Dutch study originated from the rather descriptive and qualitative method of the German guidelines for road categories (FGSV, 1988). It is recognised that urban areas differ from each other in many ways. These guidelines defined a classification of roads and centre types in a qualitative way. In the Dutch study the method of FGSV has been adjusted to a more quantitative method for the assessment of the infrastructure. The road network SPI is also a quantitative method.

The road network SPI investigates whether the actual road category meets the road category that should be present given the sizes of the urban areas that it connects. To obtain a road network SPI that allows for international comparison, an internationally harmonized road categorization is proposed. Table 8.1 presents the functional road classification that is proposed for SafetyNet together with the preferred characteristics. Roads are classified into six categories ranging from AAA to C. This classification is restricted to rural roads and motorways.

	Rural areas (outside built-up areas)					
SafetyNet road classes	AAA: Motorway	AA: A-level road 1	A: A-level road 2	BB: Rural distributor road 1	B: Rural distributor road 2	C: Rural access road
Functional road category	Through-road (road with a flow function)			Distributor road		Access road
Separation of opposing directions	Dual carriageway	Dual carriageway	Single carriageway, preferable with lane separation	Dual carriageway	Single carriageway, preferable with lane separation	Single carriage way
Lane configuration	2x2 or more	2x1, 2x2	1x2, 1x3, (1x4)	2x1, 2x2	1x2, 1x3, (1x4)	1x2, 1x1
Obstacle-free zone	Very wide or safety barrier	Wide or safety barrier	Wide or safety barrier	Medium	medium	small
Intersections	Grade-separated	Preferable grade-separated	Preferable grade-separated	Preferable roundabout	Preferable roundabout	

Table 8.1 SafetyNet Functional road classification (Source: Hakkert et al. 2007).

The road category that should be present between different types of urban areas, i.e. for different types of connections, is specified in Table 8.2. Note that none of the connections requires road category A. The A road category (single carriageway) is not considered for any type of relation because the AA road category is preferred for its dual carriageway.

Urban area (# inhabitants)	Type 1	Type 2	Type 3	Type 4	Type 5
Type 1 (>200,000)	AAA	AAA	AA	indirectly	indirectly
Type 2 (100,000 to 200,000)		AA	AA	BB	indirectly
Type 3 (30,000 to 100,000)			BB	BB	B
Type 4 (10,000 to 30,000)				B	B
Type 5 (<10,000)					C

Table 8.2 Connections between different types of urban areas.

According to the original method (FGSV, 1988) the urban area types are determined in a rather qualitative way. For the application in the Netherlands that is described in Dijkstra (2003a) the method of the FGSV was applied in a more quantitative way. The urban areas (centre types) were classified based on the number of inhabitants which was used as a measure for the production and attraction of trips from a centre. This method can be enhanced with other information such as for example industrial areas, shopping areas and recreational sites. For the present case study area the theoretical method was enhanced with industrial areas. For a more detailed description, the reader is referred to section 8.4.1. The urban area types 1 to 5 are defined according to their number of inhabitants. Type 1 is a big city, type 5 is a village, and 2-4 are in-between. The number of inhabitants for each centre type is shown in Table 8.2.

The road network SPI is the percentage of appropriate current road category (AAA-C) length per theoretical road category. Thus, connections between cities are assessed by comparing the theoretical road categories with the actual road categories.

First, it has to be determined which connections are assessed. This is done by using circular search areas. The radius of the search area depends on the distance between the investigated centre and the nearest centre of the same class (N). For all the centres of types N+1 and N+2 within the search area, the connections with the investigated centre need to be assessed. Also the connection between the two centres of the same class N needs to be assessed. There is one exception; namely when searching for type 5 centres from centre type 3. Here, the search area is adjusted: from the centre type 3 to the nearest centre type 4. This prevents taking into account centre types 5 which probably do not have a relation with the centre type 3. Table 8.3 gives an overview of the search for the centres.

Start centre	**Search for the centre of the same type (radius)**	**Centres in search area**	**Assessment of the connections between**
1	The nearest 1	2 and 3	1 and 1, 1 and 2, 1 and 3
2	The nearest 2	3 and 4	2 and 2, 2 and 3, 2 and 4
3	The nearest 3	4	3 and 3, 3 and 4
3	The nearest 4	5	3 and 5
4	The nearest 4	5	4 and 4, 4 and 5

Table 8.3 Overview of the search for centres.

Applying this procedure to all centres results in a list with all connections that need to be assessed. For all these connections, it needs to be determined which road categories they are composed of. Next, the current road category is compared to the theoretically needed road category that follows from Table 8.2. When the current road category is higher than or equal to the road category that should be present according to Table 8.2, the current road is considered to be appropriate from a safety point of view. If the theoretically needed and existing road categories are compared for all connections, the share of appropriate road category length can be calculated for each theoretical road category.

An example: the existing road category between centre X of type 3 and centre Y of type 4 is B. From Table 8.2 however it follows that road category BB is needed between type 3 and type 4 centres. The current road category is thus lower than the theoretically required category which means that the current road category is not appropriate. In case the current type was higher (e.g. AA) or similar (BB), the current connection was “appropriate”.

By following the steps of the method described before, the scores of the road network SPI can be calculated. Per theoretical road category the road network SPI is the percentage of appropriate current road category (AAA-C) length. For a representative result a sample size needs to be chosen. For more information about sample sizes, the reader is referred to Chapter 2.

8.2.2 The road design SPI

The road design SPI is based on the EuroRAP Road Protection Score (RPS). The RPS is a measure for the protection that is provided in relation to three main accident types: run-off road, head-on impacts and severe impacts at intersections. EuroRAP designed a method to calculate the RPS for each road segment or route, expressed in one to four stars. The number of stars depends on a number of road characteristics that are shown in Table 8.4. Unfortunately, the precise method for calculating the number of stars, given the road characteristics is not yet published (which combinations of scores lead to a certain number of stars). For more information on EuroRAP Road Protection Scores see Lynam et al. (2003) and Lynam et al. (2004).

Road characteristics	Classes/values
Speed	50, 60, 70 etc.
Barrier (placement)	Right, left, middle etc.
Barrier (CEN approved)	Yes/No
Median (width)	0-4 meter, 4-10 meter etc.
Hard obstacle point/stretch (distance)	0-3 meter, 3-7 meter etc.
Hard obstacle point/stretch (placement)	Right, left, etc.
Side area cut (placement)	Right, left etc.
Side area embankment (placement)	Right, left etc.
Side area embankment (type)	Gentle, steep
Junctions (not signalized)	3 of 4 arms with or without left turn lane
Junctions (signalized or roundabouts)	Traffic lights, roundabouts
Intersection merging	Long/short
Intersection access	Yes/No

Table 8.4 The classes or values that are used for the scoring of each road characteristic to obtain the RPS during a drive-through inspection, (EuroRAP in preparation; Mobycon 2006).

The input for the road design SPI is the RPS of EuroRAP. The next step is to calculate the distribution of stars for the six road categories in the functional road categorisation (as described in Table 8.1).

An example: let there be 10 km of roads of road category B. When 4 km has 3 stars and 6 km 2 stars, the distribution for road category B is: 4 stars: 0%, 3 stars: 40%, 2 stars: 60% and 1 star: 0%.

The road design SPI is the distribution of stars (1-4) per road category (AAA-C).

8.3 Data

This section deals with the collection of the necessary data. Section 8.3.1 provides an overview of the data that is needed for the calculation of the SPIs and discusses some general data requirements. Section 8.3.2 describes some case studies on data collection from various European countries.

8.3.1 Data needed for the calculation of SPIs

8.3.1.1 The road network SPI

For the calculation of the road network SPI, we need the following information:

- Location of urban centres
- Number of inhabitants per urban centres
- Location of roads that connect centres
- Road categories of actual roads (expressed in AAA to C)
- Length of roads

The locations of urban centres and roads are displayed on (digital) maps. The location of urban centres and the number of inhabitants of these centres are used to produce a list of connections that are assessed.

To be able to determine whether the road category is appropriate, the actual road category should be known. Road classifications differ by country and the present classes have to be translated to the road categories specified in Table 8.1 using the criteria from this table.

Finally, the lengths of roads of different categories have to be known in order to calculate the percentage of roads that are of an appropriate category.

8.3.1.2 The road design SPI

First one needs to select the sample size for the road design SPI. For more information about sample sizes the reader is referred to Chapter 2. For the calculation of the road design SPI we need the following information:

- EuroRAP Road Protection Score (RPS) per road segment or route
- Road length per road segment or route
- Road category of the current network

For the calculation of the EuroRAP RPS scores per road category, countries are advised to contact EuroRAP. Table 8.4 provides an idea of the data that is necessary for the calculation of the scores. In order to determine the distribution of stars per road category, the length and category of each road have to be known.

All the collected data has to meet some basic requirements. Naturally, it has to be of sufficient quality. The amount of invalid data should be minimised. Besides, it has to meet certain levels of accuracy. For the level of accuracy of the data that is necessary to calculate the RPS score per road, the reader is referred to EuroRAP (2006). The necessary levels of accuracy of the data needed for the calculation of the network SPI are discussed here. With regard to the number of inhabitants, it has to be determined to which class the urban area has to be assigned. Also regarding the length of the road, the estimation (in integer or even tens of kilometres) is sufficiently accurate. The digital map should be sufficiently detailed to take all roads and all urban centres into account. With regard to the road category, accuracy does not play a role, i.e. the assigned road category is either correct or incorrect.

The data have to be updated from time to time to account for changes in the road network. We advise to update the data once every few (3 to 5) years.

8.3.2 Case studies on the collection of data

This section describes available data sources and case studies on the collection of data that can be used for the calculation of the described SPIs. Section 8.3.2.1 describes data that is available from digital maps and corresponding geographical databases. Section 8.3.2.2 discusses a case study on the collection of data from various road authorities in a central database in the Netherlands. Finally, section 8.3.2.3 discusses a case study on the collection of data from Greece.

8.3.2.1 Digital maps and geographical databases in Europe

Two of the main manufacturers of digital road maps in Europe are NAVTEQ and Tele Atlas. Their maps are used for logistics, in traffic and fleet management, for car and personal navigation, and so on. Google Earth for instance, uses the maps of Tele Atlas to project road categories on the aerial pictures.

MultiNet is the geographical database of manufacturer Tele Atlas and is available throughout Europe and North-America. Aerial photography and satellite images are used to give overviews of road networks, land use, and urban areas. Traditional paper maps of other mapping organizations are used for cross-checking road classification data. A new development is the use of mobile mapping vehicles (Tele Atlas, 2007b) that gather data as they drive over the European road network. Each vehicle is equipped with digital cameras. The main road features that are being surveyed within MultiNet are functional road class (8 classes for example motorway, secondary road, local road, etc.) and form of way (20 classes, for example part of motorway, part of single carriageway, part of roundabout, part of a pedestrian zone, etc.). Other manufacturers produce maps with more or less similar features and attributes.

8.3.2.2 Digital road database in the Netherlands

The Dutch governmental agency AVV Transport Research Centre (part of the Ministry of Transport, Public Works and Water Management) maintains a digital road database (in Dutch: Wegkenmerken+) which is meant to contain all road segments and the main characteristics. This database is based on the digital National Road Database (NWB), which contains all road segments and junctions of the Dutch road network (AVV, 2005).

It is intended that regional and local road authorities collect specified road features of road segments and junctions in a uniform way, enter this data in Wegkenmerken+ and upload it to AVV. AVV, in return, sends all the collected data per province to enable analysis and benchmarking. Table 8.5 shows the main road features that are being surveyed within the road database Wegkenmerken+. In addition to these features, road authorities can add their own features.

Road segment			Junction
Accesses (amount per segment)	Public transport stops (amount of)		Branches (amount of)
Bicycle facilities	Road axis and edge marking (type)		Intersection type
Carriageways (amount of)	Road surface (type)		Plateau
Emergency recesses/lots	Slow traffic (closed/partly closed)		Right-of- way
Lane separation (type)	Speed enforcement		Roundabout
Lanes (amount of)	Speed limit		Traffic signals
Obstacle free zone	Speed reducing countermeasures		Traffic signs
Parallel road	Traffic Flow (AADT)		
Parking facilities	Traffic function		
Preliminary traffic signs (in advance)			

Table 8.5 Main road features within Wegkenmerken+ (AVV, 2005).

The Sustainably Safe Indicator is incorporated in Wegkenmerken+. Sustainably Safe Indicator (SSI) is an instrument with which the road authority can determine whether the roads meet the Sustainably Safe requirements (Dijkstra 2003b). If the road authorities collected enough data with road features, the Sustainably Safe Indicator will calculate the 'percentage Sustainably Safety'.

Unfortunately, the coverage of most road features is very poor so far. However, at the moment several road authorities are carrying out large inspection programs to fill the database. Currently, the database cannot be used for the calculation of EuroRAP RPS scores since not all necessary data is collected. EuroRAP needs more detailed information about certain characteristics than the features that are been collected within the road database Wegkenmerken+ (for instance the classes of obstacle-free zones or medians, side area embankments, intersection types, etc.).

8.3.2.3 Collection of extra data, Greece

In Greece a survey was carried out for the collection of extra data. It has to be noted that the data collected with the survey is not used for the calculation of the SPI scores.

From an assessment of data availability and quality for the estimation of road SPIs in Greece, it was concluded that the available data on the road network was rather poor. In order to collect the necessary data and to give some good examples for Greece, an additional survey was carried out on the main interurban road network. The main characteristics of the survey are presented below.

The survey was carried out along the two main axes of the interurban road network of Greece, starting from the Athens capital region and expanding within a 300 kilometre radius, including the north interurban axis, connecting Athens with Thessalonica, and the West main interurban axis connecting Athens to the port of Patras and the city of Pyrgos.

The survey form of Figure 8.1 was developed for the data collection. Each row of the survey form corresponds to one section with uniform design characteristics. The duration of the survey was two days, each one devoted to one main interurban axis.

National Technical University of Athens
School of Civil Engineering
Department of Transportation Planning and Engineering

SafetyNet WP3 Task 6 - Safety Performance Indicators for Roads

Survey Form for Road SPIs in Greece

*Population	Area Type
>200000	1
100000-200000	2
30000-100000	3
10000-30000	4
<10000	5

Origin area	Population
Destination area	Population
Road Type*	

Section		Median		Obstacle free zone			Speed	Intersection			
Km		Physical	Axis marking	wide	narrow	barrier	limit	at-grade	at-level		roundabout
									signalized	not signalized	

Figure 8.1 Survey form for collecting road design data.

For the data collection, one vehicle and two individuals (one driver and one passenger) were used and the following process was followed:

- The odometer of the vehicle was set to zero before departure.
- The passenger filled out the first row of the survey form with the road design characteristics of the road upon departure, stating the start of the section as zero (first column of the form).
- Whenever one of the road characteristics changed (e.g. different speed limit, end of median), the passenger asked the driver to report the odometer display. The passenger then filled out the reported kilometre as "end" of the previous section and "start" of the next section and filled out the new characteristics of the road for the next section.
- At the same time, the passenger also counted and noted the number of intersections per type included in each section (with the respective kilometre in brackets, as reported by the driver).

The same process was applied on the return trip as well, for validation purposes.

From this case study, it can be learnt that a survey can be a useful and cost-effective alternative/additional data source for data on road design characteristics. According to Lynam et al. (2004), the factors used for the RPS need to be identifiable through visual inspections. Also for the collection of data on EuroRAP RPS scores, surveys are employed. As mentioned before, countries are advised to contact EuroRAP for the calculation of RPS scores per road segment or route, so that the country surveys can be harmonised to EuroRAP requirements.

8.4 Calculation of the SPIs for the case study area

In section 8.2, the road network and road design SPIs were described. To illustrate and evaluate, these SPIs are applied to the case study area in the Netherlands. Moreover, the process was automated for this area. This section describes this application.

8.4.1 Road network SPI

The calculation of the network SPI for the case study area consists of the following steps:

1. Collection and processing of the data
2. Determination of the list of connections that need to be assessed
3. Comparison of the theoretically required road categories with the actual road categories
4. Calculation of the SPI: Percentage of roads that have an appropriate category

These steps are described in detail below.

Ad 1. Collection and processing of the data

Theoretically, the data that is needed for the network SPI is a map with urban areas and inter-urban roads, the number of inhabitants of urban areas and the road categories and road lengths. All data necessary for the application to the case study area are extracted from the MultiNet database. The data is visualized and processed using ViaStat, a GIS-application developed by VIA, a Dutch consultancy firm.

In the theory document, the number of inhabitants was used as estimation for the size of an urban area. In the application, the theoretical method is enhanced with the amount of industry as a determinant for the size of an urban area. The attraction of traffic by industrial areas is translated into an extra number of inhabitants by multiplying half of the percentage of industrial area with the number of inhabitants and adding this to the number of inhabitants. Only half of the percentage of industrial area is used since it is assumed that less traffic is attracted to industrial areas compared to urban areas in general in which various functions are combined. For the other travel motives, unfortunately no data was available in MultiNet. Therefore, the size of urban areas is determined only by the number of inhabitants and size of industrial areas.

Figure 8.2 shows the centre types and the current road categorisation in this SafetyNet case study.

As mentioned in Section 8.2 the actual road category is expressed in AAA to C. This classification is not used as the standard road classification in the Netherlands. Information on road characteristics from MultiNet is used to determine the road categories (AAA to C). The variables that are used are (also see Table 8.6):

- MultiNet road classes:
    0. Motorway, Freeway or Other Major Road (flow function)
    1. A Major Road Less Important than a Motorway (flow function)
    2. Other Major Road (flow function – distribution function)
    3. Secondary Road (distribution function)
    4. Local connecting road (distribution function – access function)
    5. Local road of high importance (access function)
    6. Local Road (access function)
    7. Local Road of Minor Importance (access function)
    8. Other Road (access function)
- Speed information: to separate the rural distributor roads from the A-level roads.
- Route number: when a route number is present (P), the road is always higher in category than a rural access road. Or there is no route number (absent = A).

- Form of roadway: information about the part of network the road belongs to:
  1. motorway
  2. multi carriage way (not a motorway)
  3. single carriage way

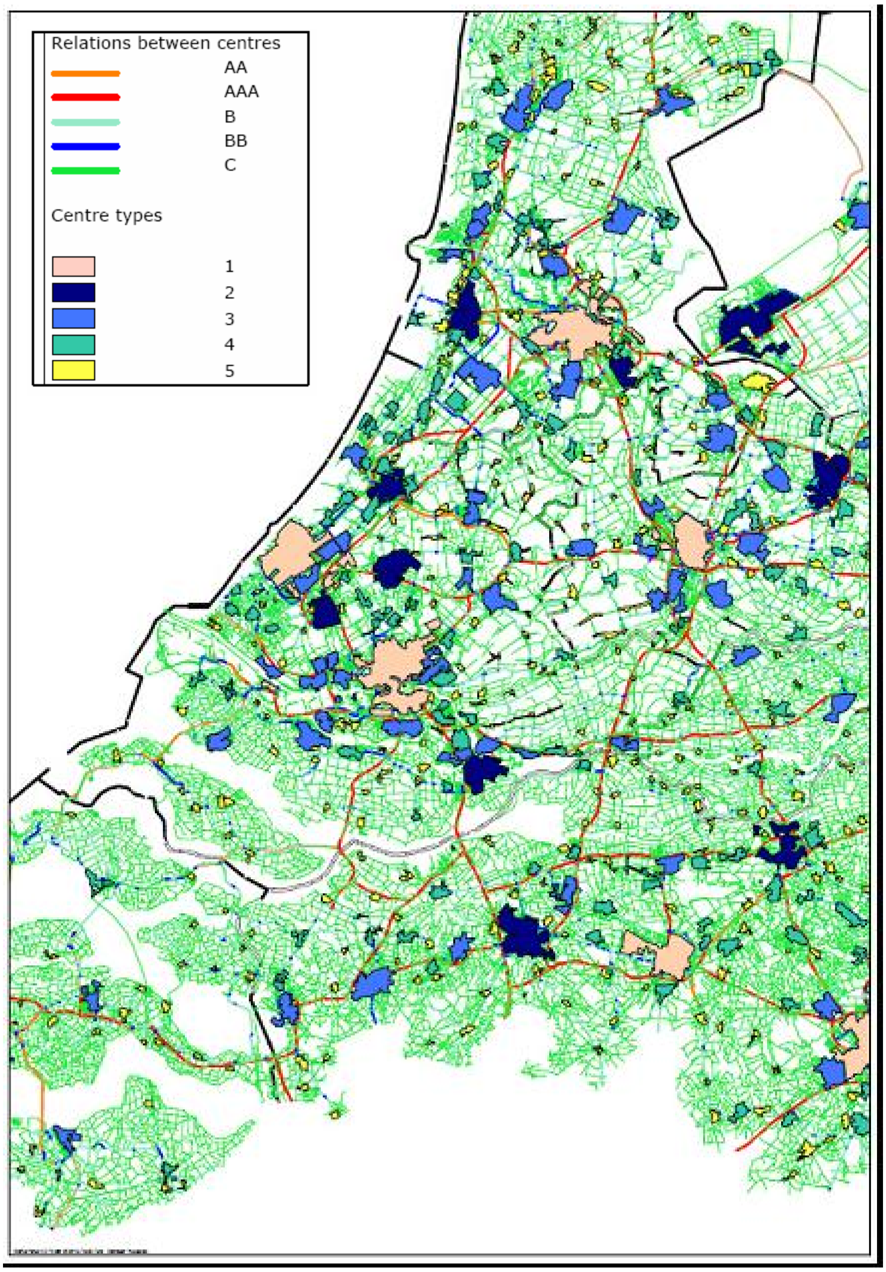


Figure 8.2 The current road categorisation in SafetyNet.

MultiNet road class	0	1		2				3		4		5-8
Speed information				>80km/h		≤80km/h						
Route number										P	A	
Form of way	1	2	3	2	3	2	3	2	3	-	-	-
SafetyNet category	AAA	AA	A	AA	A	BB	B	BB	B	B	C	C

Table 8.6 From the MultiNet road class to the SafetyNet categories.

Ad 2. Determination of the list of connections that need to be assessed

To determine which connections need to be assessed, the search area that is explained in section 8.2.1 is applied. First, for all type 1 centres, the distance to the closest type 1 centre is determined. On the basis of these distances, the search areas are defined. Within these search areas, for all type 2 and type 3 centres, the connections with the type 1 centre are assessed. The procedure is repeated for all type 2, type 3, type 4 and type 5 centres.

Figure 8.3 displays the centres and the connections between these centres, including some connections crossing the province borders. These connections are included because the centres they connect are within the search area of the centres in the case study area. The connections (expressed in required road category AAA-C) between the centres (type 1-5) are displayed by the coloured lines.

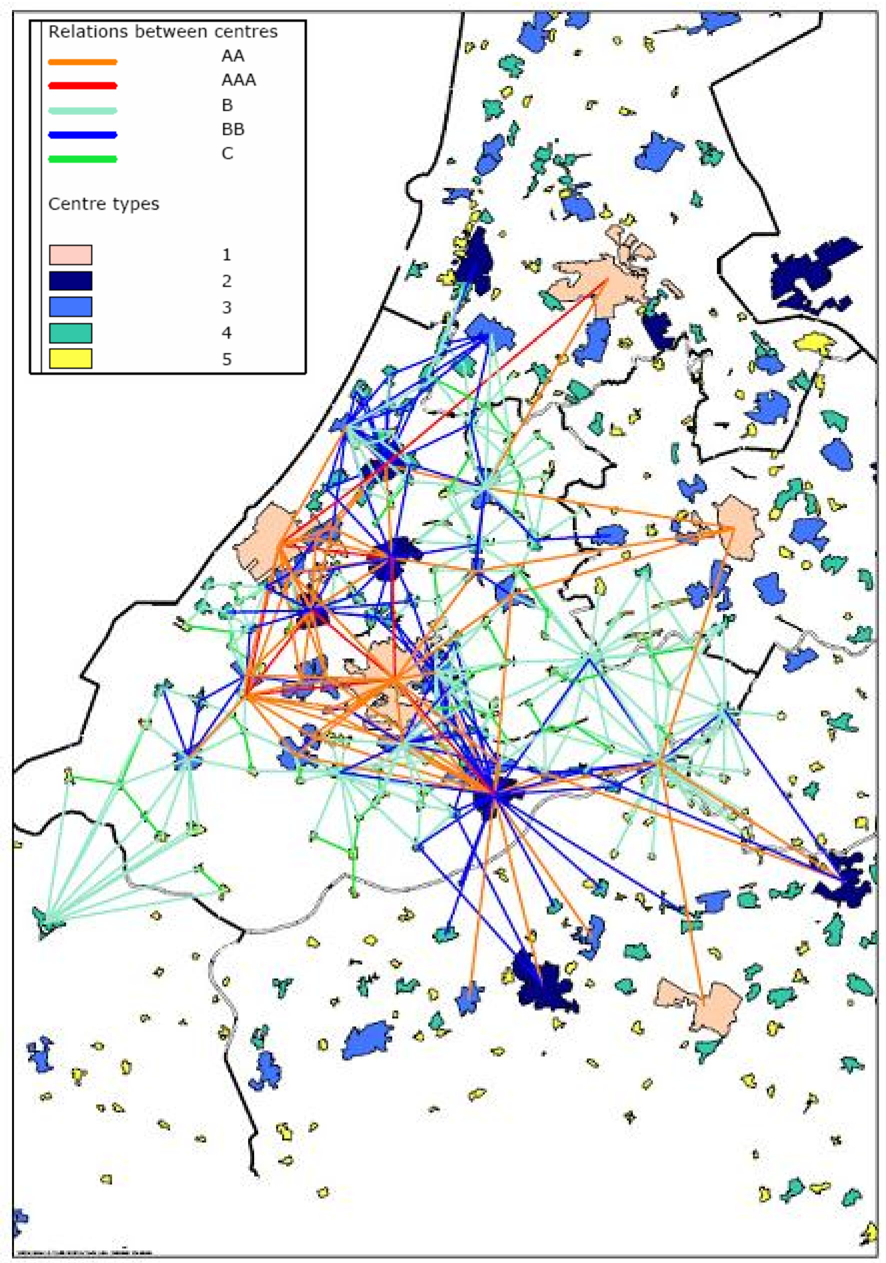


Figure 8.3 The centres and connections in the case study area.

Ad 3. Comparison of the theoretically required road categories with the actual road categories

Once it is known which connections need to be assessed, the theoretically required road categories can be compared to the current network. To be able to compare the theoretical network to the current network in an automated manner, the connections are assigned to the current network, i.e. per connection the most likely route is selected. This is done by applying a deviation factor (detour factor). This detour factor is introduced in Dijkstra (2003a) and is one of the accessibility criteria in relation to the sustainable safety (Duurzaam Veilig) concept. It is expected that traffic will take short-cuts through residential areas in cases where the safe route is longer than a certain value. In this application, the detour factor is used to search for routes that may be used by traffic between two urban centres. The detour factor is set on 1.6 times the celestial latitude. In case no route is found with a length that is less than 1.6 times the celestial latitude, a connection is not assigned to the network. The implications of this assumption are further discussed in section 8.5.

For the determination of the routes for a connection (route choice), the fastest route and the optimal route are defined. The optimal route is the route that leads as much as possible over the same or higher road categories as the theoretical category. The fastest route is the route with the shortest travel time. For the calculation of the travel time for a certain route, the maximum speed is applied. Thus, delays at intersections are not taken into account. When the optimal route is less than 5% slower than the fastest route, the optimal route is used for the connection. The implications of this assumption are further discussed in section 8.5.

An optimal route may lead over lower road categories before reaching the high category (e.g. the route to the highway). It is allowed to travel a maximum amount of kilometres over a lower order category. For this application, the following values are used:

- before reaching road category B: maximum 5 km over road category C
- before reaching road category BB: maximum 10 km over road category B and/or C
- before reaching road category AA: maximum 15 km over road categories BB, B and/or C

An example: before reaching road category BB, one has to drive over road category C and B. When driving 3 kilometres over road category C, the maximum amount of kilometres left to drive over road category B is 7 kilometres. The values are adopted from another research project (Pols, 2003) and based on interviews with employees from an area in the South of The Netherlands. The implications of this assumption are further discussed in section 8.5.

In some cases, roads might be used by more than one connection. In that case, this connection will become more important and therefore needs to be of a higher road category. In the case study area the average traffic volume of these connections together is compared to the maximum capacity of the road segment. When the necessary capacity exceeds the maximum capacity of the road segment, the connections are put together into a new and higher connection. Table 8.7 shows the maximum capacity of the road categories, which is based on the Handbook Road Design (CROW, 2002) and the ASVV (CROW, 2004).

Road Category	**Average traffic volume per connection type (vehicles/day)**	**Maximum capacity per road category (vehicles/day)**
AAA	40,001	No maximum
AA	25,000	40,000
BB	14,000	20,000
B	6,000	10,000
C	4,000	5,999

Table 8.7 Maximum capacity of each road category.

An example: in the theoretical network two connections with a C road category and one with a BB road category use the same road. This results in 2 x 4,000 + 14,000 = 22,000 vehicles movements. From Table 8.7 it can be seen that the redefined theoretically required road category should then be AA and not BB because the necessary capacity (22,000 vehicles) exceeds the maximum capacity of 20,000 vehicles.

Figure 8.4 shows the results of assigning the theoretical connections to the existing road network.

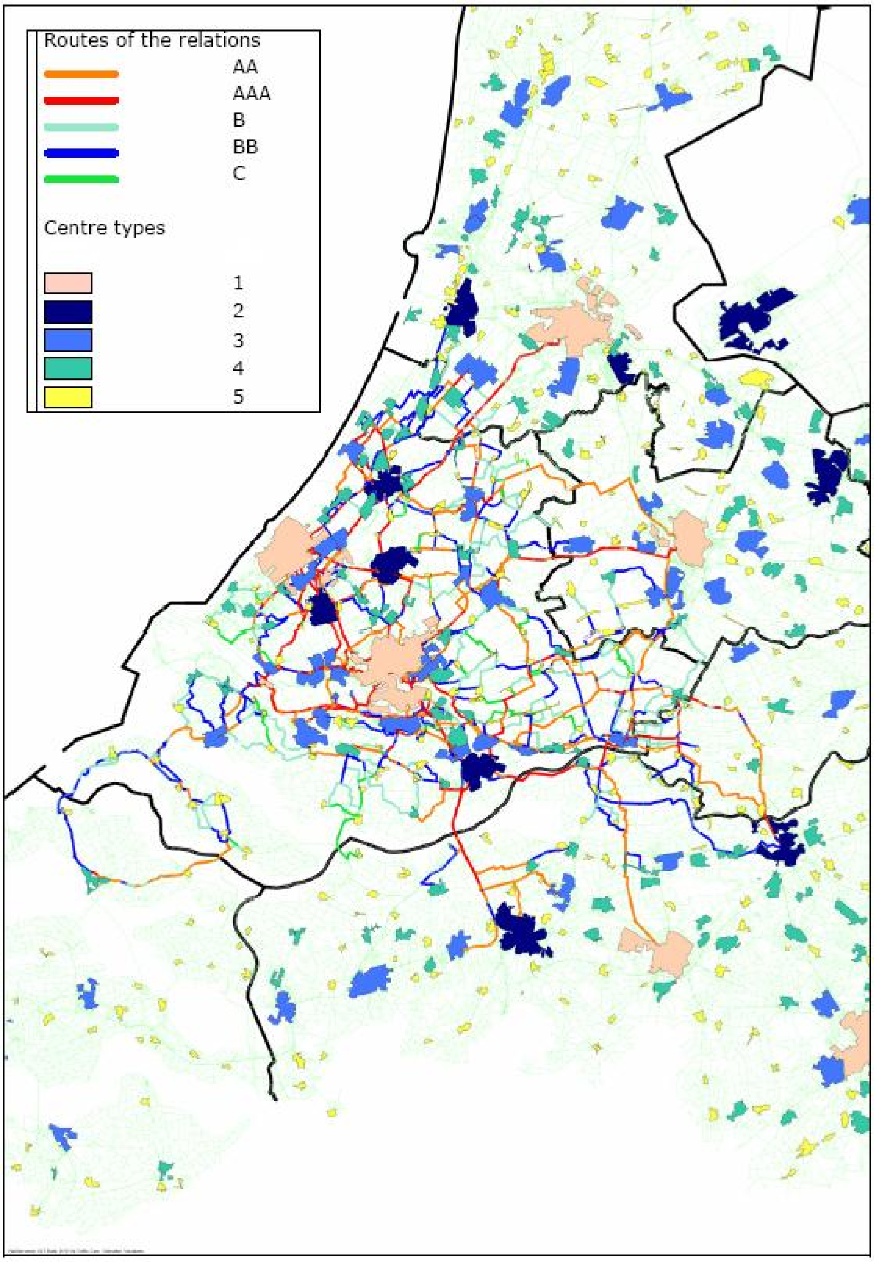


Figure 8.4 Theoretically required road categories.

Ad 4. Calculation of the SPI: Percentage of road length of appropriate category

To calculate the road network SPI, the current and theoretical networks are compared to each other. The road network SPI is the percentage of appropriate current road category (AAA-C) length per theoretical road category (AAA-C) length. Appropriate in this case is only considered from a safety point of view and means that the current road category is equal to or higher than the theoretical road category. Figure 8.5 shows for each theoretical road category the distribution of the current road categories in the case study area.

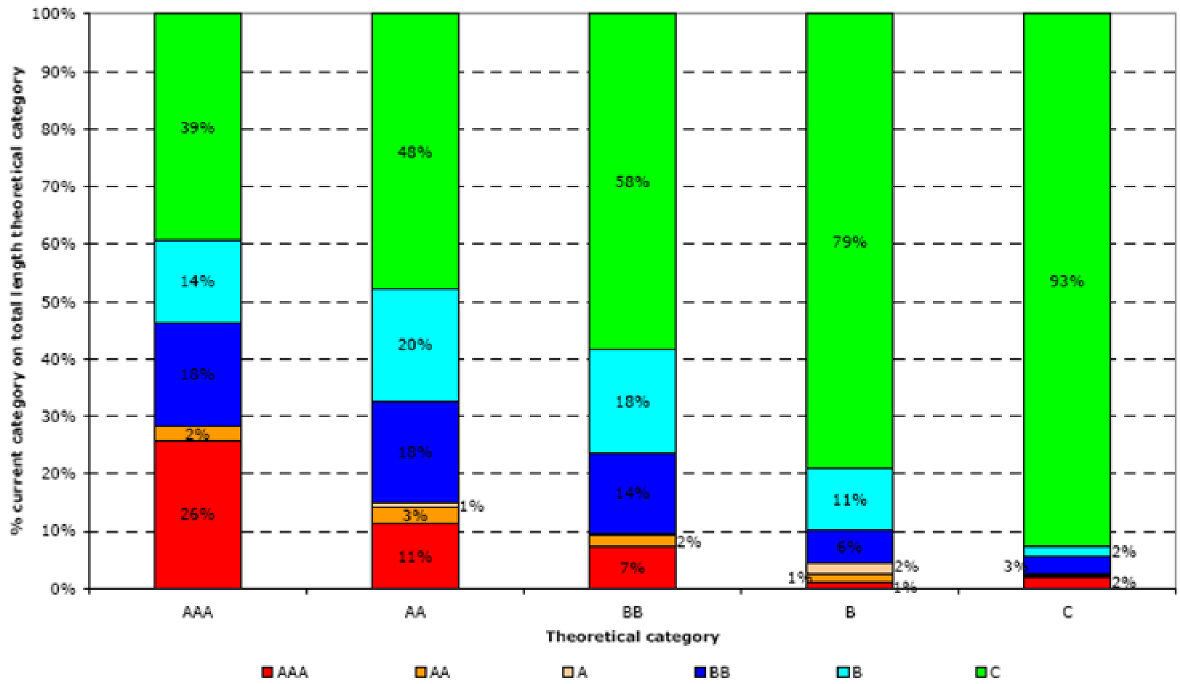


Figure 8.5 Percentage of similarity per category.

8.4.2 Road design SPI

The road design SPI is based on the Road Protection Score (RPS) of EuroRAP. This RPS needs to be converted into a score for each road category. One needs to determine the road length (km) per road category. The road design SPI is the distribution of stars per road category, see Table 8.8.

Road category	**Road Length (km)**	**Share of road length with:**			
		4 stars	**3 stars**	**2 stars**	**1 star**
AAA					
AA					
A					
BB					
B					
C					

Table 8.8 Distribution of stars per road category (table format).

Section 8.4.2.1 describes some case studies of the use of EuroRAP's RPS, section 8.4.2.2 describes the calculation of the road design SPI for the case study area using the RPS scores from EuroRAP.

8.4.2.1 Use of EuroRAP's RPS

Road inspection programmes based on EuroRAP's RPS methodology are now rolling out across Europe. Programmes are well developed in Sweden and Germany. In other countries like Great Britain, Iceland, Ireland, Finland, Netherlands, Spain and Switzerland, smaller (pilot) inspections are being executed and other countries will follow soon. This section gives a short overview of the experiences and results obtained in these countries (EuroRAP, 2006).

Sweden was the first European country to start a RPS programme and has inspected more than 10,000 km road so far. For this study, they used a specially equipped vehicle and software developed by the Swedish consultants SWECO. Overall less than 1% of the national road network received 1 star, 48% 2 stars, 29% 3 stars and 18% a 4 star rating. SWECO also was involved in the Icelandic and Finnish RPS programmes. The Icelandic Road Protection Score (RPS) programme began in 2005 and pilot road inspections covering 173 km were conducted on main roads around Reykjavik in 2006. Most routes were awarded a total score of 3 stars. In Finland initial inspections covered 150 km of the main road network in and around Helsinki.

In Germany, the motoring club ADAC began an inspection programme using EuroRAP's RPS methodology in 2004. They used a specially equipped inspection vehicle to assess road characteristics. In this first phase they inspected 6000 km of main motorways, along with 2000 km of randomly selected national and federal state roads. 70% of all motorways inspected received a rating of 4 stars. National and federal state roads provide lower protection and need improvements for run-off accidents and collisions at junctions. Of these routes, 10% achieved 4 stars, 53% 3 stars and 37% just 2 stars. In the second phase of the programme, 1200 km of motorway in Bavaria and Rheinland-Pfalz were assessed in a study conducted by ADAC and the University of Karlsruhe. The third phase, conducted in 2005 and 2006, included inspections of 3000 km of -mainly single- carriageway roads, within the federal states of Bavaria and Rheinland-Pfalz. Some of these inspections have been conducted with the new extended RPS developed by ADAC, incorporating the likelihood of accidents, and video system ARGUS.

The first full UK road inspections programme started in 2006. The inspections are being carried out by the German motoring organisation ADAC, supported by Swedish consultants SWECO, providing RPS for 7,000 km of major inter-urban roads in England, Scotland, Wales and Northern Ireland. In Ireland, pilot work on RPS began in 2006 and will be completed by the end of the year 2007. The ADAC road inspection team and vehicle were also used in Switzerland.

In the Netherlands, the motoring club ANWB started their RPS program with the inspection of the secondary road network of the province of South-Holland in 2005. A total road length of 751 km has been scored during drive-through inspections conducted by the Dutch consultant Mobycon. The inspected road network received an average score of 3 stars.

An extensive RPS programme of 7,000 km of the Spanish network was conducted in 2005 and 2006. Results showed that 54% of the motorway network was awarded less than 4 stars (47% 3 stars, 6% 2 stars and 1% 1 star).

8.4.2.2 Calculation of the road design SPI for the case study area

To obtain an SPI score, the EuroRAP Road Protection Score (RPS) needs to be converted to a score for each road category. The RPS scores for the case study area were taken from Mobycon (2006). For each road category, the distribution of the total road length over stars is determined. The results for the case study area are shown in Table 8.9.

It has to be noted that the Road design SPI could only be calculated for provincial roads, since EuroRAP RPS scores in the Netherlands are only available for these roads so far. No road design SPI is calculated for road category AAA since these are all national roads.

Road category	Road Length (km)	Share of road length with:			
		4 stars	3 stars	2 stars	1 star
AA	16 km	37%	44%	19%	0 %
A	7 km	14%	57%	29%	0 %
BB	203 km	20%	65%	15%	0 %
B	365 km	9%	61%	30%	0 %
C	108 km	38%	50%	12%	0 %

Table 8.9 Road design SPI for the case study area.

For all road categories, the majority of the roads received 3 stars. Road categories AA and C scored best and category B received the worst score.

8.5 Evaluation

This section evaluates the application of the road SPIs to the case study area. A general limitation that applies to both SPIs is that only rural roads and motorways are considered. The SPIs thus assess the safety for only part of the network. The network design SPI is further evaluated in section 8.5.1 and the road design SPI in section 8.5.2. Finally, section 8.5.3 discusses the representativeness of the case for the Netherlands in general.

8.5.1 Evaluation of the road network SPI

The four steps mentioned in Section 8.4.1 are evaluated individually.

Collecting and processing of data

- The data that is needed for the calculation of the network design SPI is available in MultiNet. From section 8.3 it can be concluded that for most countries, a large percentage of the rural and national roads are included in MultiNet. Thus, MultiNet and possibly also other geographical databases seem a good data source to calculate the network design SPI.
- The classification of the centre types for the case study area was determined by the number of inhabitants and the surface of industrial areas. It is to be questioned if the classification of centre types needs to be done using boundaries like the number of inhabitants. Other travel motives (e.g. an airport) which also attract and produce traffic, are now not taken into account. If there should be boundaries for the centre type classification, it needs to be investigated what these boundaries should be.
- Estimation of SafetyNet road categories based on road characteristics from MultiNet appears to be possible. It is not know yet if it is also possible for other countries. Only some minor problems occurred:
    - o In MultiNet the slip roads of motorways are categorised as lower type roads. This did not cause problems because slip roads only form a small percentage of the total network and part of the route may lead over lower categorised roads.
    - o Category 4 roads in MultiNet are partly category B roads and partly category C roads. Since in MultiNet the maximum speed groups 30 and 60 are not listed (yet) it is not possible to distribute these roads on the basis of this criterion. The distribution is based on the presence of a route number, but important roads of local governments could be B roads without a route

number and will be categorised as C roads. Furthermore, it is not certain that this extra criterion is appropriate for other countries as well.

Determination of the list of connections that needs to be assessed

- Drawbacks/limitations of centre type method
  - Natural barriers like mountains or administrative borders are not taken into account, while these barriers influence the number of trips between two centres (this did not occur in the present case study). A possible solution for this problem is to drop or devalue the type of connection if a barrier has to be crossed.
  - The distance between two centres of equal size can be very small, as a consequence of which very few or no connections with nearby centres exist, while those connections do exist in reality. This occurred in the case study area because the distance between two cities was very small. A possible solution is to define a set radius for the search area (one radius for each type of connection).
  - The distance between two centres of equal size can be very large, as a consequence of which connections are assumed that do not exist in reality. This problem will mainly occur in remote areas with large distances between big cities. This problem can also be solved using a set radius.
  - A smaller centre that is asymmetrically positioned between two larger centres is assumed to have equal connections with both centres, whereas in reality the connection with the nearest centre will be stronger. A possible solution is to lower the type of connection (AAA to C) with distance.
- At the moment it is not possible yet to link several countries in one database, therefore, cross-border connections cannot be assessed. In this case study this was not a problem, but problems might occur in case the study area is close to a border.

Comparison of the theoretically required road categories with the actual road categories

- Each connection is assigned to one route, whilst in reality, the traffic may be distributed over two or more routes.
- Several times a connection could not be assigned because the detour was too long as a result of geographical barriers. These connections have been classified in an extra step in which the detour factor has not been taken into account as a restriction.
- In some cases, connections were assigned to lower category roads parallel to safe routes. Maybe the 5% criterion should be adapted. This criterion was selected rather arbitrarily.
- Also the number of kilometres that may be travelled over lower category roads is selected rather arbitrarily. These values also need some further research. The sensitivity of the results for these values has to be examined and it has to be decided what values are acceptable from a policy point of view.

Calculation of the actual SPI

- It was possible to calculate the scores of the network SPI in an automated manner on the basis of the previously described steps
- The score of the road network of the example case study area is mediocre. For most road categories only a limited percentage of the roads was of a sufficiently high category. The theoretical method was also (manually) applied to another Dutch province and that province scored better (Dijkstra, 2003a). There are several possible explanations for this:

- The quality of the road network in the case study area used in this study is lower. In some cases a high category road is expected between two centres whilst in reality only a low category road is present.
- The inclusion of industry in the determination of type of urban area resulted in centres of higher classes as a result of which the theoretically required road network consists of higher road categories.
- In cases where several connections are assigned to the same road, the theoretically required road category may be higher than needed for the individual connections (see Table 8.7). In the case study described in Dijkstra (2003a), the theoretically required road network was not adapted in case more connections used one road. The network requirements are thus higher in the case study area used in the present case study area.
- Also the assumptions regarding the difference between the safest and fastest route (5% criterion) and the amount of kilometres that may be travelled over lower category roads affect the calculated SPI. As mentioned before, these assumptions need further investigation.

It would be interesting to evaluate the results of the case in more detail and to analyse the extent to which possible explanations account for the results. However, this is not the purpose of this manual.

8.5.2 Evaluation of the road design SPI

It appeared to be possible to calculate the road design SPI in an automated manner, with the data discussed in section 8.3.1. Unfortunately, the Road Protection Scores (RPS) were only available for provincial roads and not for national roads as a result of which only part of the road network could be evaluated. The roads that were evaluated scored reasonably well, the majority of the road network received three stars. The major limitation of the road design SPI is that the precise methodology by which EuroRAP calculates RPS scores on the basis of values on various road characteristics is not published.

8.5.3 Representativeness

This Chapter described the application of the network design and road design SPIs for the case study area. For this area, all rural and national roads were included in the study, i.e. no sampling process was used. The results are not representative for the Netherlands in general, since the roads that were included in the analysis are not a random sample of all Dutch roads. In this case, this is no problem, since the purpose of the application was to illustrate the method, to analyse whether the method is applicable and can be computerized and to detect bottlenecks in the application.

For a representative sample, we advise to obtain a stratified sample with a sample size of 384 connections (see Chapter 2). The connections need to be selected at random in the list of each theoretical road category (AAA-C) and proportionally to the number of kilometres within each theoretically required road category.

A complication that results from the automation process is that connections are assigned to the road network. This has some theoretical and practical implications with regard to the sampling. The practical implication is that the randomly selected connections cannot be assessed without taking account other connections that may (partially) use the same route. This is necessary since a road may have to be of a higher category than needed for the selected connection (in case that more than one connections use the same road). The theoretical implications are that (1) the same road is assessed twice in case selected connections use the same road and (2) a selected connection may consist of several theoretically required road types. The first theoretical implication can be dealt with by correcting the sample for these situations. Regarding the second implication, the shares of different road categories can be calculated for each selected connection. An example: a selected connection of type BB is assigned to the network and 20% of the route is also used by an AAA connection. The other 80% of the route is only used by the selected BB connection. In the sample, the selected connection counts for 80% as a BB connection and for 20% of an AAA connection.

Another theoretical option to solve the sampling issues mentioned above, is to take roads instead of connections as sample unit. However, this is practically not possible, since in that case a list of theoretically required road categories is needed. This implies that all connections have to be assigned to the network, before a sample can be selected.

8.6 Conclusions and recommendations

This chapter discusses the application of the network design SPI and the road design SPI to the case study area. Some additional individual examples concerning data collection are also briefly presented. As mentioned in the introduction, the purpose of the application is twofold. First, it illustrates the rather theoretical concept of the SPIs and second, it aimed to provide more insight into the practical applicability of the method.

It appeared to be possible to automate the process of the calculation of the network SPI. Moreover, most of the data necessary for the calculation of the network SPI was available in a geographical database. The process of calculating the network SPI is however quite complex and a large amount of data is needed (also in case of a sample). The SPI should be applied in a number of countries in order to investigate whether the SPI is generally applicable.

For the calculation of the road design SPI, EuroRAP Road Protection Scores (RPS) for individual road segments or routes are needed. Since EuroRAP is being applied in a growing number of countries, we expect that the road design SPI can be calculated in other countries as well.

The results from the case study seem reasonable. However, to get more insight into the meaning of the estimated values more detailed evaluation of the results is recommended, in consultation with the road authority. Moreover, it would be interesting to compare the values for different countries.

The previous section discussed some issues that need further research. These issues are summarized below:

- The quantification of the method is done quite arbitrarily. Therefore, the classification of urban areas and connections needed between them need more attention. It is to be questioned whether the classification of centre types needs to be done using boundaries like the number of inhabitants. Other travel motives (e.g. an airport) which also attracts and produces traffic, are now not taken into account. If there should be boundaries for the classification of centre types and connections types, it needs to be investigated what these boundaries should be.
- Translating the different country road classification systems into the road categories defined in SafetyNet may give some problems. Therefore a certain region of the country or a certain part of the road network could be investigated but this will then not give a representative result for the whole country.
- The centre type method has some limitations in case of natural barriers or administrative borders, or very small or large distances between centres of equal size. It would be interesting to investigate the effects of these limitations and to refine the method.
- Also the assignment of the theoretically required connections to the actual road network needs some further refinement, since some of the discussed criteria are defined rather arbitrary.

We recommend applying the method in various countries to further investigate these issues and refine the method.

In conclusion, the case study proved that it is possible to apply the theoretical concept in practice. The SPIs provide insight into the safety quality of the network as a whole and of individual roads. Therefore, application of these SPIs to assess the safety of the network is recommended. Thereby, three remarks should be noted: First, so far, the SPIs only assess the safety performance of rural roads and motorways. Extension and adaptation of the method in the future in order to include urban roads as well is recommended. Secondly, the precise methodology by which EuroRAP calculates RPS scores on the basis of values on various road characteristics is not yet published. For the transparency of the process it is however important that it is. Finally, we discussed some issues that need further research. It is recommended to investigate these issues and to refine the method.

8.7 References

1. AVV (2005) Wegkenmerken+, Een landelijk en uniform basis bestand met wegkenmerken. Ministerie van Verkeer en Waterstaat, Rijkswaterstaat, Adviestdienst Verkeer en Vervoer (AVV); Ministry of Transport, Public Works and Water Management, Directorate-General for Public Works and Water Management AVV Transport Research Centre, Rotterdam, The Netherlands. [in Dutch]
2. CROW (2002) Handboek Wegontwerp. C.R.O.W. kenniscentrum voor verkeer, vervoer en infrastructuur. [in Dutch]
3. CROW (2004) ASVV 2004: aanbevelingen voor verkeersvoorzieningen binnen de bebouwde kom, C.R.O.W kenniscentrum voor verkeer, vervoer en infrastructuur. [in Dutch]
4. Dijkstra, A. (2003a). Kwaliteitsaspecten van duurzaam-veilige weginfrastructuur. R-2003-10. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam. [in Dutch]
5. Dijkstra, A. (2003b) Testing the safety level of a road network; Contribution to the XXIIth PIARC World Road Congress, 19–25 October 2003, Durban, South Africa. Report no. D-2003-15, SWOV, Leidschendam.
6. EuroRAP (2006). Getting organised to make roads safe: second Pan-European progress report. EuroRAP AISBL. Basingstoke, Hampshire, UK.
7. FGSV (1988). Richtlinien für die Anlage von Straßen RAS. Teil: Leitfaden für die funktionelle Gliederung des Straßennetzes RAS-N. Forschungsgesellschaft für Straßen- und Verkehrswesen FGSV, Köln.
8. Hakkert, A.S., Gitelman, V. and Vis, M.A. (Eds.) (2007) *Road Safety Performance Indicators: Theory*. Deliverable D3.6 of the EU FP6 project SafetyNet.
9. Lynam, D, Sutch, T. Broughton, J, and Lawson, S.D. (2003) European Road Assessment Programme, Pilot Phase Technical Report, AA Foundation for Road Safety Research, Farnborough, UK.
10. Lynam, D, Hummel, T. Barker, J. and Lawson, S.D. (2004) European road assessment programme, EuroRAP ! (2003) Technical report, AA Foundation for Road Safety Research, Farnborough, UK.
11. Mobicon (2006). Slides from a presentation. Not published.
12. Pols, J. (2003). 0+ BOSE belicht. SRE Samenwerkingsverband regio Eindhoven. [In Dutch].

13. Tele Atlas (2007a). MultiNet country profiles. ta_ct012525.pdf; http://www.teleatlas.com/stellent/groups/public/documents/content/ta_ct012525.pdf. Tele Atlas NV, Tele Atlas North America, Inc.
14. Tele Atlas (2007b). Brochure ‘We’re mapping the world’. ta_ct012494.pdf; http://www.teleatlas.com/stellent/groups/public/documents/content/ta_ct012494.pdf. Tele Atlas NV, Tele Atlas North America, Inc.

9 Detailed manual for SPIs on trauma management

GITELMAN[A], V., AUERBACH[B], K., AVITZOUR[A], M.

[A]TECHNION, [B]BASt

9.1 Introduction

Trauma management (TM) or post-crash trauma care refers to the system, which is responsible for the medical treatment of injuries resulting from road crashes. It typically covers the initial medical treatment provided by Emergency Medical Services (EMS), at the scene of the crash and during the transportation to a permanent medical facility, and further medical treatment provided by permanent medical facilities (hospitals, trauma centres).

There is a consensus in the professional literature that the appropriate management of road casualties following the crash is a critical determinant of both the chance of survival and, on survival, the quality of life. Conversely, improper functioning of the post-crash care system leads to more fatalities and severe injuries, which could be avoided.

The ETSC report "Reducing the severity of road injuries through post-impact care" (1999) highlighted evidence-based actions for the organisation of optimal trauma care in the EU. The 2003 European action programme (CEC, 2003) stated that several thousands of lives could be saved in the EU by improving the response times of the emergency services and other elements of post impact care in the event of road traffic crashes. A recent World Report on Road Traffic Injury Prevention (Peden et al, 2004) indicated the importance of improving medical care delivered after crashes.

The ETSC report "Transport Safety Performance Indicators" (2001) highlighted a need for the development of trauma management related indicators. Such indicators would be useful:

- in estimating the quality of emergency medical services and hospital care of road crash victims;

- in recognizing weak chains and current needs of the system for decision-making purposes;

- in following-up progress through time;

- in providing evidences of efficiency of different policy measures.

In general, the international comparisons of the TM systems should be performed with caution due to a variety of definitions, legislations and systems, which are available for both the emergency and in-hospital trauma care, in different European countries. However, based on the best practice recommendations in the field of post crash care (e.g. ETSC, 1999), a number of features can be named which are definitely associated with better performance of the TM system. They are:

- shorter response time by EMS;

- higher level of the EMS staff;

- standardisation of the EMS vehicles;

- adequate hospital trauma care.

Besides, establishing a strategic approach to national trauma care is considered to be a prerequisite for high-quality hospital trauma care in the country.

Based on the above considerations, a set of TM SPIs was developed but, further, reduced accounting for the limitations of data available in different countries (Hakkert et al, 2007).

Finally, we introduced a *Minimum set of Trauma Management SPIs* which are necessary for an initial characteristic of the system's performance (see Section 9.4).

9.2 Measuring post-crash trauma care – existing practices

9.2.1 General

Recognising the importance of the post-care trauma care indicators for road safety programs and for road safety policy-making in general, ETSC (2001) could provide only a few examples of such indicators in use. Typically, national road safety programs do not include an explicit target on improving rescue services and medical care of crash injuries. An exception can be seen in the Swedish national road safety programme for 1995-2000 which included an objective of shorter response times of emergency medical services.

9.2.2 Information provided by national experts

Based on the information collected from national experts (Hakkert et al, 2007), it was concluded that EU countries generally have EMS norms/ regulations which, however, differ among the countries and, sometimes, between areas within a country (e.g. in federal states). The norms regarding EMS response time exist, in a certain form, in half of the countries. Compliance with these norms is assessed from time to time. Recent estimates of EMS response times are available in half of the countries.

EMS databases were stated as existing in many countries but their data are not easily accessible and are typically not linked to other road crash related databases, e.g. police crash files or hospital databases.

In the majority of countries, the composition of EMS teams, types of medical treatment provided at the scene, and the type of medical facility to transport the patient are regulated by internal rules. However, the quality of initial treatment provided or the extent of following these rules usually is not estimated.

Trauma Registry databases exist in some countries, but these typically cover selected hospitals only and/ or specially defined types of injury. The major problem is that available medical databases are generally not linked to the road safety research and management activities. As stated by previous research, mapping the trauma data and integrating them with the road safety data would lead to significantly improved decision making in emergency medical treatment of road crash casualties.

Most of the data provided by the countries (see Hakkert et al, 2007) are suitable for the characteristic of the *availability* of trauma care services, where quantitative data on actual *performance* of the EMS and further medical treatment, are typically lacking. The state of the data in not uniform among the EU countries: more detailed information on the TM is available for Germany, Czech Republic, Belgium, Hungary.

As data on the performance of post-crash trauma care in the country are usually not in use in the current decision-making practice, in some countries such data are lacking/ not readily available. This means that special efforts will need to be applied to provide the data requested for the calculation of TM SPIs.

Based on the information collected from national road safety experts, it can be stated that presently no European country applies trauma management indicators in the decision-making processes on road safety.

9.2.3 Trauma care indicators in medical practice

Different forms of trauma care indicators can be found in medical practice. A literature study was undertaken to review trauma care performance indicators, which were applied by empirical studies for the analysis of systems /forms of post-crash care (see Hakkert et al, 2007). A brief summary of the results follows below.

Studies analysing the relationship between the performance of the trauma care system and road crash outcomes, are not frequent. There is evidence concerning the effects of improved EMS care on the frequency of fatalities and/ or the state of severe injuries. The improved EMS care is measured in terms of lower EMS response time, higher rate of qualified emergency staff and/ or higher rate of better equipped emergency vehicles which arrive at the scene of accidents. The state of severe injuries is estimated upon discharge or some period later, and is measured by means of standard protocols, which rank the capabilities of a person to carry out basic life functions. The outcome indicator usually has a form of percentage of those patients who satisfy the protocol's demands.

In general, trauma care performance is characterized by shares/rates of different forms of treatment, with the emphasis on higher levels of treatment and on percentages of correspondence to the demands of medical protocols (for the care to be supplied). The values of EMS response time and the time values of treatment at the hospital are frequently in use. The inputs of the medical systems (EMS and hospitals/trauma centres) are typically considered together with the outcomes, which are the state of the patients treated. The mortality or survival rates (i.e. the percentage of those who survived or died out of the sample considered) and the length of stay in hospital/ intensive care unit are frequently used for comparison of different forms of initial treatment.

Indicators, which are frequently applied to the characteristics of medical treatment at permanent medical facilities, are the length of stay in hospital, the length of stay in intensive care unit, times of waiting for certain treatments, mortality rate and the quality of life of the former patients. For a comparison of the level of in-patient treatment between hospitals/ countries, an indicator of mortality rate due to poly-trauma is also applied.

Comparing both EMS and hospital treatments, a correction for injury severity is necessary as the effects of treatments can reasonably be compared only for groups of patients with similar severity levels. Besides, age and gender differences among the compared groups should be controlled for.

Some national-scope studies which sought to establish a connection between the improvements in medical care and reductions in traffic-related fatalities, applied proxies of medical cares such as: the average length of in-patient stay in the hospital, the per-capita level of National Health Service staff, the number of people per capita waiting for hospital treatment, infant mortality rates, physicians per capita, average acute care days spent in the hospital. However, these indices seem too general and not suitable for characterisation of the performance of the trauma care system.

Based on the literature considered, a summary of the evaluation parameters for the trauma care system may be as follows:

At the EMS Level -

- Type of training that EMS teams receive: Basic Life Support versus Advanced Life Support;
- Type of evacuation to trauma centre: self, regular ambulance, mobile intensive care unit, helicopter;
- Time values: arrival at scene, treatment in the field, arrival for definitive treatment in hospital;
- Type of field treatment provided;

- Treatment implementation according to protocols, to the extent that protocols exist.

At the Hospital Level -

- Level of coverage: to what extent do critical patients arrive at trauma centres and not at hospitals of other levels?
- Severity of injury according to e.g. Injury Severity Score and according to part of body injured (Barel Matrix) with emphasis on head, chest and stomach injuries;
- Performance of specific surgical procedures and evaluation of outcomes, comparisons of treatment in specific procedures;
- Speed of treatment in the hospital, speed of arrival to Emergency Rooms, extent of work according to protocols.

For outcomes –

- Death rates,
- Hospitalisation in intensive care units,
- Total length of hospitalisation.

The trauma care indicators estimated in medical studies are typically based on data from medical databases such as hospital files, trauma registry, or from national mortality files.

However, one should remember that the indicators mentioned in this Section are, mostly, of a research nature and that they are not in systematic use in the decision-making practice.

9.3 Detailed description of available best practices

In this section we present two countries which have comprehensive data on trauma care: Germany – on Emergency Medical Services and Israel – on in-hospital trauma care (Trauma Registry). In each country, the data are collected and the performance indicators are estimated systematically, which could serve as a reasonable background for the application of TM SPIs.

9.3.1 Germany: management of Emergency Medical Services

In Germany the Emergency Medical Services (EMS) is a public task, which is, according to the Basic Constitutional Law of the Republic of Germany, the responsibility of the 16 federal states (Bundesländes). Every federal state has its own quality management involving different forms of quality assessment and performance indicators.

Nevertheless, there is a general agreement that one indicator is the most important to assure a high quality standard of the EMS which is <u>response time</u>. Although the response time is defined differently by almost all 16 federal EMS-laws, it is always defined as an indicator to guarantee adequate medical care. On the basis of this definition, the number and location of EMS-stations have to fulfil the requirements and guarantee meeting the limits of response times. The following claims might be asserted for the whole country, though there are discrepancies in different states:

- The response time ought not to be longer than 15 minutes, but allowance for special cases is possible (e.g. areas with a very low density of population).
- The EMS ought to be organised in such a way that the limits of the response time are met in 95% of all sorties.
- The case that an EMS-team is needed twice ought not to occur more often than every 15 days (twice a month).

Every four years, the Federal Highway Research Institute (BASt) conducts a study to analyse the quality of the EMS system in Germany (Schmiedel & Behrendt, 2002, 2007). This is the only study which is conducted at regular intervals and which examines the German public EMS system as a whole. The study is based on extrapolation. Data collection is conducted in about 60 regions of Germany during four 1-week data collection periods. The regions cover urban, suburban and rural areas and are representative for the structure and population of Germany. The data are provided by the EMS dispatching centres according to a standardised questionnaire. As the dispatching centres collect these data routinely and normally are equipped with data processing, data collection is rather unproblematic. The data include, amongst others:

- Date and point of time of emergency call, point of time of alarm being raised
- Type of response (response with or without physician or response in case of patient transport care)
- Cause of response (road accident, accident at work, other accident, internal emergency, other emergency, patient transport care)
- Place of emergency (home, hospital, medical practice, old people's home, other, road outside build-up areas, road inside build-up areas)
- Destination of transport (home, hospital, medical practice, old people's home, other destination, just medical treatment – no transport, no medical treatment – no transport)
- Point of time of arrival at place of emergency, start of transport, point of time of arrival at destination

Based on these data response time can be estimated. The study uses the following definition of response time: "The response time is at least the time period after the emergency call, that is available for decision-making, alarming the EMS, getting ready the staff and the emergency vehicle/helicopter and driving to the scene of emergency."

This definition does <u>not</u> include the time that is needed to get to the patient (at the scene itself), because the duration of this phase is often determined by circumstances that can not be controlled by the EMS.

The latest study was finalised in 2006. The real data of the study of the year 2004/05 contain more than 188,000 rides (number of EMS vehicles alarmed) and more than 157,000 responses (number of events) for the selected areas. The extrapolated data contain 12.1 million rides and 10.2 million responses in Germany, per year.

Concerning the response time, a time series can be presented (as the study is conducted at regular intervals with the same methodology). Table 9.1 shows the results of the last decade.

Period	Response time: percentage of cases below or equal to						Mean	**95%-percentile**
	5 min	**7 min**	**10 min**	**12 min**	**15 min**	**20 min**		
1994/95	39,7 %	62,0 %	82,1 %	88,9 %	94,5 %	98,2 %	7,3 min	15,4 min
1996/97	34,7 %	57,9 %	79,8 %	87,6 %	94,0 %	98,2 %	7,7 min	15,8 min
1998/99	34,2 %	57,5 %	79,0 %	86,7 %	93,6 %	98,2 %	7,8 min	15,9 min
2000/01	33,9 %	57,5 %	79,1 %	87,2 %	93,8 %	98,0 %	7,8 min	15,9 min
2004/05	**29,9 %**	**54,0 %**	**77,3 %**	**85,8 %**	**93,2 %**	**97,8 %**	**8,1 min**	**16,3 min**

Table 9.1. Estimates of the EMS response time in Germany. Source: Schmiedel & Behrendt (2007).

Similar time series can also be presented for different types of responses. Table 9.2 shows the results for the years 2000/01 and 2004/05.

Type of calls	Period	Response time: percentage of cases below or equal to						Mean, min	95%-percen-tile, min
		5 min	**7 min**	**10 min**	**12 min**	**15 min**	**20 min**		
Traffic accidents	2000/01	35,6	56,0	77,2	85,3	91,9	97,3	8,0	16,7
	2004/05	**28,8**	**48,9**	**72,0**	**81,5**	**89,9**	**95,8**	**8,7**	**18,8**
Accidents at work	2000/01	20,1	40,4	68,0	79,1	89,2	96,3	9,6	18,2
	2004/05	21,1	43,8	67,5	81,3	89,8	95,7	9,7	18,9
Other accidents	2000/01	26,0	45,3	65,4	75,2	85,0	93,2	10,0	22,2
	2004/05	20,8	40,5	62,5	72,0	82,4	91,0	10,8	24,8
Internal emer-gencies	2000/01	28,8	50,2	72,6	81,5	89,5	95,7	9,0	19,2
	2004/05	24,3	45,8	68,7	77,8	86,7	93,7	10,0	22,0
Other emer-gencies	2000/01	29,5	49,4	69,6	78,5	87,0	93,6	9,9	22,6
	2004/05	24,3	42,3	61,5	69,4	77,1	83,9	14,9	52,0
Patient Transport Care	2000/01	17,3	25,3	38,5	46,6	57,1	69,3	20,2	63,0
	2004/05	12,4	19,6	31,8	39,1	48,9	60,8	25,6	87,5

Table 9.2. Estimates of the EMS response time (average of all EMS vehicles) for different types of responses. Source: Schmiedel & Behrendt (2007).

The time series enables recognition of negative (and positive) changes in response time and, therefore, can be used to manage unfavourable processes in the EMS.

Based on the study data, further indicators can be examined, e.g. the share of road accidents out of all EMS responses. Table 9.3 sums up the results of the year 2004/05 by type of calls and road areas. It can be seen that, depending on area, traffic accidents present 2.2-3.0% of all EMS responses.

Present data (2004/05) can be compared with the results of former studies. These comparisons indicate that the share of road accidents in the whole work of the EMS is continuously reduced.

Type of calls	**Area**			Total
	Urban	Suburban	Rural	
Road accidents	**115,240** **2.3 %**	**115,748** **3.0 %**	**29,152** **2.2 %**	**260,140** **2.6 %**
Accidents at work	29,277 0.6 %	16,896 0.4 %	7,907 0.6 %	54,080 0.5 %
Other accidents	432,635 8.7 %	299,668 7.7 %	66,766 5.0 %	799,069 7.9 %
Internal emergencies	1,250,061 25.2 %	785,156 20.3 %	237,956 17.9 %	2,273,173 22.4 %
Other emergencies	1,375,172 27.7 %	442,906 11.4 %	242,801 18.3 %	2,060,879 20.3 %
Patient Transport Care	1,757,767 35.4 %	2,215,601 57.2 %	744,049 56.0 %	4,717,417 46.4 %
Total	4,960,152 100 %	3,875,975 100 %	1,328,631 100 %	10,164,758 100.0 %
	4,960,152 48.8 %	3,875,975 38,1 %	1,328,631 13,1 %	10,164,758 100.0 %

Table 9.3. Split of the EMS responses by type of calls and area in 2004/05.

Indicators concerning the EMS infrastructure are available as well. These data are collected in longer time intervals as changes per year are rather small. An example of infrastructure data is the number of inhabitants per EMS-vehicles. Table 9.4 shows the results for the years 1994 and 2000. The results indicate that between the years 1994 and 2000 the number of vehicles in total and especially the number of MICU has increased in relation to the number of population.

Index	1994	2000
Inhabitants per EMS-vehicle (all vehicles)	10,578	9,672
Inhabitants per EMS vehicle type C (MICU)	25,599	22,070
Inhabitants per EMS vehicle type B (BLSU)	27,069	27,783

Table 9.4. The number of inhabitants per EMS-vehicles

Figure 9.1 illustrates the shares of different vehicle types out of the total EMS vehicles. It can be seen that the share of MICU (type C vehicles) has a long-term increasing trend.

(To note, the total figure of the EMS transportation units for the whole country is typically unavailable. The study applies extrapolations to estimate the shares of different vehicle types. As explained above, the methodological rules of the study are well-based and systematic, which enables the provision of robust values of the shares.)

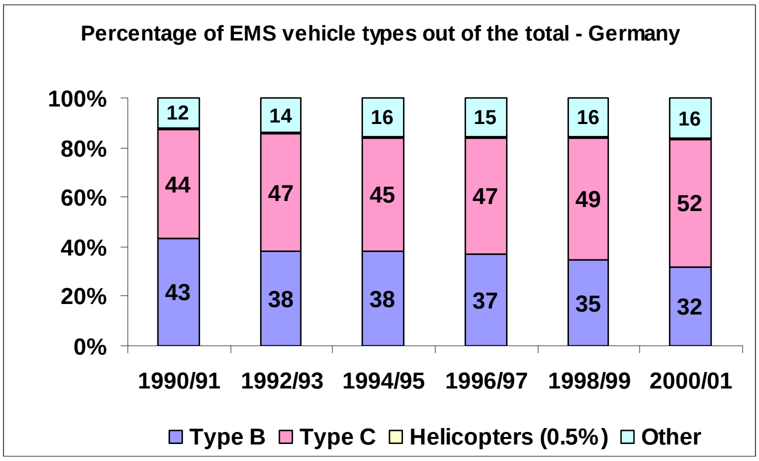


Figure 9.1. Shares of different vehicle types out of the total EMS vehicles. (Based on Schmiedel & Behrendt, 2002).

Although the long-term study presented allows many questions concerning the availability and the quality of the EMS in Germany to be answered, the estimated indices are restricted to essential time criteria (e.g. response time) and “objective” cause of the response (e.g. type of calls). The study cannot provide indicators on medical treatment applied and medical outcomes. The treatment and outcome data is much more difficult to collect as there are many differences in diagnosing and assessing injuries and diseases. In the future more efforts need to be spent in order to gain such invaluable information.

9.3.2 Israel: National Trauma Registry

Following the creation of a trauma system in the USA, it was decided in Israel to establish a national committee to study the issues of creating such a system in Israel. The Revach Committee was formed in the early 1990's with the main aims being (Shemer, Shapira, 1996):

1. To define criteria for choosing the hospitals that will become level 1 trauma centres;
2. To define standards for those centres, in terms of facilities, professional staffing, etc;
3. To recommend cost-effective working procedures and propose a system for quality control.

The Committee's final report was presented in 1992. In 1993, the Ministry of Health approved most of the recommendations. As victims of motor vehicle accidents (MVA) comprise about 25% of all hospitalized injured patients, the Ministry of Transport participated in financing the establishment of these trauma centres.

Based on the Revach Committee's report, 6 hospitals were recognized as level 1 trauma centres, with all the necessary facilities for treating multi-organ trauma, while 21 other hospitals were designated as regional trauma centres. The trauma system in Israel includes all emergency services, the ambulance system and the Israel Defence Forces. Besides, it includes a rehabilitation system, training medical and para-medical staff and investments in infrastructure including a Trauma Registry (TR).

The main aim of the TR is to maintain a database covering the whole trauma system thereby providing the ability to review and examine its components, and ongoing quality improvement (Shemer, Shapira, 1996).

The initial TR was established in 1997, under the auspices of the ICDC (Israel Center for Disease Control). The increasing interest in the quality and cost of trauma care led to the establishment of trauma registries along side trauma centres. Today, a hospital can not be a recognised trauma centre without a TR.

Continuous quality improvement of the trauma care system is achieved by collecting relevant data about the injury, the medical care received and the outcome; this is in addition to individual quality assessment. The data collected can be used as indicators for monitoring problematic issues. Such indicators should be sensitive enough to note deviation from normal care; specific enough to pinpoint those situations needing clarification; and efficient enough to justify the cost in terms of money and time. Most importantly these indicators should be valid, and based on accepted measurements and scores (Cryer et al, 2005).

The Israeli National Trauma Registry is maintained by the Israel Center for Trauma and Emergency Medicine Research of the Gertner Institute. The ongoing work of the TR includes:

- Gathering non-identifiable data from each of the trauma centres
- Quality improvement of the data collected
- Data analysis for each of the hospitals in order to assist in quality assurance for the treatment of trauma patients
- Performing research using the data collected
- Supplying necessary data needed to make health policy decisions both on a hospital and on a national level.

Complete cooperation is needed between the pre-hospital and in-hospital staff, in order to allow the TR to fulfil its mission. This function is usually performed by a trauma coordinator. There is an obligation of confidentiality between all parties.

As of 2006, the National TR contains data on hospitalized patients from 10 trauma centres: all six level I Trauma Centres and four regional trauma centres (the number may grow in the future). The TR includes data of all those patients that were injured and hospitalized as well as those that were injured and died in the emergency rooms (ER). The physicians in the trauma centres are responsible for identifying eligible cases. Data are collected by designated registrars. Over 200 data fields are included in the registry such as: demographic information on the patients, details of the injury including diagnoses, severity indicators, and details on the external cause of injury; treatment; hospital resource utilization; and outcome.

The data items of the TR, which are relevant to MVA, are as follows (Heruti et al, 1997):

1. Injured person's identification and demographic variables: age, sex, population group, country of birth, date of immigration, home address;
2. Circumstances of the MVA: timing (date, hour) and location of the event;
3. Road accident details: type of accident, type of vehicle, position in the vehicle, involvement of pedestrian, use of safety devices;
4. Pre-hospital treatment: vital signs, procedures performed on the scene (resuscitation, airway, etc.), GCS[7], type of transfer;
5. ER including transfer from another hospital: time of arrival, vital signs, GCS, consciousness status, procedures;
6. Hospitalization: diagnosis (up to 20 per patient), procedures, ISS[8], intensive care unit (ICU), other wards; and
7. Outcome: death, length of stay (LOS), functioning, rehabilitation.

At present, the TR collects approximately 20,000 records per year and the accumulated database from 1997-2002 contains over 100,000 records; about 25% of them are data on MVA. These data are used to identify populations at risk, potential risk factors and patterns of injury in the population as well as to make and disseminate recommendations for policy and action (Trauma in Israel 2000, 2002).

As effective injury prevention requires a multifaceted and multidisciplinary approach, the TR data summaries and estimated indices are provided to the National Council on Trauma, the Ministry of Health, the National Road Safety Authority, etc. The TR provides data to the Israeli Parliament for discussions and policy-making with regard to road safety issues.

Concerning the usability of the TR data, a recent Israeli study (Peleg, Aharonson-Daniel, 2004) compared the trends in the numbers of road accident injuries based on the TR versus the officially published accident statistics (based on the police reporting). In the study, significant discrepancies were identified between the official (police) numbers and the TR numbers of injured people who are hospitalized due to road accidents. The TR was found to be a more comprehensive source of data. For example, over the years 1998-2000, the TR numbers were higher that the police numbers despite the fact that the latter covers the whole country, while the TR includes only 8 hospitals (out of 24). Another important finding was that while the police figures showed a reduction in the number of injured, the TR data indicated an increase. One of the conclusions of the study was that road safety decision-making should be based on data gathered from multiple data sources.

Examples of post-crash trauma care indicators which can be estimated based on the TR data and compared over the years, are given below.

a. Distribution of the MVA injuries by severity, in 2003 versus 1998 (Table 9.5).
It can be seen that the share of severe injuries (ISS≥16) was similar in both years – over 18%[9].

---

[7] See Glossary - Sec.9.7.
[8] The same
[9] The difference is not statistically significant: p=0.51for χ^2-test; z<1.96 while testing the hypothesis on difference between two sites' proportions (see Sec.5 in the *Statistical Appendix*).

ISS	**1998**		**2003**	
	Number	**%**	Number	**%**
1-14	3693	**81.3**	4149	**80.7**
16+	837	**18.4**	974	**18.9**
Unknown	14	**0.3**	19	**0.4**
Total	4544	**100**	5142	**100**
Where:				
16-24	384	**8.5**	496	**9.6**
25+	453	**9.97**	478	**9.3**

Table 9.5. MVA injuries: arrivals to hospitals by severity, in 1998 and 2003

b. Distribution of the MVA severe injuries (with ISS≥16) by mode of arrival, destination of emergency room (ER) exits, outcome, length of stay in intensive care units (ICU), total length of stay (LOS) and medical procedures, in 2003 versus 1998 (Table 9.6).

It can be seen that about 60% of severe patients are brought to hospitals by MICU, 2-3% by helicopters, and 27-30% by other ambulances. However, the share of those who came by private cars increased in 2003 as opposed to 1998: 6.4% versus 3%, accordingly[10].

A third of the severe patients are treated by ICU, 25%-29% go though an operation; slightly over 2% die in the ER. In 2003 as opposed to 1998: a higher share of injuries were hospitalized: 35.5% versus 28% accordingly[11].

Comparing the outcome of hospitalizations of severe cases, it can be seen that in 2003 versus 1998 *a lower share of patients died* (about 12% versus 19% accordingly) but a higher share went to rehabilitation facilities (about 31% versus 23% accordingly)[12].

The share of cases treated in ICU and the mean length of stay in ICU were similar in 2003 as opposed to 1998[13]. The total length of stay for an average patient was shorter in 2003 versus 1998[14]; the share of patients treated by procedures was similar in both years[15].

[10] The difference is statistically significant: $p<0.001$ in χ^2-test.

[11] The difference is statistically significant: $p<0.01$ in χ^2-test.

[12] The differences are statistically significant: $\chi^2=17.059$, $p<0.001$ for mortality; $\chi^2=14.062$, $p<0.001$ for rehabilitation.

[13] The differences are not statistically significant: $p=0.729$ in χ^2-test for the number of cases treated in ICU; $|z|=1.52 < z_{\alpha/2}$ ($\alpha=0.05$) for the lengths of stay in ICU.

[14] The difference is statistically significant: $|z|=2.13 > z_{\alpha/2}$ ($\alpha=0.05$).

[15] The difference is not statistically significant: $p=0.59$ (χ^2-test).

Characteristic		**1998**		**2003**	
		Number	%	Number	%
Total MVA injuries		837	100%	974	100%
Mode of Arrival	Regular ambulance	231	27.6	247	25.4
	MICU	501	59.9	569	58.4
	Army ambulance	22	2.6	16	1.6
	Helicopter	24	2.9	21	2.2
	Private vehicle	25	3.0	62	6.4
	Other/unknown	34	4.0	59	6.0
Exit from ER	Operation room	244	29.2	242	24.9
	ICU	272	32.5	324	33.3
	Hospitalization	238	28.4	346	35.5
	Another hospital	37	4.4	37	3.8
	Death	20	2.4	20	2.1
	Other	26	3.1	5	0.5
Outcome	Death	156	18.6	114	11.7
	Rehabilitation	192	22.9	300	30.8
	Other	489	58.4	560	57.5
Stay in ICU	0 (days)	343	41.0	442	45.4
	1+ (days)	489	58.4	529	54.3
	Unknown	5	0.6	3	0.3
	Mean, days	7.8±8.8 (482 cases)		8.7±9.9 (523 cases)	
	Median, days	4		4	
Length of stay (LOS)	0-14 (days)	523	62.5	665	68.3
	15+ (days)	300	35.8	301	30.1
	Unknown	14		8	
	Mean (days)	15.3±15.0 (762 cases)		13.8±13.5 (906cases)	
	Median (days)	11		9	
Procedures	Without	383	45.8	458	47.0
	With	454	54.2	516	52.3

Table 9.6. MVA severe injuries by mode of arrival, destination of ER exits, outcome, length of stay in ICU, total LOS and procedures, in 1998 and 2003.

c. A set of indicators which can be applied for the characteristic of quality of further (in-hospital) treatment of road crash injuries, based on data from the TR, is presented in Table 9.7. The values estimated belong to 2002 where, in total, 5,171 road accident casualties were registered by the Israeli TR, of which 26.4% were pedestrians, 12.1% bicycles and the rest – motor vehicle occupants.

Indicator	**Values**
share of severe cases (ISS ≥ 16)	18.2%
share of patients delivered by MICU or ambulance	71%
mean LOS in the hospital	6.2 days
share of those who died during hospitalization	2.9%
share of treated in ICU	14.4%
average number of days in ICU	7.5 days
share of those who were in surgery rooms	34.8%
share of transferred to rehabilitation facilities upon discharge	8.4%

Source: Israeli TR, 2002 (total: 5,171 casualties)

Table 9.7. Possible set of indicators for the characteristic of quality of in-hospital treatment of road crash injuries, based on the TR data.

d. A set of indicators which can be applied for along-time comparisons of the quality of in-hospital treatment of road crash injuries, based on the TR data, is presented in Table 9.8. The consideration in based on the characteristics and outcomes of treatment of severe road crash injuries which were registered by the Israeli TR in 1998 and 2003 (more detailed data were presented above in Tables 9.5-9.6).

Indicator	**Year: 2003**	**Year: 1998**
share of severe cases out of the total	18.9%	18.4%
share of delivered by MICU + ambulance	58.4% + 25.4%	59.9% + 27.6%
mean LOS in the hospital	13.8 days	15.3 days
share of those who died during hospitalization	11.7%	18.6%
share of treated in ICU	54.3%	58.4%
average number of days in ICU	8.7 days	7.8 days
share of those who were in surgery rooms	53%	54.2%
share of transferred to rehabilitation facilities upon discharge	30.8%	22.9%

Source: Israeli TR, severe* road accident casualties (1998: 837; 2003: 974).

* with ISS≥16

Table 9.8. Possible set of indicators for along-time comparisons of the quality of in-hospital treatment of road crash injuries, based on the TR data.

9.4 Recommended system for producing national TM SPIs

9.4.1 Setting up a national system

Trauma management SPIs should estimate the speed and the quality of the post-crash care, both initial and further, in the country. Accounting for the limitations of data available in the countries, a Minimum set of Trauma Management SPIs was introduced. The Minimum set of TM SPIs is applicable for an initial characteristic of the TM system's performance.

The minimum set of the TM SPIs includes fourteen indicators which are estimated based on seven data items provided for the country. Data requirements and calculation rules for estimating the TM SPIs are presented below.

9.4.2 Data to be provided and calculation rules

The minimum set of the TM SPIs includes fourteen items as follows:

1. EMS stations per 10,000 citizens

2. EMS stations per 100 km length of rural public roads

3. Percentage of physicians out of the total EMS medical staff

4. Percentage of physicians and paramedics out of the total EMS medical staff

5. EMS medical staff per 10,000 citizens

6. Percentage of MICU out of the total EMS units

7. Percentage of BLSU, MICU and helicopters/ planes out of the total EMS units

8. EMS transportation units per 10,000 citizens

9. EMS vehicles per 100 km road length of total public roads

10.-11. Percentage of EMS responses which meet the demand for response time; accompanied by "The demand for a response time, min"

12. Average response time of EMS, min

13. Percentage of beds in certified trauma centres and trauma departments of hospitals out of the total

14. Number of the total trauma care beds per 10,000 citizens

Separate SPIs are further combined into ranks which enable comparison of the levels of trauma care system performance in different countries (see *SPI Theory*).

To estimate the TM SPIs, *a minimum data set* to be provided for the country is as follows:

1. Total number of EMS stations
2. Number of EMS staff in service (according to categories)
3. Number of EMS transportation units in service (according to categories)
4. The demand for a response time (min)
5. Percentage of EMS responses which meet the demands for response time
6. Average response time of EMS (min)
7. Total number of beds in permanent medical facilities (according to categories).

Besides, *General information* is required for the country, including:

a. population size,

b. road length – total,

c. road length – outside built-up areas.

Table 9.9 provides a detailed description of data items to be collected and Table 9.10 provides rules for calculating TM SPIs, for a country.

General information

No	Item	To be filled in
A	Population, million	
B	Road length - total, km	
C	Road length - public, outside built-up areas, km	

Data on TM

No	Item	To be filled in	Description
1	No of dispatching centers		Absolute number, for the whole country Average annual figure
2	No of EMS stations		Absolute number, for the whole country Average annual figure
4a	Number of EMS staff in service: No of physicians		Absolute numbers, for the whole country; Average annual figures
4b	No of paramedics		
4c	No of nurses		
4d	No of medical technicians		
4f	Total (including others)		
7a	Number of EMS transportation units in service: No of BLSU		Absolute numbers, for the whole country; Average annual figures
7b	No of MICU		
7d	No of helicopters/ planes		
7e	Total (including others)		
19	The demand for EMS response time, min		A value reported by the EMS authorities
20	Percentage of EMS responses meeting the demand		An estimate reported by the EMS authorities; according to a special algorithm to assess the value
21	Average response time of EMS, min		An estimate reported by the EMS authorities; according to a special algorithm to assess the value
22a	Number of trauma beds in permanent medical facilities: In certified trauma centres		Absolute numbers, for the whole country; Average annual figures
22b	In trauma department of hospitals		
22d	Total (including others)		

Table 9.9. Data to be collected

Topic	No	SPI	Calculation rules
Quality of initial treatment by EMS: Staff and equipment in service	3a	EMS stations per 10000 citizens	Number of EMS stations (No 2) divided by the population size (A) and divided by 100
	3b	EMS stations per 100 km of rural road length	Number of EMS stations (No 2) divided by the length of public roads outside built-up areas (No C) and multiplied by 100
	5a	Percentage of physicians out of EMS staff	Number of physicians (No 4a) divided by the total sum of EMS medical staff in service (No 4f) and multiplied by 100
	5	Percentage of physicians + paramedics out of EMS staff	Number of physicians (No 4a) plus number of paramedics (No 4b) divided by the total sum of EMS medical staff in service (No 4f) and multiplied by 100
	6	EMS medical staff per 10000 citizens	Total sum of medical staff in service (No 4f) divided by the population size (A) and divided by 100
	8b	Percentage of MICU out of the total EMS units	Number of MICU (No 7b) divided by the total number of EMS transportation units (No 7e) and multiplied by 100
	8	Percentage of BLSU + MICU + Helicopters/ planes out of the total EMS units	Number of BLSU (No 7a) plus number of MICU (No 7b) plus number of helicopters/ planes and (No 7d) divided by the total number of EMS transportation units (No 7e) and multiplied by 100
	9	EMS transportation units per 10000 citizens	Total sum of transportation units (No 7e) divided by the population size (A) and divided by 100
	11	EMS transportation units per 100 km of road length	Total sum of transportation units (No 7e) divided by the total length of public roads (B) and multiplied by 100
Speed of initial treatment by EMS: Time values of initial treatment	19	The demand for EMS response time, min	A value as reported by EMS authorities
	20	Percentage of EMS responses meeting the demand	A value as reported by EMS authorities
	21	Average response time of EMS, min	A value as reported by EMS authorities
Quality of further treatment: facilities in service	24a	Percentage of beds in certified trauma centres and trauma departments of hospitals out of the total number of trauma beds	Number of beds in certified trauma centres (No 22a) plus number of beds in trauma departments of hospitals (No 22b) divided by total number of beds (No 22d) and multiplied by 100
	25	Number of the total trauma care beds per 10000 citizens	Total number of beds (No 22d) divided by population size (A) and divided by 100

Table 9.10. TM SPIs estimated: calculation rules

9.4.3 Data requirements

General requirements for the data (see Table 9.9) are as follows:

a. The data should be provided, and the estimates should be given, *for the whole country*. If the data for the whole country are unavailable or different policies are applied in different regions of the country (e.g. lands, cantons, counties), then the information is requested for *two representative sub-areas* of the country. As such areas it is recommended to select one highly-populated and one scarcely-populated area of the country. In this case, the "General Information" items should also be provided for the sub-areas selected.

b. Most of the data, except for estimates of the EMS response time, are annual figures, which should be extracted from the *national statistics*.

c. The estimates of EMS response time (No 20, No 21) are reported by the EMS authorities. The average response time of EMS (No 21), is estimated by each country using a special algorithm.

<u>Sampling rules recommended for estimating average EMS response time</u>

1. The sampling units are the EMS calls on road crash related injuries.
2. In order to ensure a nationally representative sample, the EMS stations to be surveyed need to be geographically spread over the country. To satisfy this demand, two stratification factors are suggested for building the sample:
   (a) the type of area in which the EMS station is located (e.g. 3 strata: urban, suburban, rural) and

   (b) the magnitude of the EMS station's fleet (e.g. two strata: high and low).
3. Within each stratum, a random sample of the EMS stations needs to be selected. It is suggested to use results from 10 different EMS stations within each stratum, and to ensure that at least 5% of the total number of EMS stations is represented.
4. Within each of these stations, at least 30 calls related to road crash injuries should be selected. It is recommended to collect 400 cases per stratum: if 400 cases are selected randomly from the total list of road crash related injuries, a sampling error of 5% and a confidence level of 95% are guaranteed at the level of each separate stratum, so certainly on the national level.
5. To estimate the total SPI - EMS average response time for the whole country, the shares of EMS stations belonging to different strata, throughout the country, are required.

   The calculation will be as follows:

   Let *SPI_i* be the SPI value estimated for the *i-th* stratum. Let λ_i be the share of total EMS stations attributable to stratum *i*, where $\Sigma\lambda_i = 1$.

   Then, the total SPI equals:

$$SPI = \sum_i \lambda_i SPI\_i$$

   Since we expect the *SPI_i*'s for the different strata to be independent, the variance for the total SPI then equals:

$$Var(SPI) = \sum_i \lambda_i^2 Var(SPI\_i).$$

   A 95% confidence interval for the total SPI is equal to

$$SPI - 1.96\sqrt{Var(SPI)} \leq trueSPI \leq SPI + 1.96\sqrt{Var(SPI)}.$$

d. The data collection should be accompanied by a Glossary (see Sec.9.7).

e. The data should preferably be provided by a medical expert, who is involved in trauma care activities in the country and is well familiar with medical (and especially EMS and trauma) databases and statistical publications in the country.

f. The data should be collected and the TM SPIs be calculated on an annual/ bi-annual basis.

9.4.4 Example of calculating TM SPIs

Based on the data collected during the project, following is an example of the data provided and the TM SPIs estimated for a country: Belgium, for 2003.

General data

Population, million	10,356
Road length - total, km	149739
Road length - public, outside built-up areas, km	120407

Data on Trauma management

(1) No of dispatching centers	16
(2) No of EMS stations	300
Number of EMS staff in service: (4a) No of physicians	1400
(4b) No of paramedics	0
(4c) No of nurses	450
(4d) No of medical technicians	7500
(4f) Total	9350
Number of EMS transportation units in service: (7a) No of BLSU	357
(7b) No of MICU*	100
(7d) No of helicopters/ planes	n/a
(7e) Total	480
Comments	*Called MOG in the flanders region and MSUR in the waals region

(19) The demand for EMS response time, min	15 min
(20) Percentage of EMS responses meeting the demand	100%*
(21) Average response time of EMS, min	6 min*
Number of trauma beds in permanent medical facilities: (22a) In certified trauma centres	0
(22b) In trauma department of hospitals	356
(22d) Total	356
Comments	* based on local EMS service, region Antwerp

SPI values estimated for Belgium

(3a) EMS stations per 10000 citizens	0.29
(3b) EMS stations per 100 km of rural road length	0.25
(5a) Percentage of physicians out of EMS staff	15%
(5) Percentage of physicians + paramedics out of EMS staff	15%
(6) EMS medical staff per 10000 citizens	9.0
(8b) Percentage of MICU out of the total EMS units	21%
(8) Percentage of BLSU + MICU + Helicopters/ planes out of the total EMS units	95%
(9) EMS transportation units per 10000 citizens	0.46
(11) EMS transportation units per 100 km of road length	0.32
(19) The demand for EMS response time, min	15 min
(20) Percentage of EMS responses meeting the demand	100%*
(21) Average response time of EMS, min	6 min*
Comments	* based on local EMS service, region Antwerp

(24a) Percentage of beds in certified trauma centers and trauma departments of hospitals out of the total number of trauma beds	100%
(25) Number of the total trauma care beds per 10000 citizens	0.34

It should be noted that in the example above the values of the EMS response time cannot be considered as representative for the whole country.

9.4.5 Analysis and reporting

The whole set of the TM SPIs is recommended for a characteristic of the trauma management system in the country. Time-series of the SPIs can be used for monitoring the system's development through time – see examples for Germany and Israel in Sec. 9.3.

The countries can be compared using selected TM SPIs, e.g. by availability of the EMS stations, by availability and composition of the EMS medical staff, by availability and composition of the EMS transportation units, by characteristics of the EMS response time or by availability of trauma beds in permanent medical facilities (see Vis and van Gent, 2007).

For comparing countries, a combined indicator may be applied. The combined indicator is estimated by means of ranking the values of separate TM SPIs and weighting the results together (see Hakkert et al, 2007). Several rankings are applied, which finally attribute each country to one of five levels of the TM system's performance such as: "high", "relatively high", "medium", "relatively low" or "low". This way, the combined indicator provides an overall characteristic of the system, indicating a "higher" or "lower" level of the trauma management system's performance in a certain country *relatively* to other countries considered.

9.5 Discussion

The suggested trauma management SPIs are recommended for application as a minimum set for the initial characteristic of the trauma care system's performance in the country. The minimum set was introduced recognizing that the data are not easily available in most European countries. In the majority of countries, trauma care indicators are currently not in use in the decision-making practices on road safety. Hence, for estimating the TM SPIs special efforts will be required for the data collection.

The suggested set of trauma management SPIs enables characterisation, to a certain extent, of the scope and the quality of the post-crash care in the country. This comes in terms of the EMS treatment potential, EMS response time and the treatment potential of permanent medical facilities. However, one should admit that the SPIs' message is limited mostly to the availability of trauma care services and, partly, to their quality (e.g. in terms of shares of higher-quality resources). In other words, they characterise the *possibilities* of trauma care to be provided for the road crash injuries.

The suggested SPIs do not include values on actual performance of the trauma care system, i.e. on trauma care, which was *actually applied*, e.g. in terms of EMS units which treated the casualties or the forms of treatment provided in the hospitals. Such indices to be estimated need data from a trauma registry and other hospital databases. This option was considered at the beginning of the TM SPIs' development but later denied due to a lack of access to the data required, in the majority of countries (see Hakkert et al, 2007). In the future, with further development of combined road crash related databases, an extended set of TM SPIs with both figures on the availability of services and characteristics of quality of the treatment supplied, should be considered for application.

The suggested trauma management SPIs and the combined indicator are applicable for the comparison of TM systems in different countries or in different regions of the same country. However, one should remember that the estimates and the comparisons are performed in terms of characteristics which are associated with the treatment of road crash victims. In other words, the TM SPIs should not be considered as an overall estimate of the trauma care system in the country, which covers other types of injury and diseases as well.

Moreover, we realise that, in general, the trauma care system is a matter of strategic approach with necessary guidelines, standards and regulations; distribution of emergency care; education and training of trauma teams; definition of clinical capabilities of hospitals, etc, where the system's performance is followed up in a long-term and is estimated in terms of actual treatments applied and their outcomes (changes in mortality and the quality of life). In this sense, the suggested SPIs can be seen as a component of the measurable part of the trauma care system's performance, mostly in the context of road crash related injuries.

9.6 References

1. CEC (2003). European road safety action programme: Halving the number of road accident victims in the European Union by 2010: A shared responsibility. Commission of the European Communities (CEC), Communication from the Commission Com (2003) 311 final.
2. Cryer C., Langley J. D., Jarvis S. N., MacKenzie S. G., Stephenson S. C. R., Heywood P. (2005) Injury outcome indicators: the development of a validation tool. (Includes statistical tables). Injury Prevention 11(1): 53-5.
3. ETSC (1999). Reducing the Severity of Road Injuries through Post Impact Care. European Transport Safety Council, Brussels.
4. ETSC (2001) Transport Safety Performance Indicators. European Transport Safety Council, Brussels.
5. Hakkert A.S., Gitelman V. and M.A. Vis (Eds.) (2007). Road Safety Performance Indicators: Theory. Deliverable D3.6 of the EU FP6 project SafetyNet.
6. Heruti R. J., Stein M., Barel V., Shemer J. (1997) Trauma registry - the concept and its application in Israel. Harefuah 133(3-4): 155-9 (in Hebrew).

7. Peden M, Scurfield R, Sleet D, Mohan D, Hyder A, Jarawan E and C Mathers eds. (2004): *World Report on Road Traffic Injury Prevention*, World Health Organisation, Geneva.
8. Peleg K., Aharonson-Daniel L. (2004) Road Traffic Accidents - Severe Injuries. How missing data can impair decision making? Harefuah 143(2):111-115 (in Hebrew).
9. Schmiedel R. & Behrendt H. (2002). Leistungen des Rettungsdienstes 2000/01. Zusammenstellung von Infrastrukturdaten zum Rettungsdienst 2000 und Analysen des Leistungsniveaus im Rettungsdienst für die Jahre 2000 und 2001. Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit, Heft M 147.
10. Shemer J., Shapira S.C. (1996) Establishing a nationwide trauma system in Israel. Harefuah 129(11): 526-528 (in Hebrew).
11. Trauma in Israel 2000 (2002). Israel National Center for Trauma and Emergency Medicine Research, Publication no. 6002, Ed: Peleg K, Tel-Aviv (in Hebrew).
12. Trauma in Israel 2002 (2004). Israel National Center for Trauma and Emergency Medicine Research, Publication no. 6003, Ed: Peleg K, Tel-Aviv (in Hebrew).
13. Vis, M.A. and A.L. van Gent (Eds.) (2007) Road Safety Performance Indicators: Country Comparisons. Deliverable D3.7a of the EU FP6 project SafetyNet.

9.7 Glossary

Trauma management/ post-crash trauma care: No common definition is available. The "Glossary of terms for burden of injury studies" suggests a definition of term *Trauma (Care) system* as follows: "A system of health care provision that integrates and coordinates prehospital emergency medical service resources and hospital resources to optimize the care and therefore the outcome of traumatically injured patients"[16].

Notification time: The time interval between the crash occurrence and the emergency call is made.

Response time: The time interval between emergency call and the response of the EMS (thus the time of arrival of the EMS at the scene of crash).

Arrival time: The time interval between the crash occurrence and the response of the EMS (thus the time of arrival of the EMS at the scene of crash).

Medical terms:

Emergency Medical Services (EMS) System includes the emergency dispatch system and the emergency units. The dispatch system takes incoming calls for emergency care.

A dispatching centre is an office which is informed in case of emergencies (mostly by telephone calls) to ask for medical assistance. The dispatching centre then alarms and coordinates the EMS units.

An EMS station is the location/base station where at least one EMS vehicle or helicopter/plane (and in most cases its crew) are positioned.

The EMS units are mostly ambulances but also helicopters/planes/boats, which arrive at the scene of crash and provide initial medical assistance to injured patients. There are different forms of EMS units, which depend on the type of a transport means (helicopter, ambulance); EMS vehicle equipment (mobile intensive care unit; basic life support unit); medical staff arriving with the vehicle.

[16] See: http://www.eurosafe.eu.com/csi/eurosafe2006.nsf/wwwVwContent/l2injuryglossary.htm

The medical staff may include a physician, a paramedic, a “critical care” nurse, and an emergency medical technician.

Advanced life support (ALS): medical care given by medical doctors and nurses trained in critical care medicine with the use of specialized technical equipment, infusion of fluids and drugs aimed to stabilize or restore vital functions

Basic life support (BLS): consists of emergency medical care to restore or sustain vital functions (airway, respiration, circulation) without specialized medical equipment and to limit further damage in the period preceding advanced medical care.

Mobile intensive care unit (MICU): a unit with a medical doctor or paramedic and a nurse transported to the scene of the crash with the knowledge, skills and equipment necessary for performing advanced life support.

Basic Life Support Unit (BLSU): a transportation unit with personnel and equipment necessary for performing basic life support.

Emergency medical technician: a person who received training in emergency medical care for sick or injured patients in need of transportation to a hospital. This training includes BLS and the ability to assist doctors and nurses in the delivery of ALS.

Paramedic: an emergency medical technician who received further training for the delivery of some aspects of ALS care.

The term *„emergency call“* includes all calls which are answered by EMS dispatching centre and which lead to an emergency response by the EMS. The term includes false and abusive alarms, but excludes calls due to patient transportation requests.

EMS rides are rides of the EMS in consequence of emergency calls, including false and abuse alarms**.**

EMS vehicles according to European Norms EN 1789:

The European norm EN 1789[17] "Medical vehicles and their equipment - Road ambulances" (2007) standardises the minimum equipment of EMS vehicles. According to this norm there are three types of EMS vehicles:

Type A_1/A_2: A vehicle that is appropriate to transportation of one or more patients - transportation ambulances

Type B: A vehicle that is equipped for transportation, basic life support and medical monitoring of patients (similar to *BLSU*).

Type C: A vehicle that is equipped for transportation, advanced life support and medical monitoring of patients (similar to *MICU*).

Meanwhile, there is no European norm for helicopters, planes, and boats. Thus any helicopters, planes, and boats that are in use by EMS can be mentioned.

Definitions of crash injury severity used by the police and national crash databases:

Killed (fatality): a person who died as a result of the crash, or died of his injuries within 30 days of the crash.

Seriously injured: a person who was hospitalized as a result of the crash for a period of 24 hours or more.

Slightly injured: a person who was injured as a result of the crash and was not hospitalized, or was hospitalized for a short period (up to 24 hours).

Hospitalized[18]*:* non-fatal victims who are admitted to hospital as in-patients.

---

[17] See: http://www.cen.eu/catweb/43.160.htm

Definitions of crash injury severity using medical scales:

Abbreviated Injury Scale (AIS): a score from 1-6, for anatomically different injuries, indicating the chance that such injuries lead to death. AIS 6 injuries are usually considered to lead to inevitable death, AIS 5 to probable death (life-threatening), AIS 4 to possible death (very serious), AIS 3 are serious injuries; other grades (1=slight injuries, 2=moderate injuries) rarely lead to death. AIS 0 means "no injury". ***AIS 3-6*** correspond to patients which are ***hospitalized***.

Injury Severity Score (ISS): a score based on the AIS, which accounts for multiple injuries in one patient; calculated as a sum of the squares of the highest AIS grades in each of the three most severely injured body regions (out of 6 body regions). Groups of ISS values, which are usually applied for a qualification of injury's severity, are: ISS 1-8 for slight injuries, ISS 9-14 for medium injuries, ISS 16-25 for serious injuries, ISS 25+ for very serious injuries. ***ISS 16+*** indicates ***severe*** injuries.

Glasgow-Coma Scale (GCS): a score that focuses on the neurological situation of the patient by the item „eyes open" and on the verbal and motoric reactions of the patient. Maximum value: 15 (no neurological disorders), minimum value: 3 (severe neurological disorder). Groups of values, which can be applied for a quantification of injury's severity, are: GCS 13-15 - slight craniocerebral injury, GCS 9-12 - „medium severe" craniocerebral injury, GCS < 9 - severe craniocerebral injury, possibility of long-term/lasting disorders. ***GCS < 9*** indicates ***severe*** injuries.

Trauma beds:

No common definition is available. Typically are considered:
1) all beds in trauma centres/ trauma departments of hospitals;
2) beds in surgery departments of regular hospitals.

Trauma Centre:

No common definition is available. The minimum threshold of basic clinical capabilities to be provided by a trauma centre is as follows[19]:

[18] In use by IRTAD – International Road Traffic and Accident Database

[19] ETSC (1999)

In-house 24 hours a day:
Emergency Medicine
Anaesthesiology
General Surgery and any life saving surgery (such as urgent external fixation for pelvic fractures, vascular surgery)
Radiology: a mobile X-ray apparatus should be located in the resuscitation room and the other X-ray facilities such as CT-scan should located near the emergency department

On call promptly available:
ESSENTIAL:
Anesthesiology (2nd team)
General Surgery (2nd team)
Neurosurgery (2nd team)
Orthopaedic Surgery
Maxillo facial Surgery
Interventional Radiology

DESIRABLE :
Pediatric Surgery
Vascular Surgery
Urologic Surgery
Plastic Surgery
Thoracic surgery

Facilities and resources: available in-house 24 hours a day:
X-ray and Ultrasonography
CT-scan
Trauma operating room with staffed personnel
Clinical laboratory service
Blood bank with adequate storage facilities
Rehabilitation team for the acute trauma phase

The facilities and medical instruments for every clinical procedure must be recorded on dedicated checklists which are monitored every day by trained nursing staff overseen by the trauma coordinator.

10 Summary and Conclusions

Safety performance indicators (SPIs) are defined as the measures (indicators), reflecting those operational conditions of the road traffic system, which influence the system's safety performance. Basic features of SPIs are seen in their ability to measure unsafe operational conditions of the road traffic system and in their independence from specific safety interventions. The SPIs' purpose is to serve as assisting tools in assessing the current safety conditions of a road traffic system, monitoring the progress, measuring impacts of various safety interventions, making comparisons, and other uses.

Based on the potential of different road safety domains for promoting road safety, seven problem areas were stated as central for road safety activity in Europe and, therefore, selected for the development of SPIs. They are: alcohol and drug-use; speeds; protective systems; daytime running lights; vehicles (passive safety); roads (infrastructure) and the trauma management system.

For each area, SPIs were developed based on the proven relationships between accident/injury occurrences and the area's characteristics as well as on the experiences and data available in the EU (and some other) countries.

The Road SPI Theory report (Hakkert et al, 2007) dealt with the theory behind the development of SPIs in each of the seven safety areas. Two other reports: the Road SPI Country Comparisons (Vis and van Gent, 2007a) and the Road SPI Country Profiles (Vis and van Gent, 2007b) - provided results on the data collected so far for each of the 27 cooperating countries (the EU at the end of 2006, Norway and Switzerland) and the SPIs estimated, based on the data submitted by each of the countries.

This report is called a Manual as it should help the countries to establish the necessary systems of data collection for producing national SPIs, in each one of the predefined safety fields, and to make them comparable on a European level. For each safety area, the report defines quantitative SPIs, demonstrates existing practices for their measurements, provides best practice examples (when available), and details the procedures which are necessary to collect and process the required data for the estimation of the SPIs' set on a national level.

Besides, the report provides a general theoretical background concerning the sampling issues in estimating SPIs (e.g. in Chapter 2 and in the Appendix). Regarding setting up an SPI survey, the main questions discussed were: sampling procedures to obtain a national sample; sampling size; sampling error; stratified sampling (combination into a single SPI by weighting); representativeness of the results and estimating confidence intervals of the SPI values.

Recognizing the potential for road safety improvements coming from the use of harmonized SPIs across the EU, enabling benchmarking as a proven tool in road safety policy, the Member States are encouraged to seek ways leading to the application of a uniformed methodology for producing national SPIs. The procedures and methods presented in the *Manual* should be treated as minimum quality requirements for producing national SPIs, in each one of the predefined safety fields.

Alcohol and drugs

In the area of alcohol and drugs, the proposed SPI is the number and proportion of severe and fatal injuries resulting from crashes involving at least one active road user who was impaired by psychoactive substance abuse (concentration above a predetermined impairment threshold).

The proposed SPI can be implemented step by step, starting with the BAC of fatally injured drivers and gradually extending to a larger set of psychoactive substances used by all active road users involved in severe injury crashes. Chapter 3 detailed the successive requirements for each step as well as the method, measurement tools and quality control issues.

It is emphasized that ethical and privacy problems might arise when body fluids are collected from active road users. Hence, the use of data on impairment by psychoactive substances for judicial purposes should be discussed in the country as well as other ethical and privacy problems.

Essential prerequisites for the SPIs' production system on alcohol and drugs to be established is that the police should ensure that blood, breath and/or saliva samples are taken from all drivers involved in on-the-scene fatal accidents. Police usually collect such samples for judicial purposes. The police should report the results to the agency responsible for road accident statistics or produce and publish the SPIs themselves.

The national road transport authorities should present to the national parliaments a proposal for the legislation necessary to require the police to collect the necessary samples.

The national forensic laboratory or nationally authorized laboratory should carry out the analyses of the samples. These laboratories across Europe should cooperate to establish international norms for analysis methods and routines and ensure that these norms are complied to.

The national agency responsible for road accident statistics should incorporate the alcohol and drug data from involved drivers into their road accident database.

Speed

The Speed chapter contains a manual for guiding the planning and implementation of representative vehicle speed surveys. It is aimed at providing guidance to professionals, practitioners, and policymakers in the field of transportation and road safety for the planning, collection, processing, and analysis of vehicle speed data.

When setting up the survey, two issues should be considered: the purpose of speed measurements and expected data outcomes. The indicators have to be chosen according to road safety targets and the chosen methodology must allow computing these indicators. Chapter 4 details which locations are suitable for speed measurement, which road types should be considered and how the set of measuring locations can be sampled; which time periods are valid for speed measurements; how to determine speeds for different types of vehicles on the basis of identified requirements for speed measurements, and so on.

The Manual provides detailed recommendations on the selection of measurement sites (random procedure, what "appropriate location" means, the road types to be considered); periods of measurement; practical considerations for the measurements (sample size, measuring devices, additional information required – on traffic count, speed limits, etc); data control and analysis; documentation and reporting.

The speeds are analysed for free flow hours only. The minimum set of speed indicators, for each road type, should include:

- Average speed for light vehicles during day
- Average speed for light vehicles during night
- Standard deviation of speed for light vehicles during day
- Standard deviation of speed for light vehicles during night
- 85th percentile of speed for light vehicles during day
- 85th percentile of speed for light vehicles during night
- Percentage of light vehicles over the speed limit during day
- Percentage of light vehicles over the speed limit during night
- Percentage of light vehicles 10 km/h over the speed limit during day
- Percentage of light vehicles 10 km/h over the speed limit during night

Protective systems

The SPIs developed in this field are:

– daytime wearing rates of seat belts at front seats (passenger cars + vans /under 3.5 tons), at rear seats (passenger cars + vans /under 3.5 tons), by children under 12 years old (restraint systems use in passenger cars), and at front seats (HGV + coaches /above 3.5 tons); and

– daytime usage rates of safety helmets by cyclists, moped riders, motorcyclists.

The SPIs are estimated by means of a national observation survey, where the measurements should be classified according to motorways, other rural roads and urban roads. The values for major road types should then be aggregated into one indicator (of each type) for the country.

The national system for producing SPIs on protective systems should fulfil the requirements as detailed in Chapter 5, which concern the sampling demands and procedures, data collection and processing, documentation requirements, evaluation and aggregation rules. Besides, calculation rules are suggested for estimating over time progress, e.g. the annual increase in the protective systems' usage rate, the conversion rate – a year-to-year decrease of non-use of the devices. Many calculation examples accompany the procedures and the requirements presented.

It is stated that assessing protective systems' use is a prerequisite for recognising and evaluating weaknesses of road safety systems related to certain types of poor road users' behaviour. Identifying the lack of protective systems' use among road users allows for setting of priorities and the launching of intervention programmes.

International or regional comparisons of protective systems' use rates are vital tools for recognising deficiencies, setting priorities and stimulating efforts at the political level. Applying the common rules presented in this Manual allows the production of reliable and accurate indicators that are comparable among the countries (regions).

Daytime Running Lights (DRL)

The DRL SPIs are defined in the form of percentage of vehicles using daytime running lights, where the value is estimated for different road categories and for different vehicle types.

A recommended system for producing national DRL SPIs is presented in Chapter 6. The system is based on a national observation survey of the DRL use. The Manual provides relevant details on the method of data collection (periods, procedures), definition of observation sites, data collected, sampling demands, rules for estimating SPIs (per road categories, vehicle types and in total), quality control and reporting issues. An example of calculating SPIs is provided for Hungary.

The suggested system of DRL SPIs enables estimation of the DRL usage rates at the national level. The system may serve as a background for both the countries' comparisons and along-time considerations on the DRL-related issues.

The recommended system for producing DRL SPIs may also be applicable on a regional/ local level, if necessary adjustments are performed in the sampling procedures.

It is stated that the background information on the DRL legislation is essential for a correct interpretation and comparison of the results. For example, comparing the countries' DRL usage rates it is reasonable to take into account whether the countries have a law/ regulation on obligatory use of DRL and if they do, for how long.

In countries, where the automatic DRL was introduced a long time ago (e.g. Sweden, Norway), according to expert estimates, current DRL usage rate is close to 100%, thus the DRL usage rate as a behavioural safety performance indicator does not have practical implications any more. In general, once the option of automatic DRL is introduced Europe-wide, the DRL indicators will lose their importance.

Vehicles (passive safety)

National vehicle fleet's crashworthiness and compatibility are suggested as SPIs in this field.

EuroNCAP is widely used as an indicator of passive safety for individual vehicles to give consumers a guide to the crashworthiness of specific makes and models. To provide a crashworthiness measure for entire vehicle fleet, for each country a EuroNCAP score is attributed to eligible vehicles. An average figure is then calculated for each year and weighted by the number of vehicles present in the current fleet from that year. An overall average EuroNCAP score is then awarded for each country and together, with the median age of passenger cars in the fleet, these two figures make up the SPI for each country.

Compatibility means that passenger vehicles of disparate size provide an equal level of occupant protection in car-to-car collisions. Vehicle mass is one of the most significant factors affecting driver injury in car-to-car injury, and an incompatible vehicle induces high risk for the occupants in the other vehicle.

In the case of compatibility, an SPI related to vehicle fleet composition is provided. Fleet composition indicates the size of vehicle types within the total fleet. More particularly, since collisions between the very smallest and the very largest vehicles (powered two-wheelers and heavy goods vehicles, respectively) are the most problematic, proportions of these two vehicle types are compared. Besides, a “relative gravity” is calculated for each fleet. This uses the vehicle fleet data to calculate the gravity of the possibility of collisions between incompatible vehicles (namely cars, HGVs and motorcycles).

Monitoring these issues is related to the accuracy of the vehicle fleet database as a record of the vehicles that are actually on a country’s roads at a point in time. Additionally, there are minimum requirements for the level of detail contained within the database, to make calculation of the SPIs for vehicles possible, such as:

1) Provide detailed and accurate descriptions of vehicle makes and models.

2) Classify vehicles according to vehicle-types compatible with CARE definitions.

3) Distinguish between smaller (less than 3.5 tonnes) and larger goods vehicles, since these are significantly different when assessing their compatibility in collisions with passenger cars or vulnerable road users.

Roads

SPIs for roads aim to assess the safety hazards by infrastructure layout and design. Two SPIs for roads were developed: the road network SPI and the road design SPI. Because there were no best practices yet that used these SPIs, a case study was executed. In the case study, the SPIs were calculated for the province of South Holland, The Netherlands. The purpose of the case study was twofold: first, to illustrate the application of the SPIs' concept, and, second, to attain more insight into the practical applicability, strengths and weaknesses of the method, aiming to provide guidelines for its future use.

The evaluation procedure (Chapter 8) included details on: collecting and processing of data; determination of a list of urban area connections that need to be assessed; comparison of the theoretically needed road categories with the actual road categories, and calculation of the actual SPIs. All rural and national roads were included in the study, no sample was taken. The results are not representative for the Netherlands in general, since the roads that were included in the analysis are not a random sample of all Dutch roads. (For a nationally representative sample, 384 randomly selected connections need to be analysed, proportionally distributed over the connection types on the basis of the number of kilometres of each connection type).

It appeared to be possible to automate the process. Moreover, most of the data necessary for the calculation of the SPIs was available in MultiNet. For the calculation of the road design SPI, EuroRAP Road Protection Scores for individual road segments or routes were applied.

The results from the case study seem reasonable. However, to get more insight into the meaning of the found values consideration of the results in more detail, in consultation with the road authority, is recommended. Moreover, it would be interesting to compare the values for different countries. The list of issues, that need further research should include: a more detailed definition of types of urban areas; considering the areas' limitations in case of natural barriers or administrative borders; the assignment of the theoretically needed connections to the actual road network, etc.

Applying the method in various countries is essential for further investigation of the above issues and further method's development.

In general, the case study proved that it is possible to apply the road SPIs' concept in practice. Furthermore, the process could be automated and thus can be applied on a larger scale as well. The SPIs provided insight into the safety quality of the network as a whole and of individual roads. Therefore, they can be recommended for application to assess the safety of the road network.

Trauma management

Recognising the importance of the trauma management indicators for road safety programs and for road safety policy-making in general, presently only a few examples of such indicators can be found in use.

In Chapter 9, two countries having comprehensive data on trauma care were presented: Germany – on Emergency Medical Services (EMS) and Israel – on in-hospital trauma care (Trauma Registry). In each country, the data are collected and the performance indicators are estimated systematically, which could serve as a reasonable background for the application of trauma management SPIs.

Trauma management SPIs should estimate the speed and the quality of the post-crash care, both initial and further, in the country. Accounting for the limitations of data available in the countries, a minimum set of Trauma Management (TM) SPIs was developed, which referred to the availability of EMS stations; availability and composition of the EMS medical staff; availability and composition of the EMS transportation units; characteristics of the EMS response time and the availability of trauma beds in permanent medical facilities.

The minimum set of the TM SPIs includes fourteen indicators which are estimated based on seven data items provided for the country. Data requirements and calculation rules for estimating the TM SPIs were presented in Chapter 9. Besides, general requirements for the data were detailed, including sampling rules recommended for estimating the average EMS response time.

The suggested TM SPIs are recommended for application as a minimum set for the initial characteristic of the trauma management system's performance in the country. In the majority of countries, special efforts will be required for the data collection for estimating the TM SPIs.

It was emphasised that the suggested set of SPIs enables to characterise mostly the *possibilities* of trauma care to be provided for the road crash injuries (namely, the EMS treatment potential, EMS response time and the treatment potential of permanent medical facilities). The suggested SPIs do not include values on *actual performance* of the trauma care system, e.g. on EMS units which treated the casualties or the forms of treatment provided in the hospitals. For such indices to be estimated data are needed from a trauma registry and other hospital databases. This option was discussed at the beginning of the TM SPIs' development but later denied due to a lack of access to the data required, in the majority of countries.

In the future, with further development of combined road crash related databases, an extended set of TM SPIs with both figures on the availability of services and characteristics of quality of the treatment supplied, should be considered for application.

Annexes

Annexes to Chapter 4 on Speeds

Annex 1: Definition of sampling units for a speed survey

To define the sampling units, one must first define what the population (in statistical terms) is. If one wants to compute annual indicators, the population is constituted by all the speeds observed for all the vehicles driving on the road network during the year. Obviously, it would be impractical to monitor the speed of all vehicles constantly and impossible to store and analyse the resulting data. A selection of time slices, vehicles and places must, thus, be made.

For the sake of simplification, we will not consider the road network as continuous, but as an aggregation of small road segments. These segments should be the most internally homogenous as possible. In that way, the population becomes finite because the speed of a vehicle does not have to be measured continuously but only one time per homogenous road segment.

As we consider that each vehicle has its own specific speed on each of road segments, we can now define the population by selecting a road network and a time period. In space, the population is limited by a specific road network and in time by a specific time period (e.g. one year). This way of defining the population is confirmed by Isaksson (2003) and Walters (2001). We can visualize the population as a matrix whose columns would represent each road segment and lines each vehicle passage on a segment during the period of interest (e.g. one year).

Still, it would not be appropriate to take a random sample from the elements of this imaginary matrix. Indeed, it would result in having to measure speed of some isolated vehicles on a multitude of different parts of the road network and moments of the year, which is impractical. The only realistic way to proceed is to concentrate on a smaller quantity of road units and to measure speed of several vehicles on each of these units during fixed time slices (i.e. sampling some of columns of the matrix and then selecting elements of each columns characterised by consecutive time of passage).

We argued in Section 4.4.2.2 of this Manual that the selection of time slices should not be random but based on the knowledge of typical speed variations. The selection should ensure that the chosen times of measurement are representative for the whole period of interest (i.e. the selected elements in each column of the matrix are representative for the whole column).

That is why the sampling problem can be simplified to sampling of road segments amongst the whole road network and the road segments can, therefore, be considered as primary sampling units.

Annex 2: Sample size calculation and example

Determination of the minimum sample size is a difficult task as variability in speed data can occur at different steps. For the purpose of calculation, we will suppose that the measure of speed at each location is correct and, thus, determine the sample size only as a function of the variability that is observed *between* the measurement locations.

As always, the sample size is determined by the maximum margin error that one wants to accept and by the desired confidence interval of estimate (typically 95%).

As a simplification, we will imagine that all locations have similar traffic counts, so that the SPI can be computed as a simple mean of values obtained for each location. In reality, in the process of aggregating the data from the locations to a national SPI, the location indicators will be weighted by traffic counts measured at these locations (see Section 4.4.3 for details).

For an SPI that is determined as a mean value, the simple random sampling theory instructs that a confidence interval for the population mean can be determined by (see Groves *et al.*, 2004; Levy and Lemeshow, 1999, amongst others):

$$\bar{x} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z\frac{\sigma}{\sqrt{n}} \qquad \text{Equation A}$$

with μ the population mean,

$\bar{x}$ the survey mean,

n the sample size (i.e. number of measuring locations),

σ the population variance,

z the z-value of the desired confidence interval (e.g. 1.96 for a 95% confidence).

The maximum error margin with which the sample mean and the true population mean may differ from one another is given by:

$$md \geq z\frac{\sigma}{\sqrt{n}} \qquad \text{Equation B}$$

Isolating n from this formula leaves us thus:

$$n \geq \frac{z^2\sigma^2}{md^2}. \qquad \text{Equation C}$$

Obviously, the more accuracy and the more confidence are desired, the bigger the sample size should be. This formula can be used when the population variance is known. However, we do not know the variance between the mean speeds of all the segments that constitute the road network for a road class. A general rule of thumb, however, is that when the sample size is larger than 30, one can replace the population variance (σ^2) by the sample variance (s^2). This population/sample variance needs to be determined based on results from a prior study on the same population or from a small pilot study. A sampling error of 1km/h and a confidence level of 95% are reasonable figures to work with.

However, the inherent variability in speeds may not be the only reason of variability of speed between the measuring locations. The variability may be increased by the fact that some sites do not correspond exactly to the criteria that define a good measuring location (see Section 4.4.2.1.2). Before willing to calculate the sample size, one must first ensure that all the measuring locations are good and, thus, belong to the same statistical population.

Example

Data below represent average speeds (in km/h) at 30 fictional measuring sites. The sites are considered as homogenous (same road class and road design) and thus are part of one coherent "population". As stated above, it is also assumed that all sites have similar traffic counts so that the SPI can be computed via a simple arithmetic mean of the sites' indicators. The SPI "average speed" is, thus, equal to 88.5 km/h.

Site 1	83.8	Site 7	80.5	Site 13	86.1	Site 19	90.7	Site 25	87.2
Site 2	87.7	Site 8	88.7	Site 14	85.7	Site 20	95.2	Site 26	82.0
Site 3	94.4	Site 9	88.0	Site 15	99.0	Site 21	102.5	Site 27	88.5
Site 4	94.6	Site 10	84.8	Site 16	88.4	Site 22	78.1	Site 28	82.9
Site 5	94.8	Site 11	85.6	Site 17	93.5	Site 23	77.0	Site 29	92.9
Site 6	94.6	Site 12	84.6	Site 18	79.8	Site 24	88.6	Site 30	95.3

Equation B allows estimating the margin of error in this estimation of the population average speed. We chose here to use a confidence level of 95% (z=1.96) and to approximate the population standard deviation by the sample standard deviation (6.2 km/h):

$$md \geq 1.96 \frac{6.2}{\sqrt{30}} = 2.2 km / h$$

Thus, there is a 95% probability that the population average speed lies between 86.3 and 90.7 km/h.

Inversely, equation C allows assessing the ideal sample size for a given desired margin of error. The results are given in the below table.

Sample size estimation for σ=6.2

		Accuracy/Sampling error (*md*)			
Confidence level	**z-value**	**0.5 km/h**	**1 km/h**	**1.5 km/h**	**2 km/h**
0.90	**1.65**	423	106	47	26
0.95	**1.96**	597	149	66	37
0.98	**2.33**	844	211	94	52
0.99	**2.58**	1034	259	115	64

Annex 3: Example of documentation for a measuring site (inspired by the form used in Belgium)

Internal code of the site °°°

Start of measurement (D/M/Y – H:M:S): °°° End of measurement (D/M/Y – H:M:S): °°°

Address of the site
- ZIP code / City °°°
- Road: °°°
- Nearest House number: °°°
- Kilometre marker: °°°
- Geographic coordinates N°°° E°°°
- Other information °°°

Characterization of the road
- Road class: °°°
- Situation ☐ Built-up area ☐ Outside built-up area
- Number of directions: °°°
- Directions separated by
  - Road marking ☐ Yes ☐ No
  - Central reservation ☐ No ☐ Yes
  - Note: °°°
- Direction of measured flow: °°°
- Number of lanes in direction of measurement: °°°
- Number of lanes in direction opposite to direction of measurement: °°°

Speed regime ¨°°° km/h

Are there any permanent infrastructural measures that could affect speed behaviour?

☐ Speed ramps ☐ Road axis shifts ☐ Flower boxes ☐ Other

Are there any temporary interventions that could affect speed behaviour?

☐ Parking cars ☐ Bus lanes ☐ Road works ☐ Other

Is there some enforcement in the proximity of the site?

☐ Yes ☐ No

Was it possible to respect the following conditions of measurement?

- ❖ Site not on bend but on straight road segment ☐ Yes ☐ No
- ❖ Site on a small gradient road (< 3%) ☐ Yes ☐ No
- ❖ Site at (minimum) 500 m of crossing (junction, crossroads) ☐ Yes ☐ No
- ❖ Site at (minimum) 500 m of traffic calming device ☐ Yes ☐ No
- ❖ Site at (minimum) 500 m of road works ☐ Yes ☐ No
- ❖ Site at (minimum) 1000 m of speed regime change ☐ Yes ☐ No
- ❖ Site away from parking zone ☐ Yes ☐ No

Note °°°

Code of annexed picture: °°°

Note:

Annexes to Chapter 5 on Protective systems

Annex A: Recommendations on designing survey - sampling

Given the complexity of statistical sampling for the observation of protective systems' use in road traffic, it is recommended that an experienced statistician assists in setting up the national system. The following recommendations are based on the experience of several Member states, which regularly carry out surveys on the protective systems' use.

Each category of road users defined in the Manual (see Chapter 5) represents, from the statistical point of view a **statistical population**, i.e. a set of entities concerning which statistical inferences are to be drawn, based on a random sample taken from the population. For example, if we are interested in helmet wearing among motorcyclists then we should describe the set of motorcyclists that is of interest.

Survey sampling is a random selection of a sample (individual observations) from the finite population (all road users in the category considered).

The most elementary methodology is called simple random sampling. Most of the theory of statistics assumes this kind of sampling unless otherwise noted. It assures that the same probability of selection is given to every possible subset of the population which has the desired sample size. Because simple random sampling is a fair way to select a sample, it is reasonable to generalize the results from the sample back to the population. Statistically, the simple random sampling is not the most efficient method of sampling and one may, just because of luck of the draw, not get a good representation of all subgroups of the population. It is, however, a method, which is easily understood and performed in practice. It requires **assuming a rate measured at an observational site to be a sampling unit**.

If the observational sites are attributed to road types according to the proportion of relevant traffic, which they carry, then one can yield the final value directly as an average of the rates measured at all observational sites. One disadvantage of such method is that it will in turn require a relatively large number of observational sites. Assuming the worst case, when the indicator estimated is 50%, i.e. the proportion p=0.5, sampling error of 5% and a confidence level of 95%, 384 sample units seem a minimum in the safest option (Table 1). Drawing conclusions about strata, such as road types, will require a further increase in the number of observation sites.

Considering simple random sampling, the formulas of classical statistics can be applied in order to determine a confidence interval for the population proportion:

$$p - z\sqrt{\frac{p(1-p)}{n}} < \pi < p + z\sqrt{\frac{p(1-p)}{n}},$$

with *p* - the survey proportion, *n* - the sample size and z - the z-value of the desired confidence interval. Based on this calculation, one can determine a maximal deviation *(md)*, with which the survey proportion *p* can deviate from the population proportion π:

$$md \geq z\sqrt{\frac{p(1-p)}{n}},$$

Isolating *n* leaves us: $n \geq \frac{z^2 p(1-p)}{md^2}$.

Based on this formula, we can observe that the sample size *n* depends on the survey proportion *p*, the accuracy with which we want to draw conclusions via the value of *z* and the accuracy itself via *md*. If we take the most ‘safe' case, i.e. p=0.5, then the following tables give possible sample sizes for some classical values of confidence and maximal deviation:

p=0.5		Accuracy/Sampling error (*md*)			
Confidence level	**z-value**	**0.1**	**0.05**	**0.02**	**0.01**
0.90	**1.65**	68	271	1 691	6 764
0.95	**1.96**	96	384	2 401	9 604
0.98	**2.33**	135	541	3 382	13 530
0.99	**2.58**	166	663	4 147	16 587

p=0.8 (or p=0.2)		Accuracy/Sampling error (*md*)			
Confidence level	**z-value**	**0.1**	**0.05**	**0.02**	**0.01**
0.90	**1.65**	44	175	1089	4356
0.95	**1.96**	62	246	1537	6147
0.98	**2.33**	87	348	2172	8687
0.99	**2.58**	107	427	2663	10651

Table 1 a,b: Minimal sample in function of confidence level and sampling error

To deal with the issues of low accuracy by a limited number of observational sites, we have to turn to other sampling methods. Sophisticated sampling techniques that are both economical and scientifically reliable have been developed and should therefore be used here. The above series of formulas is not valid in the case of non-simple-random samples (such as clustered samples/stratified samples) and there is no universal formula suitable for all situations. The use of statistical software is necessary.

Stratified Random Sampling, also sometimes called proportional or quota random sampling, involves dividing the individuals into homogeneous subgroups and then taking a simple random sample in each subgroup. In more formal terms: the individuals are divided into non-overlapping groups (i.e., strata) N_1, N_2, N_3, ... N_i, such that $N_1 + N_2 + N_3 + ... + N_i = N$. Then a simple random sample of f = n/N is applied in each strata. The strata should be mutually exclusive: every element in the population must be assigned to only one stratum. The strata should also be collectively exhaustive: no population element should be excluded. Then random sampling is applied within each stratum.

Stratified sampling allows representing not only the overall population, but also key subgroups of the population. If there is a need to cover and make a judgement about any subgroup, this may be the only effective way. If the subgroup is extremely small, then different sampling fractions (f) within the different strata to randomly over-sample the small groups can be used. When we use the same sampling fractions within strata we are conducting a proportionate stratified random sampling. When we use different sampling fractions in the strata, we call this a disproportionate stratified random sampling. Second, a stratified random sampling will generally have more statistical precision than a simple random sampling. This will only be true if the strata or groups are homogeneous. If they are, then we expect that the variability within-groups is lower than the variability for the population as a whole. Stratified sampling capitalizes on that fact.

For the observational survey of protective systems' use, we distinguish two kinds of stratification:

- **Prior stratification**, depending on the subgroups of interest, such as driver and front seat passenger in the case of SPI_A indicator, or the three different road types, which are to be considered before designing a survey.
- **Post stratification**, allowing rough conclusions to be drawn based on subgroups of the population observed, such as gender, road type, etc.

Following is a description of a sample design that meets the guidelines and presents an example of a reasonably accurate and practical design. Depending on the data available in a country, substitutions in this design can be made without a loss of accuracy. This information is intended to serve only as an example of a complying survey design and to provide guidance for countries concerning the recommended design options. These are not the design requirements. It is strongly recommended that national surveys of protection systems' use to be designed by qualified survey statisticians.

SAMPLE DESIGN

A. In countries with known regional differences in the use of protective systems, the survey should be designed in such a manner that the regions are assumed to be independent strata. Ideally, a national survey is designed covering all regional units but if the absolute differences in the indicator's value found do not exceed 5%, and the standard deviation does not exceed 5%, the survey can be carried out on a limited number of randomly chosen regions.

It is straightforward to assume that no significant regional differences are presented in some Member states, such as the Baltic countries, Malta, or Cyprus, because these countries exhibit a high degree of homogeneity in terms of demographic and other characteristics, which have an influence of the rate of protective systems' use. (This is also the case of the U.K., where only subtle regional disparities in the protective systems' use have been recorded.) There are, however, other countries, which require taking regional heterogeneities into account when designing a survey for country level.

Example:

Three (linguistic) regions having different social and cultural background can be identified in Switzerland. A more detailed survey on seat belt wearing rates by passenger car occupants on front seats unveiled significant differences in the assessed values. So the A indicator values were in 2005 as follows: Deutschweiz 85%, Westschweiz 77% and Tessin 55%. The average national value (weighted by population and traffic exposure) is 82%. So the standard deviation of the three regional values is 11,55>5%. Therefore, the first stage of multi-stage sampling must include the three strata (regions), for which the random sampling procedure is applied.

B. Regions are the best candidates for prior stratification. In large countries with differing geographic areas, it is recommended that stratification of sampling units by geographic region be employed prior to sampling units' selection. Regions should be randomly selected, preferably with the probabilities proportional to the vehicle kilometres travelled (VKT) in each region. If the VKT is not available by region, sampling units can also be selected with the probabilities proportional to region's population. Selecting the sampling units, countries should ensure that an adequate mix of rural and urban regions is represented.

C. Within sampled sampling units, it is suggested that road profiles will be stratified by road type. For example, the design strata might be rural, urban roads and motorways as the three major types of roads presented in the country. The sample should be allocated to these strata by estimated annual traffic performance in each stratum. The sample of road segments within a stratum should be selected with a probability proportional to average daily traffic.

Table 3 shows the recommended strata to be considered in prior sampling process at country level. (Strata numerated with indices as XY, for different road types and regions.)

	Road type 1	Road type 2	…	…	Road type n
Region 1	N_{11}	N_{12}	…	…	N_{1n}
Region 2	N_{21}	N_{22}	…	…	N_{2n}
…	…	…	…	…	…
…	…	…	…	…	…
Region n	N_{n1}	N_{n2}	…	…	N_{nn}

Table 3: Possible prior strata to be considered when designing survey

If the accurate estimates of the indicator are to be determined for any other sub-group, then new strata are to be considered, which will basically augment the total number of strata geometrically, that will in turn lead to a higher number of observed individuals.

D. *Sample size:* the following rules provide rough guidelines for determining the sample size for estimating protective systems' use rate with the required level of precision.

(1) The sample should include approximately half of all considered regions (fulfilling the requirements on homogeneity specified in clause A).
(2) To attain required accuracy of estimates, 30 observation sites on which at least 30 individuals are observed, should be considered for each particular stratum.
(3) Ideally, some 50 observational sites should be considered and attributed to different road types according to the share of relevant traffic carried by them. The rule proposed by Tanner (1962) can be applied here: Number of points/km of road (of that sample) ≈ (average flow on that type of road) x with x ∈ [1/2;1] AND the minimum number of points must be equal to 3.
(4) At each location at least 100 (50) road users should be observed.

For each stratum, at least 3 observation points should be attributed (Tanner, 1962), i.e. at least 3 points at each road type in each region considered.

If a minimum of 50 and a sample of 100 road users are surveyed at each location point, this would lead to a total sample size of at least 1 500 within each stratum, if considering 3 road types and an even distribution of observation points across road types. At the national level, this strategy would already provide a sampling error of 2% given a confidence level of 95%, if the axiom of simple random sampling is considered.

The optimum number of observed individuals at each location depends on the number of locations and targeted accuracy of estimates for both strata and overall indicator values.

Example:

To achieve the required level of precision on SPI-A indicator, a country with 10 regions would cover five of them, but not as prior strata. Assuming the following traffic performance proportions for passenger cars and vans: 50/40/10% on rural roads/urban roads/motorways, altogether 100 sites will be distributed on the roads as follows: 50 on rural roads, 40 on urban roads and 10 on motorways. At each location, 100 individuals will be observed at times randomly chosen across the measurement period. Formulas for simple random sampling can be used here and conclusions can be drawn for all the three road types with sufficient precision (except motorways).

SPI computation

Assuming simple random sampling within considered strata and not accounting for the design effects, one can estimate the SPI values and their confidence intervals (CI) by applying the following template.

a. The value of the SPI can be computed for each stratum considered (coming from either prior or post stratification) applying the following formula:

$$SPI\_X = \frac{N_{use}}{N_{total}} \cdot 100$$

where:

***N**$_{use}$* is the number of road users using protective systems in road traffic, as observed during the roadside survey in order to determine a nationally representative SPI_X value,

***N**$_{total}$* is the sample size, meaning the total number of road users observed during the roadside survey in order to determine a nationally representative SPI_X value.

The 95%-confidence interval of the above estimate will be:

$$[p - z\sqrt{\frac{p(1-p)}{N}}; p + z\sqrt{\frac{p(1-p)}{N}}]$$

or

$$[p - 1.96SE; p + 1.96SE]$$

with $SE = \sqrt{\frac{p(1-p)}{N}}$,

where:

SE stands for standard error;

p is the proportion estimate, i.e. the SPI value divided by 100,

z is a z-value for considered confidence interval, here for 95%CI *z=1.96* and

N is the number of observed individuals N_{totalx}.

b. The value of the SPI can be further computed for any combination of strata, applying rules for aggregation as follows:

$$SPI\_X = \sum_{i=1}^{n} \lambda_i SPI\_X_i$$

where:

SPI_i are SPIs of different strata (e.g. road categories or regions),

λ_i designates the portion of total exposure to traffic attributable to individuals travelling on different road categories (or in different regions).

The 95%-confidence interval of the above estimate will be:

$$p - 1.96\sqrt{\sum_{i}^{n} \lambda_i^2 \cdot SE_i^{\ 2}}; p + 1.96\sqrt{\sum_{i}^{n} \lambda_i^2 \cdot SE_i^{\ 2}} .$$

$$w = \sum_{i}^{n} \lambda_i^2 \cdot SE_i^{\ 2}$$

As an example, the computation of SPI_A values for different strata is shown for Switzerland. Two ways of aggregation are considered, using different regions and different road types.

Region	Road category	Loc	N	N_{use}	SPI_A	p	z*SE	TS	λi	p* λi	w	w(total)	CI 95%	
German	urban	16	8036	6671	83.0	0.830	0.008	0.239	0.319	0.265	1.78E-06	1.00E-06	0.822	0.838
	rural	10	4733	4218	89.1	0.891	0.009	0.259	0.345	0.308	2.44E-06	1.37E-06	0.882	0.900
	motorway	8	4016	3680	91.6	0.916	0.009	0.252	0.336	0.308	2.16E-06	1.21E-06	0.908	0.925
	all roads	34	16785	14569	-	-	-	0.750	p	0.880	6.38E-06	3.59E-06	0.875	0.885
French	urban	6	3034	2102	69.3	0.693	0.016	0.064	0.320	0.222	7.18E-06	2.87E-07	0.676	0.709
	rural	6	3014	2422	80.4	0.804	0.014	0.069	0.345	0.277	6.23E-06	2.49E-07	0.789	0.818
	motorway	4	2000	1743	87.2	0.872	0.015	0.067	0.335	0.292	6.28E-06	2.51E-07	0.857	0.886
	all roads	16	8048	6267	-	-	-	0.200	p	0.791	1.97E-05	7.88E-07	0.782	0.800
Italian	urban	3	1513	892	59.0	0.590	0.025	0.016	0.320	0.189	1.64E-05	4.09E-08	0.565	0.614
	rural	2	1008	786	78.0	0.780	0.026	0.017	0.340	0.265	1.97E-05	4.92E-08	0.754	0.805
	motorway	2	1000	867	86.7	0.867	0.021	0.017	0.340	0.295	1.33E-05	3.33E-08	0.846	0.888
	all roads	7	3521	2545	-	-	-	0.050	p	0.749	4.94E-05	1.24E-07	0.735	0.762
Tota		57	28354	23381						0.856		4.50E-06	0.852	0.860

Region	Road category	Loc	N	N_{use}	SPI_A	p	z*SE	TS	λi	p* λi	w	w(total)	CI 95%	
Urban	German	16	8036	6671	83.0	0.830	0.008	0.239	0.749	0.622	9.85E-06	1.00E-06	0.822	0.838
	French	6	3034	2102	69.3	0.693	0.016	0.064	0.201	0.139	2.82E-06	2.87E-07	0.676	0.709
	Italian	3	1513	892	59.0	0.590	0.025	0.016	0.050	0.030	4.02E-07	4.09E-08	0.565	0.614
	Country total	25	12583	9665	-	-	-	0.319	p	0.791	1.31E-05	1.33E-06	0.783	0.798
Rural	German	10	4733	4218	89.1	0.891	0.009	0.259	0.751	0.669	1.15E-05	1.37E-06	0.882	0.900
	French	6	3014	2422	80.4	0.804	0.014	0.069	0.200	0.161	2.09E-06	2.49E-07	0.789	0.818
	Italian	2	1008	786	78.0	0.780	0.026	0.017	0.049	0.038	4.14E-07	4.92E-08	0.754	0.805
	Country total	18	8755	7426	-	-	-	0.345	p	0.868	1.41E-05	1.67E-06	0.861	0.876
Highways	German	8	4016	3680	91.6	0.916	0.009	0.252	0.750	0.687	1.07E-05	1.21E-06	0.908	0.925
	French	4	2000	1743	87.2	0.872	0.015	0.067	0.199	0.174	2.23E-06	2.51E-07	0.857	0.886
	Italian	2	1000	867	86.7	0.867	0.021	0.017	0.051	0.044	2.95E-07	3.33E-08	0.846	0.888
	Country total	14	7016	6290	-	-	-	0.336	p	0.905	1.33E-05	1.50E-06	0.898	0.912
Tota		57	28354	23381						0.856		4.50E-06	0.852	0.860

Annex B: Examples of observation forms

B.1: Observation form for assessing seat belt wearing by front seats occupants of light vehicles (SPI-A)

Survey on the use of seat belts by front seats occupants of light vehicles Nr: ....

Observation site:	[Code, Name]		Observer:	[ID, Name]		
Date:	[DD.MM.YYYY]		Weather:	☐ sunny	☐ rainy	☐ dull
Time:	From:	[XX:XX]	Temperature:	[XX] °C		
	To:	[XX:XX]	Remarks:	…………………………………………………….		
				…………………………………………………….		

Vehicle Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Driver																									
yes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
no	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Front passenger																									
yes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
no	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Vehicle Nr.	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Driver																									
yes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
no	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Front passenger																									
yes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
no	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Vehicle Nr.	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Driver																									
yes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
no	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Front passenger																									
yes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐								
no	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐												

etc.

Instructions recall: XXX vehicles to be observed in total
Taxi, emergecy vehicles, police cars, armed forces vehicles to be excluded
Observation time must fall between 8 and 18 o'clock
If police controls on the same road, observations must be stopped

B.2: Observation form for assessing helmet use by pedal cyclists (SPI-F)

Survey on the use of helmets by pedal cyclists [Year, Period, Purpose] Nr: ………

Observation site:	[Code, Name]		Observer:	[ID, Name]
Date:	[DD.MM.YYYY]		Weather:	☐ sunny ☐ rainy ☐ dull
Time:	From:	[XX:XX]	Temperature:	[XX] °C
	To:	[XX:XX]	Remarks:	………………………………………………………
				………………………………………………………

Person Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Helmet																									
yes, correct use	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
yes, but misuse	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
no, but helmet present	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
no	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Gloves																									
yes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
no	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sportswear																									
yes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
no	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
dont know	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Bicycle type																									
standard/city/trekking	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
mountain-bike	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
racing-bike	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
childrens bike	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Age																									
< 14	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
15–29	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
30–44	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
45–59	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
60+	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sex																									
female	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
male	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Person Nr.	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Helmet																									
yes, correct use	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
yes, but misuse	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐				
no, but helmet present	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐								
no	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐												

etc.

Appendix

Introduction: comments on the statistical report attached

GITELMAN[A], V.

[A]TECHNION

1. Background

The report presented below was prepared by a statistician from the Technion Statistical Laboratory - Dr. Etti Doveh. The issues to be explored were: what is a correct evaluation of an indicator and its confidence interval, where the indicator comes from an observation survey and presents a share of cases with a certain feature (e.g. vehicles using DRL out of the total number of cases).

As described in the SPI Manual, a method accepted for estimating SPIs in the areas of protective systems and DRL, applies to an observation survey, with a stratified sample. The SPI estimation includes two steps: (a) estimating an SPI value per stratum, using the data collected at 10-30 observation sites (with 30-100 measurements per site), and (b) weighting the stratum values to provide a final estimate, with its confidence interval.

The procedure described in the SPI Manual can briefly be presented as follows:

SPI computation

The value of the SPI can be computed for each stratum considered applying the following formula:

$$SPI\_X = \frac{N_{use}}{N_{total}} \cdot 100$$

where:

N_{use} is the number of road users with the feature considered (e.g. using a protective system in road traffic), as observed during the roadside survey,

N_{total} is the sample size, meaning the total number of road users observed during the roadside survey.

The 95%-confidence interval (CI) of the above estimate will be:

$$[p - z\sqrt{\frac{p(1-p)}{N}};\, p + z\sqrt{\frac{p(1-p)}{N}}]$$

or

$$[p - z \cdot SE;\, p + z \cdot SE],$$

$$SE = \sqrt{\frac{p(1-p)}{N}}$$

where: SE stands for standard error;

p is the population proportion estimate, p=SPI_X/100;

z is the z-value for a considered confidence interval (for 95%CI z=1.96),

and *N* is the number of observed individuals N_{total}.

The value of the SPI can be further computed for any combination of strata, as follows:

$$SPI\_X = \sum_{i=1}^{n} \lambda_i SPI\_X_i$$

where:

SPI_i are the SPIs of different strata (e.g. road categories), *i=1..n*,

n – the number of strata,

λ_i designates the portion of total exposure (e.g. traffic volumes) attributable to *i* stratum.

The 95%-confidence interval of the above estimate will be:

$$p - 1.96\sqrt{\sum_i^n \lambda_i^2 \cdot SE_i^2}\,;\, p + 1.96\sqrt{\sum_i^n \lambda_i^2 SE_i^2}\ .$$

As can be seen from the above procedure, estimating an SPI value per stratum we do not apply the number of sites which served for the data collection. On the other hand, the SPI Manual recommends a certain range for the number of sites (per stratum) for the survey's performance. Thus, it would be useful to demonstrate how the number of sites (per stratum) influences the accuracy of the final estimate.

2. A practical summary of the report

The report provides answers and comments on the following questions:

1) How to evaluate the confidence interval of an estimate (SPI) for one site? – see Section 3.

Due to the small *n* (number of measurements per site) the assumption of an approximately normal distribution of the sample proportion may be incorrect. Hence, estimating the variance of the value, a correction should be applied.

The practical recommendations are:

- When the expected proportion (SPI per site) lies in the range of 30%-70%, for estimating confidence internal use **formula 1**;

- When the expected proportion (SPI per site) is higher or lower than the range mentioned above, for estimating upper (P_U) and lower (P_L) limits use **formulae 2**.

Besides, in Section 4 a formula for estimating **the required number of measurements per site** is presented. Section 5 provides a method for comparing the (SPI) values estimated **for two sites**, including a way for estimating the required number of measurements for such a comparison.

2) How to estimate a confidence interval of the value (SPI) per stratum? – see Section 6.

Three methods are compared:

1 – The so-called "Normal approximation" when the P (final proportion) is weighted through the sites – **formula 3**;

2 – The "Ad hoc Wilson method" which is known as an alternative to the normal approximation – **formula 4**;

3 – "A simplified approach" where, per stratum, mixing all the data and counting the number of cases with a feature out of the total cases, is applied – **formula 5**.

Simulations of observation results were performed for the cases with 5 and 30 sites, 30-50-100 measurements per site and two ranges of proportions (SPI values per site), i.e. 0.925-0.975 and 0.6-0.65. Per each case defined, 10000 runs were performed.

The comparison refers to the three methods of estimation of confidence interval. It considered the proportion of runs where the true P (SPI per stratum) was inside the confidence interval estimated by a method.

In the results (see Section 6.3) one should look at the values:

p_cover_norn for method 1;

p_cover_WIL for method 2;

p_cover_all for method 3.

It can be seen that the proportions of "correct answers" in some cases are slightly lower for method 3 (a simplified approach) than for other methods (mostly, when the weights of sites are not equal). Strictly based on the simulation results, the statistician recommends to apply method 2 (Ad hoc Wilson method) as the most accurate one in the majority of cases considered.

At the same time, one can note that the difference in the estimates provided by the "worst method" (a simplified approach) in comparison with the "best method" (Ad hoc Wilson) is not large: on average, the accuracy of estimate (a probability that the true value will be inside the estimated confidence interval) is 0.90 versus 0.95, accordingly.

As soon as equal weights are used, or if the number of sites or the sample size per site are enlarged, the results provided by the "Normal approximation" are usually very close to the results provided by Ad hoc Wilson method.

Therefore, for the estimation of the SPI confidence interval per stratum, the following summary recommendations can be drawn:

1) "A simplified approach" (the one introduced in Chapters 5, 6 of the Manual) is suitable for application when **equal weights** for different sites in the same stratum are considered (and the number of observations per each site is similar).

2) When the weights of sites are **not equal** (and/or the number of observations varies between the sites), the "Normal approximation" method should be applied (a calculation example in provided in Chapter 6 of the Manual; the formulae can be found in Section 6 of this Appendix).

3) **In any case**, the calculation can be performed using the Ad hoc Wilson method (see formulae in Section 6 of this Appendix).

Statistical Report on:

Statistical issues in estimating SPIs and their confidence intervals

DOVEH[A], E.

[A]TECHNION

Table of content page

1. Problem definition

In this document we consider the problem of estimating proportions and appropriate confidence intervals for these proportions, for a specific sampling paradigm that was presented to us. Let us assume our target is to estimate the proportion of a given population, which belongs to class C. We assume that the rest of the population belongs to C', which is the complement of C.

The sampling program is:

- The researchers choose H specific road sites.
- The collection of these road sites represents the population, according to the impression of the researchers, maybe after some weighting.
- n_h units are sampled from the *h*-th road site. Actually the intention is to sample the same sample size from each road site, but there are dropouts.
- Our assumption is that road sites are independent

We translate this information as follows.

Let:

P_h = the true proportion of the population of site h, which belong to C

W_h = the weight of site *h*. We assume that the sum of W_h over all the H strata equals 1.

Then the true population proportion that we want to estimate is: $\sum_{h=1}^{h=H} W_h P_h$

We estimate it by: $\sum_{h=1}^{h=H} W_h p_h$, where p_h is the sample proportion from the *h* site.

Note: if each stratum *h* has l_h substrata with weights $W_{h1},\ldots,W_{hl_h}$, which sum to 1, then, estimating P can be done as above, by forming a new set of weights, by multiplying the substratum weights by the corresponding main stratum weight.

2. Contents of this work

In the original proposal (that was presented to us) it was suggested to sample 30 individuals per site and 30 sites in order to estimate the population proportion. Here we will discuss these numbers. First we will demonstrate one site proportion estimate, confidence interval and needed sample size. Then comparing proportions of two sites will be discussed and only after that the stratified sampling will be discussed.

Comment:

In the discussion below we show which *n* we should use to assure normality of each estimate from each site. But each site has different proportion (and hence different variance also). In order to perform t-test between two populations (e.g. two countries – each one with its own sites) we have to assume that each one of the sites from the same country has the same expected mean and variance, **which is not the case in our problem.** On the contrary, we use stratification, since it is assumed that the expected values are different for each chosen site. Moreover, we do not use randomly sampled sites from the population of sites, but choose them according to their characteristics.

3. Confidence intervals for a single site proportion and when and how the Normal approximation, for single site proportion estimation, may be used?

This paragraph uses Cochran 1977 pp. 57-58 and Fleiss et al pp 26-29.

Let P be the true site proportion, and *p* its estimate in a sample of size *n*. When *n* is large, the central limit theorem implies that the sample proportion is approximately normally distributed with mean P and variance (P*(1-P))/n.

In order to bring the normal curve tail probabilities into closer agreement with binomial tail probabilities we correct for continuity with $\frac{1}{2n}$. Hence, we use $z = \frac{|p-P| - \frac{1}{2n}}{\sqrt{\frac{P(1-P)}{n}}}$ to get confidence intervals for P and define an approximate 100*(1-α)% confidence interval as all P satisfying $z < z_{1-\frac{\alpha}{2}}$.

($z_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quantile of standard normal distribution)

3.1. Traditional normal approximation

Actually, $z = \frac{|p-P| - \frac{1}{2n}}{\sqrt{\frac{P(1-P)}{n}}} = \frac{|p-P| - \frac{1}{2n}}{\sqrt{V(p)}}$. If we replace V(p) by its unbiased estimator $\frac{p(1-p)}{n-1}$, we get that the normal approximation to the confidence limits for a single P is:

$$p \pm \left(z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n-1}} + \frac{1}{2n} \right)$$

Formula 1

Comment:

Let y_i=1 if *i*-th unit sampled=1 and 0 otherwise. *n* is the sample size.

The expected value of y_i is P and its variance is P(1-P).

$$E\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{1}{n} \sum_{i=1}^{n} E(y_i) = P$$

$$V\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{1}{n^2} \sum_{i=1}^{n} V(y_i) = \frac{1}{n} V(y)$$

We estimate $V(y)$ by its unbiased estimator:

$$\hat{V}(y) = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{n} y_i(y_i - \bar{y})}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} y_i - n\bar{y}^2}{n-1} = \frac{n\bar{y} - n\bar{y}^2}{n-1} = \frac{n\bar{y}(1-\bar{y})}{n-1} = \frac{np(1-p)}{n-1}$$

since: $V\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{1}{n}V(y)$

We can calculate its unbiased estimator as: $\hat{V}\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{p(1-p)}{n-1}$.

The rules in the following table are constructed so that with 95% confidence limits the true frequency with which the limits fail to enclose P (the population proportion) is not greater than 5.5%.

P	*np* = Number Observed in the Smaller Class	*n* = Sample Size
0.5	15	30
0.4	20	50
0.3	24	80
0.2	40	200
0.1	60	600
0.05	70	1400
~0*	80	∞

*This means that *p* is extremely small, so that *np* follows the Poisson distribution.

This table demonstrates that we need a large sample size for small probability (of the smallest group). The confidence intervals which we used in this section suit for 0.3≤P≤0.7.

Note: the table gives an idea about the sample size needed within a road site in order to achieve normality.

3.2. Better confidence intervals

Solving for *P* which satisfies $\frac{|p-P| - \frac{1}{2n}}{\sqrt{\frac{P(1-P)}{n}}} < z_{1-\alpha/2}$ without substituting *p* instead of *P* in the standard deviation gives:

$$P_L = \frac{(2np + z^2_{1-\alpha/2} - 1) - z_{1-\alpha/2}\sqrt{z^2_{1-\alpha/2} - \left\{2 + \frac{1}{n}\right\} + 4p(nq+1)}}{2\left(n + z^2_{1-\alpha/2}\right)}$$

Formulae 2

$$P_U = \frac{(2np + z^2_{1-\alpha/2} + 1) + z_{1-\alpha/2}\sqrt{z^2_{1-\alpha/2} + \left\{2 - \frac{1}{n}\right\} + 4p(nq+1)}}{2\left(n + z^2_{1-\alpha/2}\right)}$$

4. Sample size for a single site proportion

The first consideration - normality achievement, is discussed above.

Here we will discuss sample size for confidence intervals. The width of the confidence interval, given by the above equations (P_L and P_U) is not greater than:

$$\frac{1 + z_{1-\alpha/2}\sqrt{z^2_{1-\alpha/2} + 2 + nk}}{\left(n + z^2_{1-\alpha/2}\right)}$$ where k=4pq.

If we require that the latter will be smaller than a given size *d*[20], we can solve for *n*:

$$n \geq \frac{kz^2_{1-\alpha/2}}{d^2} + \frac{2}{d} - 2z^2_{1-\alpha/2} + \frac{z_{1-\alpha/2} + 2}{k}$$

Since k=4pq ≤ 1 an upper bound to *n* can be achieved by setting 4pq=1. If we have some information about *P*, or assumed it to lie in a given interval, we select *k* to be:

If P satisfies	Then use
0≤P<d/2	k=4d(1-d)
d/2≤P<0.3	k=4(P+d/2)(Q-d/2)
0.3≤P≤0.7	1
0.7≤P≤1-d/2	k=4(P-d/2)(Q+d/2)
1-d/2< P≤1	k=4d(1-d)

5. Difference between two sites' proportions – confidence intervals and sample sizes

Let:

P_1 = the true proportion for site 1

[20] *d* defines the width of confidence interval. For example if we require a 2% accuracy of estimate then d=0.04.

P_2 = the true proportion for site 2

p_1 = the sample proportion for site 1

p_2 = the sample proportion for site 2

n_1 = the sample size for site 1

n_2 = the sample size for site 2

5.1. Confidence interval

An estimate of the standard error of (p_1- p_2) is:

$$\hat{s}e(p_1 - p_2) = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

If both n_1 and n_2 are big enough then (as discussed above) the 100*(1-α)% confidence interval for $P_1 - P_2$ is given by:

$$Low = (p_1 - p_2) - z_{1-\alpha/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} - \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

$$High = (p_1 - p_2) + z_{1-\alpha/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} + \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

5.2. Hypothesis testing

Here the hypothesis being tested is whether the proportions in the two sites from which we sampled, say P1 and P2, are equal.

We test the difference using the following test statistics:

$$z = \frac{|p_1 - p_2| - \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ where $\bar{p} = \frac{p_1 + p_2}{2}, \bar{q} = 1 - \bar{p}$.

If z exceeds the standard normal 1- α/2 quantile, P1 and P2 are inferred to be unequal.

5.3. Sample size for hypothesis testing

We assume that the sample sizes from the two sites being compared are equal to a common *n* (i.e. the total sample size from both samples is *2n*).

We find the value for the common sample size *n* so that:

(1) if in fact there is no difference between the two underlying proportions, then the chance is approximately α of falsely declaring that the two underlying proportions are different;

(2) if in fact the proportions are P1 and P2≠P1 then the chance is approximately 1-β of correctly declaring the two proportions are different.

The investigator has to hypothesize P1 and P2.

Let $\overline{P} = \frac{P_1 + P_2}{2}$ and $z_{1-\beta}$ is the 1-β standard normal percentile then according to Fleiss (2003), pp 72, the following sample size *n* required in each group provides an excellent degree of approximation:

$$n = \frac{n'}{4}\left(1 + \sqrt{1 + \frac{4}{n'|P_2 - P_1|}}\right)^2$$

where:

$$n' = \frac{\left(z_{1-\frac{\alpha}{2}}\sqrt{2\overline{P}\,\overline{Q}} + z_{1-\beta}\sqrt{P_1Q_1 + P_2Q_2}\right)^2}{\left(P_2 - P_1\right)^2}$$

6. Stratified sampling

Let:

P_h = the true proportion of the population of site h, which belongs to C;

W_h = the weight of site *h*.

Then the true population proportion that we want to estimate is: $P = \sum_{h=1}^{h=H} W_h P_h$.

We estimate it by: $\hat{P} = \sum_{h=1}^{h=H} W_h p_h$, where p_h is the sample proportion of the *h* site.

The sample variance (assuming all sites are independent) of $\hat{P}$ is:
$V(\hat{P}) = V\left(\sum_{h=1}^{h=H} W_h p_h\right) = \sum_{h=1}^{h=H} W_h^2 V(p_h) = \sum_{h=1}^{h=H} W_h^2 \frac{P_h(1-P_h)}{n_h}$

The variance above can be estimated in two ways (which give approximately the same results for large enough n_h).

- Using unbiased variance estimators for each stratum

$$\hat{V}(\hat{P}) = \hat{V}\left(\sum_{h=1}^{h=H} W_h p_h\right) = \sum_{h=1}^{h=H} W_h^2 \hat{V}(p_h) = \sum_{h=1}^{h=H} W_h^2 \frac{p_h(1-p_h)}{n_h - 1}$$

- Plug-in estimators, p_h, for P_h

$$\hat{V}(\hat{P}) = \hat{V}\left(\sum_{h=1}^{h=H} W_h p_h\right) = \sum_{h=1}^{h=H} W_h^2 \hat{V}(p_h) =$$

$$\sum_{h=1}^{h=H} W_h^2 \frac{\hat{P}_h(1-\hat{P}_h)}{n_h} = \sum_{h=1}^{h=H} W_h^2 \frac{p_h(1-p_h)}{n_h}$$

If n_h is equal for all *h* and W_h=1/H for all *h* then: $P = \sum_{h=1}^{h=H} \frac{1}{H} P_h = \frac{1}{H} \sum_{h=1}^{h=H} P_h$

And its estimator is:

$$p = \frac{1}{H}\sum_{h=1}^{h=H} p_h = \frac{1}{H}\sum_{h=1}^{h=H}\frac{x_h}{n_h} = \frac{1}{H}\sum_{h=1}^{h=H}\frac{x_h}{n} = \frac{1}{nH}\sum_{h=1}^{h=H} x_h =$$

$$\frac{total\ number\ of\ successes}{total\ number\ of\ units\ sampled}$$

$$where$$

$$x_h = number\ of\ successes\ in\ site\ \ h\ sample$$

And:

$$V\left(\sum_{h=1}^{h=H} W_h p_h\right) = \frac{1}{H^2}\sum_{h=1}^{h=H}\frac{P_h(1-P_h)}{n}$$

Let y_{ih}=1 if *i*-th unit of the *h*-th site succeeded and 0 otherwise. If each n_h is large enough, such that we can use the normal approximation, we get:

$$p_h = \frac{\sum_{i=1}^{n_h} y_{ih}}{n_h} \approx N(P_h, \frac{P_h(1-P_h)}{n_h}) \Rightarrow \hat{P} \approx N(P, V(\hat{P}))$$

We can use this approximation for hypothesis testing and confidence intervals building. We will discuss confidence intervals building more thoroughly, since this subject is now discussed in statistical literature.

6.1. Methods for estimating confidence intervals

We will compare the following methods:

- Normal approximation:

$$Plow = \hat{P} - z_{1-\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{P})},$$

Formula 3

$$Phigh = \hat{P} + z_{1-\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{P})}$$

where $\hat{V}(\hat{P})$ is any one of the choices above of $\hat{V}(\hat{P})$ estimator.
We will refer to this estimate with the extension "norm".

- Ad hoc Wilson method -

  The effective sample size n* is defined as: $n* = \frac{\hat{P}(1-\hat{P})}{\hat{V}(\hat{P})}$

  The Ad hoc Wilson confidence intervals are:

$$\hat{P} + \frac{z^2_{1-\frac{\alpha}{2}}}{2n*}(1-2\hat{P}) \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{z^2_{1-\frac{\alpha}{2}}}{(2n*)^2} + \hat{V}(\hat{P})}$$

Formula 4

We will refer to this estimate with the extension "WIL".

- Normal approximation with $\hat{V}(\hat{P})$ estimated as $\frac{\hat{P}(1-\hat{P})}{Ntot}$, where *Ntot*=total number of units sampled in all strata, which is the estimator we were asked to check:

$$Plow = \hat{P} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{P})} \; .$$

$$Phigh = \hat{P} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{P})}$$

Formula 5

We will refer to this estimate with the extension "all".

To note, there are many methods, which are alternative to the normal approximation, but we took a simple one - Ad hoc Wilson method.

Further, we will look both at the coverage and width of confidence intervals estimated by the above three methods for some examples of P (true population proportion), weights and sample sizes (per site).

6.2. Comparing the confidence intervals estimated

In the simulations we used the plug-in $\frac{p\ (1-p\ )}{n}$ estimator as $\hat{V}(\hat{p})$.

- Sample size for each site is 30 -50-100 (n_h)
- Number of sites: 5 or 30 (H)
- Proportions per sites:
    - o sample from the range 0.925 – 0.975
    - o sample from the range 0.6 – 0.65

Note: The same sampled proportions are used in all the repetitions, i.e. we sample before the simulations, and use it during all the simulations.

- Weights:
    - o Equal
    - o Uniform 0.1 – 0.9 (divided by their sum)

Note: The sampled weights are used in all the repetitions, i.e. we sample before the simulations, and use it during all the simulations.

- Each combination of sample size *with* the number of sites *with* the proportions per site and *with* certain weights runs 10000 times.
- For each combination we calculate confidence intervals by each of the three methods. Then for each combination we calculate: proportion of runs where the true *P* was inside the confidence interval estimated (coverage rate), and the average confidence interval width.

The results are presented in the next section.

6.3. Results

6.3.1. 5 sites

========True P = 0.9580834 ==========

p= 0.9393789 0.9644153 0.9454488 0.9691509 0.9720234

w= 0.2 0.2 0.2 0.2 0.2

h (site sample size)= 30	h = 50	h = 100
p_cover_all= 0.9486 p_cover_norm= 0.9385 p_cover_WIL= 0.9639	p_cover_all= 0.9415 p_cover_norm= 0.9415 p_cover_WIL= 0.9454	p_cover_all= 0.9453 p_cover_norm= 0.9453 p_cover_WIL= 0.9437
wide_all= 0.062661 wide_norm= 0.061682 wide_WIL= 0.06667	wide_all= 0.049089 wide_norm= 0.048592 wide_WIL= 0.050923	wide_all= 0.034932 wide_norm= 0.034717 wide_WIL= 0.035546

========True P = 0.9541808 ==========

p= 0.9393789 0.9644153 0.9454488 0.9691509 0.9720234

w= 0.2752816 0.2360762 0.2614177 0.03932139 0.1879031

h (site sample size)= 30	h = 50	h = 100
p_cover_all= 0.8931 p_cover_norm= 0.9013 p_cover_WIL= 0.955	p_cover_all= 0.8996 p_cover_norm= 0.9238 p_cover_WIL= 0.9523	p_cover_all= 0.911 p_cover_norm= 0.9356 p_cover_WIL= 0.9512
wide_all= 0.06528 wide_norm= 0.071417 wide_WIL= 0.077898	wide_all= 0.051177 wide_norm= 0.056292 wide_WIL= 0.059344	wide_all= 0.03636 wide_norm= 0.040165 wide_WIL= 0.041255

========True P = 0.6247413 ==========

p= 0.6022778 0.6264053 0.644621 0.6275718 0.6228307

w= 0.2 0.2 0.2 0.2 0.2

h (site sample size)= 30	h = 50	h (site sample size)= 100
p_cover_all= 0.9493 p_cover_norm= 0.9469 p_cover_WIL= 0.957	p_cover_all= 0.9447 p_cover_norm= 0.9433 p_cover_WIL= 0.9455	p_cover_all= 0.9495 p_cover_norm= 0.9444 p_cover_WIL= 0.9445
wide_all= 0.15442 wide_norm= 0.15226 wide_WIL= 0.15428	wide_all= 0.11972 wide_norm= 0.1187 wide_WIL= 0.11966	wide_all= 0.084795 wide_norm= 0.084425 wide_WIL= 0.084768

```
========True P =  0.6238995 ==========

p= 0.6022778 0.6264053 0.644621 0.6275718 0.6228307

w= 0.2752816 0.2360762 0.2614177 0.03932139 0.1879031
```

h (site sample size)= 30	h = 50	h = 100
p_cover_all= 0.9236 p_cover_norm= 0.9424 p_cover_WIL= 0.954	p_cover_all= 0.9258 p_cover_norm= 0.9439 p_cover_WIL= 0.9503	p_cover_all= 0.9239 p_cover_norm= 0.9456 p_cover_WIL= 0.9486
wide_all= 0.15443 wide_norm= 0.16576 wide_WIL= 0.16836	wide_all= 0.1198 wide_norm= 0.12928 wide_WIL= 0.13051	wide_all= 0.0848 wide_norm= 0.091865 wide_WIL= 0.092306

6.3.2. 30 sites

```
========True P =  0.9529468 ==========

p= 0.9728417 0.9476667 0.9588785 0.9536317 0.9301462 0.9699912 0.9373044
0.927103 0.941396 0.9727252 0.969477 0.9596402 0.9570253 0.9747135
0.9577853 0.9604265 0.9522033 0.9547071 0.939458 0.9323557 0.9731512
0.970115 0.9595353 0.9647734 0.9262307 0.9488898 0.962923 0.9358204
0.940909 0.9365813

w= 0.033 0.033 0.033 0.033 0.033 0.033 0.033 0.033 0.033 0.033 0.033 0.033
0.033 0.033 0.033 0.033 0.033 0.033 0.033 0.033 0.033 0.033 0.033 0.033
0.033 0.033 0.033 0.033 0.033 0.033
```

h (site sample size)= 30	h = 50	h = 100
p_cover_all= 0.9507 p_cover_norm= 0.9507 p_cover_WIL= 0.9456	p_cover_all= 0.9443 p_cover_norm= 0.9443 p_cover_WIL= 0.9516	p_cover_all= 0.9459 p_cover_norm= 0.9459 p_cover_WIL= 0.9497
wide_all= 0.027572 wide_norm= 0.027057 wide_WIL= 0.027369	wide_all= 0.021402 wide_norm= 0.021142 wide_WIL= 0.021289	wide_all= 0.01514 wide_norm= 0.01503 wide_WIL= 0.015083

```
========True P =  0.952147 ==========

p= 0.9728417 0.9476667 0.9588785 0.9536317 0.9301462 0.9699912 0.9373044
0.927103 0.941396 0.9727252 0.969477 0.9596402 0.9570253 0.9747135
0.9577853 0.9604265 0.9522033 0.9547071 0.939458 0.9323557 0.9731512
0.970115 0.9595353 0.9647734 0.9262307 0.9488898 0.962923 0.9358204
0.940909 0.9365813

w= 0.01914609 0.0280056 0.04092917 0.02645122 0.01310004 0.02044982
0.04399618 0.0301043 0.05066117 0.01264120 0.03106105 0.06157685 0.05647821
0.05611306 0.01664596 0.01418518 0.0431666 0.02599179 0.04336943 0.02470788
0.0173471 0.05033377 0.01212695 0.03283 0.03531128 0.04022006 0.02539858
0.03404128 0.05988574 0.03372447
```

h (site sample size)= 30	h = 50	h = 100
p_cover_all= 0.926 p_cover_norm= 0.9405 p_cover_WIL= 0.948	p_cover_all= 0.924 p_cover_norm= 0.9433 p_cover_WIL= 0.9462	p_cover_all= 0.9271 p_cover_norm= 0.9469 p_cover_WIL= 0.9486
wide_all= 0.027777 wide_norm= 0.029706 wide_WIL= 0.030106	wide_all= 0.021568 wide_norm= 0.023244 wide_WIL= 0.023434	wide_all= 0.015264 wide_norm= 0.016528 wide_WIL= 0.016596

```
========True P =  0.6221979 ==========
```

p= 0.60714 0.6207273 0.6206862 0.6184423 0.6076222 0.6069403 0.6116517 0.6232981 0.6132986 0.6428914 0.6022916 0.62211 0.6399462 0.606095 0.6280474 0.6103266 0.6063766 0.6376654 0.6447523 0.6187231 0.6332558 0.604742 0.6191985 0.6137192 0.640732 0.6224258 0.6405032 0.6406195 0.6397171 0.6219916

w= 0.033

h (site sample size)= 30	h = 50	h = 100
p_cover_all= 0.9454 p_cover_norm= 0.9454 p_cover_WIL= 0.9499	p_cover_all= 0.9458 p_cover_norm= 0.9458 p_cover_WIL= 0.9458	p_cover_all= 0.9491 p_cover_norm= 0.9465 p_cover_WIL= 0.9491
wide_all= 0.063308 wide_norm= 0.062256 wide_WIL= 0.062393	wide_all= 0.049061 wide_norm= 0.048565 wide_WIL= 0.04863	wide_all= 0.034693 wide_norm= 0.034511 wide_WIL= 0.034535

```
========True P =  0.6223959 ==========
```

p= 0.60714 0.6207273 0.6206862 0.6184423 0.6076222 0.6069403 0.6116517 0.6232981 0.6132986 0.6428914 0.6022916 0.62211 0.6399462 0.606095 0.6280474 0.6103266 0.6063766 0.6376654 0.6447523 0.6187231 0.6332558 0.604742 0.6191985 0.6137192 0.640732 0.6224258 0.6405032 0.6406195 0.6397171 0.6219916

w= 0.01914609 0.0280056 0.04092917 0.02645122 0.01310004 0.02044982 0.04399618 0.0301043 0.05066117 0.01264120 0.03106105 0.06157685 0.05647821 0.05611306 0.01664596 0.01418518 0.0431666 0.02599179 0.04336943 0.02470788 0.0173471 0.05033377 0.01212695 0.03283 0.03531128 0.04022006 0.02539858 0.03404128 0.05988574 0.03372447

h (site sample size)= 30	h = 50	h = 100
p_cover_all= 0.9275 p_cover_norm= 0.9459 p_cover_WIL= 0.948	p_cover_all= 0.9294 p_cover_norm= 0.9479 p_cover_WIL= 0.9494	p_cover_all= 0.9275 p_cover_norm= 0.9464 p_cover_WIL= 0.947
wide_all= 0.063285 wide_norm= 0.06794 wide_WIL= 0.068117	wide_all= 0.049049 wide_norm= 0.053008 wide_WIL= 0.053093	wide_all= 0.034688 wide_norm= 0.037673 wide_WIL= 0.037703

6.4. Conclusion

As expected, the biggest gap between the expected coverage of 0.95 to the coverage attained in simulations was when weights are non-equal, sample size and number of sites is small, and the true proportion is close to 1. Coverage was as low as 0.9 instead of 0.95 when using the normal approximation, and was even worse with the third option that we were asked to check, but was 0.955 using the Ad hoc Wilson method.

7. References

1. Cochran, W.G. (1977), *Sampling Techniques*, Third Edition, New York: John Wiley & Sons, Inc.
2. Fleiss, J.L. et al. (2003), *Statistical Methods for Rates and Proportions*, Third Edition, John Wiley & Sons, Inc.
3. Amang Sukasih and Donsig Jang (2005), "An Application of Confidence Interval Methods for Small Proportions in the Health Care Survey of DoD Beneficiaries". Proceedings of the Survey Research Methods Section, ASA. (www.amstat.org/Sections/Srms/Proceedings/y2005/Files/JSM2005-000468.pdf)