# Methodological framework for the evaluation of road safety measures

**Deliverable 3.3**

SafetyCube

# Methodological framework for the evaluation of road safety measures

Work Package 3, Deliverable 3.3

**Report Author(s):**
Elvik, R. (TOI); Perez, C. (ASPB), Filtness, A.; Talbot, R.; Thomas, P. (LOUGH); Theofilatos, A.; Papadimitriou, E.; Yannis, G. (NTUA); Saadé, J.; Phan, V.; Krishnakumar, R. (CEESAR), Hermitte, T. (LAB); Jaensch, M. (MUH); Wijnen, W.; Weijermars, W. (SWOV); Lassarre, S.; Carnis, L. (IFSTTAR); , Martensen, H.; Focant, N.; Daniels, S; Van den Berghe, W.; Diependaele, K. (VIAS);

| Due date of deliverable: | 01/02/2018 | Submission date: | 29/03/2018 |
|---|---|---|---|

Co-funded by the Horizon 2020
Framework Programme of the European Union

# Table of contents

# Executive summary

## CHAPTER 1: INTRODUCTION

This deliverable gives the theoretical framework for the SafetyCube Decisions Support System. It addresses the three main steps that have led to generating the content of the DSS.

1.  Repository of studies
2.  Synopses that summarize the studies on a particular topic
3.  Economic efficiency evaluation of countermeasures.

While the repository and the synopses concern risk factors as well as countermeasures, the economic efficiency analysis only concerns potential countermeasures.

In the SafetyCube Decision Support System, coded studies and synopses are divided into analyses of risk factors and evaluation of countermeasures. There is a common core to evaluating risk factors and countermeasures: both are characterized by how they change the probability for a crash or a casualty to occur. Nevertheless, the most important challenge for setting up a joint methodology is the wide variety of topics addressed, with research based on different principles, employing different designs and different reporting conventions. The differences *within* the domains of studying risk factors and *within* the domain of studying countermeasures were, however, larger than the differences between these two domains. Therefore this methodology treats different research traditons and different study designs, but does not structurally differentiate between methods for evaluating risk factors and those evaluating countermeasures.

## PART 1 SCIENTIFIC BACKGROUND

## CHAPTER 2: THREE APPROACHES TO IDENTIFY RISKS AND COUNTERMEASURES

As a background for the literature search on risks and measures, we have investigated how risks and measures are studied in the road safety domain. Three main approaches to the study of risk factors and countermeasures are presented. These approaches are the sequential approach, the epidemiological approach, and the systemic approach. The three approaches are complementary rather than competitive and can fruitfully be applied in combination.

**Sequential models** categorize and identify the phases of crashes, indicate where failures occur and what the consequences are. *A particular aspect of the traffic system is considered a risk factor if it has been identified as a contributing factor to crashes*. An important tool for that is crash reconstruction. Countermeasures can be evaluated by simulations based on the reconstructions. Crash reconstruction is also essential to define crash scenarios that have a chance to be prevented by a particular countermeasure (we call these crashes the target group for a countermeasures.

In the **epidemiological approach**, *a risk factor is defined as a factor which is overrepresented in crashes compared to its occurrence in normal traffic.* In the same way, countermeasures are considered effective when crashes are less likely in their presence than in their absence. The most common epidemiological measure for both is **relative risk**. Effects of countermeasures are also often quantified as crash modification factors (**CM factors**), the percentage by which the number of crashes is re-

duced due to the measure. Odds ratios are considered good estimators for relative risk as well as CM factors.

The systemic approach considers the interaction between user, vehicle, and environment. **Driving behaviour** is studied in naturalistic driving studies as well as on-road and simulator based driving tests. **Biomechanics** help to model possible impacts on the human body, to better understand injury mechanisms and allow the tests of countermeasures. *A particular aspect of the traffic system is considered a risk factor if it is associated with a worse performance of the system*.

## CHAPTER 3: STUDY DESIGNS IN RISK ANALYSIS AND THE EVALUATION OF ROAD SAFETY MEASURES

To be able to register different types of studies together in one template in a meaningful way, it is necessary to take a step back from the study itself and look at it in a methodological way: which study design was applied? What is compared to what? How were the results quantified? Only if properly characterised, results entered into the data-base will let the user to decide whether they are comparable to each other. To support this process, this chapter of the guidelines forms a methodological "crash course". To code studies it is important to indicate the possible biases of a study. At the end of Chapter 3, a table summarized the main biases of each study design type.

### Principal epidemiological measures

To estimate which effect the exposure to a risk factor or a countermeasure has on the crash or injury risk, epidemiological methods for the study of injury or diseases are used, in particular **measures of association**. The most important measures are *relative risk* (comparing the risk of an exposed group to the risk of an un-exposed group) and the *odds-ratio* (comparing the exposure of cases with a negative outcome to that of control cases without the outcome). For rare events, like road crashes, relative risk and odds ratios can be considered equivalent.

### A taxonomy of study designs in risk analysis and evaluation of road safety measures

A compact overview of the most common study designs is given and their typical biases are briefly mentioned. Each study design is characterised by a number of principles. Important differentiations are those between experiments and observational studies. Experiments can be divided into randomized control trial experiments and quasi-experiments. Typical experimental designs are: Between-group comparisons and before-after studies (which is a within-group comparison).

Observational studies can be analytical -- when different outcomes are linked to different exposures to a risk-factor or a countermeasure – or they can be descriptive giving the prevalence of either road safety outcomes (e.g., particular crash scenarios) or the prevalence of a risk factor. The most important observational analytical designs in road safety are case-control studies (comparing crash cases to non-crash controls) and cross-sectional studies, where outcomes and exposures are linked to each other by means of a (mostly multivariable) statistical model.

### Experimental study designs

In experimental studies the exposure to a risk factor or a countermeasure is *assigned* to the units under study (e.g., road users or road segments). The road-safety outcome is then compared between exposed and non-exposed subjects. In randomized control trial experiments the study-units are assigned to the conditions *at random* while in *quasi-experiments* this assignment was not under the control of the researcher.

**Randomized experiments** are often laboratory experiments or simulator studies and their main threat is their validity (experimental setting, tasks) and generalisability (sample selection). Other biases can be induced by comparing groups that are not a priori the same, repeated measurements on the same subject, and unwanted side effects.

The most frequent type of quasi-experiments are **before-and-after** studies. The main biases are regression-to-the-mean, long term trends, exogenous changes (e.g. in traffic volume). A control group is the minimum requirement to control at least for long term trends. The Empirical Bayes method helps to correct for all three biases.

### Analytical observational study designs

Most often quantitative studies in road safety are observational studies. Cross-sectional studies are widely used both to identify risk factors and to evaluate the effects of road safety measures. Unless they are corrected by multivariable accident prediction models, they are subject to many confounding factors. The most important variable to correct for is the distance driven (for road segments as well as road users) as well as gender and age (for road users).
An important threat to the validity of cross-sectional studies of measures is the endogeneity bias, the tendency for measures to be applied to units with extreme outcomes. This bias can lead to the reversion of the observed outcome and is difficult to control for in multivariable models.

**Cohort studies** follow-up a population over time registering the presence of risk factors or countermeasures and also whether a crash occurred. Because crashes are rare events, a very large population has to be studied over time to collect a meaningful number of crashes. Although this type of studies is considered ideal to link the occurrence of crashes to the presence of risk factors or countermeasures it is relatively rare in the road-safety domain.
**Case control studies** compare cases with a particular outcome (usually crash or injury) to controls without that outcome with respect to the distribution of a potential risk factor or countermeasure in both groups. An over-representation among the cases is taken as evidence for an increased risk and vice versa. In classic case-control studies, **prevalence of a risk factor in crashes** data is compared to the **prevalence among the non-crash** population (e.g., measured in road side surveys, travel survey's or odometer readings).
In **induced exposure** studies the control group is sampled from the crash population as well. It is sampled among neutral cases such that the control group can nevertheless be considered representative for the general population. The validity of induced exposure studies, therefore, depends crucially on the definition of the control group.
Generally, the most important threats in case control studies concern poor data on exposure to a risk factor, a-priori differences between cases and controls or inadequate control for such differences, and the simultaneous exposure of cases to several highly correlated risk factors.

### Meta-analysis

With an increasing amount of scientific evidence becoming available, meta-analyses are becoming more and more important. A joint estimator is computed taking into account the precision of each study's results. The main threat to validity is summing up studies that are not comparable.

### PART 2 RISK ANALYSIS AND ASSESSMENT OF MEASURES WITHIN SAFETYCUBE

This part gives the instructions to partners for literature search, coding studies, and summarizing them and, therefore, ensures the consistency within the DSS.

## CHAPTER 5: SELECTING AND PRIORITISING STUDIES

To identify relevant studies for the inclusion into the DSS, a systematic scoping review was conducted for each item in the taxonomy. The aim of this approach is to represent the body of literature in a scientific way. While the criteria applied differed between research fields, there was a schematic approach followed for each review, consisting of initial search, screening, identifying additional papers, and prioritizing papers for coding.

### Initial search

Initially, several relevant literature databases were searched, e.g., Scopus, Medline, and Google scholar based on well-defined logical strings of keywords (See Table 3, as an example). The keywords as well as the resulting number of studies were documented.

### Screening

The potentially relevant studies were then screened to assess their eligibility for further analysis. Generally, only studies with quantitative results were coded for repository. Important qualitative results were, however, included in the Synopses (see Chapter 8). Moreover topic-specific inclusion and exclusion criteria were applied and documented. This was done first on the basis of the abstract, then on the basis of the full paper. If few relevant papers had been retrieved, the reference lists of the selected papers were examined to identify any additional relevant papers.

### Prioritising

For several risk factors and measures, meta-analyses were already available. If this was the case, the most recent meta-analysis was used as the basis, and completed with additional studies published after, and consequently not included in that meta-analysis. Studies included in a meta-analysis were not included individually.

If there were too many other papers, they were listed in descending order of importance for the road safety DSS, based on outcome, transferability, recent publication date, language and source. Note that these criteria were applied flexibly depending on how many studies were available and the field of research. Papers that evaluated measures and risks in terms of observed crashes were considered more relevant than those based on observed road safety behaviour (e.g. speeding), which again were considered more relevant than studies that had other indicator variables as outcomes (e.g., self-reported behaviour, driving simulator data, simulated crash data, etc.). SafetyCube is focused on Europe, therefore prioritizing European studies above US/Australian/Canadian studies. The latter are prioritised above studies from other countries. Other criteria were publication date (recent studies before older studies, though older studies of particular relevance were included), language (papers in English before papers in other languages), and source (peer reviewed papers before non-peer reviewed papers).

## CHAPTER 6: CODING STUDIES

One of the main objectives of the SafetyCube project is to create a repository of estimates of risk factors and safety effects. While there are already a number repositories of safety effects around, these are tailored to infrastructural measures. In SafetyCube a much broader scope is applied, comparable e.g. to the Handbook of Road Safety Measures (Elvik et al., 2009), where measures directed towards infrastructure, vehicles, human behaviours and post impact care are evaluated. In contrast to all existing repositories, SafetyCube departs from the perspective of risk factors which makes the type of studies included into the repository even more diverse.

The collected studies investigated the effect on different outcome variables: crash counts, simulated crash data, injury severity, on-road driving, driving in a simulator, crash simulations, and so on. They

employed a large variety of research designs: before-after studies, cross-sectional designs, case-control, induced exposure, time-series; and statistical methods: simple comparisons of counts or means, different types of regression analyses, Empirical Bayes, hazard rate, to name just a few. The enormous differences between studies constitute a big challenge for the creation of a joint database. The structure is general enough to allow coding different kinds of safety or risk effects and flexible enough to capture all important details of different types of studies. For each study, therefore, the template includes general information of the sampling frame and study conditions (e.g. road user types, severity of crashes, road types included), but also allows for the inclusion of conditions that are relevant to the specific area only (e.g. the differentiation between different injury types or details of the roadway design). Furthermore, for each estimated effect the following specifications were registered:

• what was compared to what
• analysis method/model
• measure of effect (often odds ratio but also many other less used measures of effect)
• statistical results (standard error, confidence interval)
• conclusion (significant effect on road-safety or not).

The selected studies are individually coded in an Excel coding template. The coding template consists of several sheets, requiring the researcher to provide information, mostly in predefined categories. On the basis of the study features coded, a result table shapes itself in which the results for all conditions can be entered.

Another important issue is the quality of research results. Possible biases of a particular study are coded with an indication of how severe this possibility is believed to be. To this end, common biases for the major research designs are described and included into the coding template, so that these (or other) problems can be flagged if necessary. Finally, the researcher included a brief verbal summary of the study, including the main findings, as well as an assessment of their reliability and usefulness, given the study design and potential biases.

## CHAPTER 7: EXAMPLES OF CODING STUDIES

A number of studies are presented, their design and results shortly described together with the result tables contained in the paper. Subsequently it is shown how these studies have been coded in the SafetyCube Coding Template. Examples address the following types of studies:

• before after
• case-control
• experimental
• cohort
• accident prediction models
• cross-sectional
• meta-analysis

## CHAPTER 8: SUMMARIZING STUDIES

For each risk factor or road safety measure, a synopsis has been compiled. The synopsis provides a synthesis of the findings for a specific risk factor or road safety measure, including both quantitative information from the coded studies and more qualitative information from previous review studies. The synopsis aims to complement other outputs of the DSS, like lists of available studies and direct access to the results of individual studies (see Chapters 6 and 7).

Each synopsis consists of three parts:
• Summary: In maximum two pages, the summary very briefly reports some background of the topic concerned, and the main results and conclusions based on the analysis.

- Scientific overview: In approximately four to five pages, the scientific overview describes the essence of the way the reported effects have been estimated, including a full analysis of the methods and results, and its transferability conditions in order to give the user all the necessary information to understand the results and assess their validity.
- Supporting documentation: The supporting documentation gives a more elaborate description of the literature search strategy, as well as the details of the study designs and methods, the analysis method(s) and the analysis results. Here, also a full list of coded studies and their main features is provided.

When writing a synopsis, researchers proceed from the detailed analysis described in the supporting documentation to giving an overview of the results, to writing a short summary. This is why this chapter is written in the opposite order as the resulting synopses. Synopsis writers are however instructed not to report the same material twice. If contents, tables, etc. are included in one of the more global parts (overview or summary), it is deleted from the supporting documentation. Often, this means that the majority of tables described in the supporting documents (if they are the most compact result available) are actually moved "up" either to the overview or even to the summary.

### Supporting Document: the analysis

After having coded all selected studies, the researchers analysed the results. Three ways had been defined to analyse and summarise the results, in a decreasing order of priority:

- Meta-analysis, if there is a sufficiently large number of studies that are comparable in terms of both their scientific design features and the type of results they produced. A meta-analysis combines the numerical results of multiple studies and yields a weighted average from the results of the individual studies.
- Vote-count analysis, if a meta-analysis is not possible due to large differences between studies, but if there is a sufficient number of studies. A vote-count analysis compares the share of studies that showed a positive effect, no effect, or a negative effect.
- Review-type analysis, if the number of studies is small or if the studies are so heterogeneous that a vote-count analysis is not meaningful. In a review-type analysis the results are summarised in a more qualitative way, generally including a table of study descriptions (e.g. sample, method, outcome), the observed effects and their interpretation.

In each type of summarising analysis attention was dedicated towards the identification of ***modifying conditions*** (e.g. a countermeasure that works in urban, but not in rural settings or a risk-factor that is more dangerous for novice drivers). In meta-analyses or vote count analyses, this was addressed by sub-group analyses. In this context, ***transferability*** of the results is also discussed, giving an indication of situational characteristics that can impair the validity of the reported estimates for another implementation of a particular countermeasure.

### Scientific overview (5 pages)

In this part, literature review that accompanied the coding of studies is described in the necessary detail. This includes a description of modifying conditions (when does the factor show its largest effect, when the least?) and possibly a brief summary of relevant theoretical models. If applicable crash data are included (target scenarios for measures, for risk-factors frequencies in crashes), and an overview is given over the study results with respect to the factor/measure in question.

### Summary (2 pages)

Within the summary, the information proceeds once more from very compact to relatively detailed. The most compact information is given in the **colour code** that is based on the results of the (majority of) the studies' outcomes to indicate the overall conclusion about the effect. Each colour code is

supported by a ***short statement of two to three sentences***, which is visible to the DSS users before they even click on the PDF file containing the synopsis.

| Risk factor | | Countermeasure | |
|---|---|---|---|
| **Red** | Results consistently show an increased risk when exposed to the risk factor concerned. | **Green** | Results consistently show that the countermeasure reduces road safety risk. |
| **Yellow** | There is some indication that exposure to the risk factor increases risk, but results are not consistent. | **Light green** | There is some indication that the countermeasure reduces road safety risk, but results are not consistent. |
| **Grey** | No conclusion possible because of few studies with inconsistent results, or few studies with weak indicators, or an equal amount of studies with no (or opposite) effect. | **Grey** | No conclusion possible because of few studies with inconsistent results, or few studies with weak indicators, or an equal amount of studies with no (or opposite) effect. |
| **Green** | Results consistently show that exposure to the presumed risk factor does not increase risk. | **Red** | Results consistently show that this measure does NOT reduce road safety risk and may even increase it. |

In the **abstract** of no more than half a page, the measure/risk-factor is shortly described giving possible restrictions (e.g. effect is only temporary, risk factor is particularly dangerous for inexperienced drivers).

A little more information on how the measure/risk-factor "works" is given in the **background** (1.5 pages). This section also includes a very compact presentation as to when it has its largest effect and when the least, as well as a short indication of how well the effect has been studied so far.

### CHAPTER 9: META-ANALYSES

Very compact instructions to conduct a meta-analysis are given in Chapter 9. This includes the data-inspection methods addressing possible heterogeneity of the data, different model types (fixed vs. random effects), as short introduction to meta-regression and an illustration of an analysis in R, a freely available software.

### PART 3 ECONOMIC EFFICIENCY OF COUNTERMEASURES

Economic efficiency evaluation ($E^3$) is included in the SafetyCube Decision Support System (DSS) by example analyses for a large range of measure and by the ***$E^3$ calculator***, that allows users to run their own analyses or adjust SafetyCube examples to their own situation.

# CHAPTER 10: THE E³ CALCULATOR

| **User input**<br>**Info on measures** | *Output E³-calculator* | *SafetyCube input*<br>**Info per country** |
|---|---|---|
| **Effectiveness**<br>Saved crashes per unit<br>*(fatal, serious, slight, pdo)* | **Cost Effectiveness Analysis**<br>• **Costs per crash prevented**<br>*(fatal, serious, slight, pdo)* | **Crash & casualty costs**<br>*(fatal, serious, slight, pdo)* |
| *Time horizon* | **Cost Benefit Analysis**<br>• **Net present value**<br>*(benefits − costs)*<br>• **Benefit-cost ratio**<br>*(benefit / costs)* | Discount rate |
| **Costs of measures** | | |

The E³ calculator allows partners, to combine the information about the effectiveness of a measure (i.e. the percentage of crashes or casualties prevented) with the costs of these measures. The calculator also integrates the information of crash costs collected in the SafetyCube project, allowing to express all costs and benefits of a measure in monetary values and conducting benefit-cost analyses. The description of the calculator in this deliverable form hands-on instructions for partners, who use the existing excel template. The theoretical background to cost-benefit analyses, as well as instructions how to interpret and proceed with the results are documented in D3.4 and D3.5.

# CHAPTER 11: INPUTS TO THE E3 CALCULATOR

As input to the calculator for the Economic Efficiency Evaluation, the following is needed

- Measure costs
  - Initial costs
  - Annual costs
Number of crashes / casualties prevented (for each level of severity)
  - Target crashes of countermeasure
  - % reduction
- Time horizon of a measure

This is joined with the cost of crashes, because the costs of the saved crashes from the benefits in a cost-benefit analysis. This information is provided by SafetyCube for each European country.
- Crash/casualty costs per unit
  - Fatal crash/ fatality
  - Severe crash/ severe injury
  - Slight crash / slight injury
  - Damage only crash (not applicable to casualties)
- Discount rate

The discount rate is a percentage that is detracted from benefits and costs for each year that they are delayed into the future. It is a fixed value for each country. The users of the E3 calculator can choose whether to use the countries own reported crash-costs (e.g., if the analysis is intended for official use within the country) or the crash-costs estimated in SafetyCube applying the common methodology (e.g., if comparability to other countries is desired). The SafetyCube analyses are by default conducted with the EU standardized crash-cost values.

## CHAPERT 12: CALCULATIONS IN THE E3 CALCULATOR

For each year of the horizon, the number of saved crashes (casualties) are calculated per severity category. The percentage reduction is considered to remain constant and the target group is diminished by the saved crashes in the previous years. Subsequently, for each year, the benefits (saved crashes * crash costs) and annually recurrent costs for the countermeasures are calculated and discounted . The implementation costs are assumed to take place in the present and are not discounted. Subsequently all saved crashes, costs, and benefits are summed up and compared with the summed up costs.

## CHAPTER 13: OUTPUT OF THE E3 CALCULATOR

On the basis of this input and the crash or casualty costs, the calculator adds for each year within the time horizon the present value of all costs and benefits, resulting into the following outputs:
- Number of crashes / casualties prevented (per unit of implementation)
- Cost effectiveness: cost per prevented crash / casualty
  - Costs per prevented fatality / fatal crash
  - Costs per prevented severe injury / severe crash
  - Costs per prevented slight injury / light crash
  - Cost per prevented damage only crash (if applicable)
- Total benefits
- Cost benefit ratio (benefits/costs)
- Net effect (benefits – costs)

If no measure costs are entered, the break-even costs are calculated: the costs of the measure at a cost-benefit ratio of 1. This indicates how much a measure could maximally cost and still be cost-effective.

The results of economic efficiency analyses depend critically on the input values and consequently a *sensitivity analysis* is advised. Apart from the crash costs (which are imported from SafetyCube), the most influential parameters are the measure costs and the effectiveness of the measure. For each SafetyCube example, the following variations were calculated:
- Low measure effect (lower confidence interval from source study)
- High measure effect (upper confidence interval from source study)
- High measure costs (+ 100%)
- Low measure costs (-50%)

These were combined to two scenario's:
- Worst case: low effectiveness + high costs
- Ideal case: high effectiveness + low costs

## CHAPTER 14: CONCLUSION AND OUTLOOK

The main achievements of the methodology are summed up, pointing to widely differing research traditions, methods, and standards in the field of road safety. Collecting studies in a repository for estimates of risk factors and safety effects (countermeasures) is an important first step to consider results from different types of research jointly and thus to gain the maximum of information on each risk factor and each countermeasure. We also discuss remaining steps to derive comparable crash modification estimates from different types of studies.

# 1 Introduction

This chapter describes the SafetyCube project and the role that the present deliverable, the methodological framework played. A short description of the Work Package which produced the deliverable is also provided.

## 1.1    SAFETYCUBE

Safety CaUsation, Benefits and Efficiency (SafetyCube) is a European Commission supported Horizon 2020 project with the objective of developing an innovative road safety Decision Support System (DSS) that will enable policy-makers and stakeholders to select and implement the most appropriate strategies, measures and cost-effective approaches to reduce casualties of all road user types and all severities.

SafetyCube aims to:
1. develop new analysis methods for (a) Priority setting, (b) Evaluating the effectiveness of measures (c) Monitoring serious injuries and assessing their socio-economic costs (d) Benefit-cost analysis taking account of human and material costs
2. apply these methods to safety data to identify the key crash causation mechanisms, risk factors and the most cost-effective measures for fatally and seriously injured casualties
3. develop an operational framework to ensure the project facilities can be accessed and updated beyond the completion of SafetyCube
4. enhance the European Road Safety Observatory and work with road safety stakeholders to ensure the results of the project can be implemented as widely as possible

As yet there is no systematic pan-European in-depth study of crash causation and it is very difficult for policy-makers and other road safety stakeholders to assemble a clear evidence base of the causation paths and associated risks. In a similar manner there is also no systematic catalogue of measures and their safety effects. There are many individual studies of well-established measures in the literature, but the measured effectiveness, limitations and applicability can be highly varied. It is therefore difficult for road safety stakeholders to form conclusions over the most appropriate measures to be deployed.

SafetyCube addresses this gap by providing a comprehensive analysis of both crash risks as well as the effectiveness and benefit-cost of safety measures. SafetyCube's objective to prepare a Decision Support System that will enable countries to adopt the best possible approach to road safety is highly challenging, because it requires a comprehensive and consistent evaluation of crash causation factors and quantified risks as well as a clear presentation of the effectiveness of road safety measures – on which the information is currently highly diverse, unstructured and often incomplete. Even the best performing countries do not have available an evidence-base of the breadth and depth to which SafetyCube will work.

The data that is available, and which is deployed within the project, has been gathered for a variety of purposes using a range of protocols and selection criteria. It is therefore a significant challenge to bring this data together to form a single coherent analysis of crash causation mechanisms and risks. SafetyCube focuses on road users, infrastructure, vehicles and injuries framed within a systems ap-

proach with road safety stakeholders at the national level, EU and beyond having involvement at all stages.

The SafetyCube team includes an impressive group of data analysts, researchers and policy advisors who are highly experienced in transferring the research results into well-founded policy-support information. To do this a series of new procedures has been developed to combine and analyse the safety effect of a wide range of measures, thereby extending the current level of knowledge and simplifying and making accessible what is currently a very large body of knowledge.

A further area where the project will develop the state of the art to a new level of understanding concerns analysis of the costs and benefits of measures. There is currently a lack of systematic information on the cost-effectiveness of measures when implemented in the European context. Cost information is scarce, particularly when concerning vehicle based measures. There is currently no method available that enables comparable calculations of cost-effectiveness for crash avoidance, crash mitigation and injury mitigation technologies.

SafetyCube will address each of these challenges within one compressive online tool, called the DSS (Decision Support System). This will advance the state of the art in the understanding and access to information for informing evidence-based road safety policy making.

A systems approach provides a framework within which the work of other Work Packages is integrated into the DSS. A road collision is rarely the result of a single factor. Risk and problems from road user behaviour, infrastructure and vehicle deficiencies interact with each other resulting in environments within which a crash may occur. Understanding these risks and the most appropriate measures and solutions to mitigate them is central for evidence based policy making. In order to provide policy-makers and industry with comprehensive and well-structured information about measures, it is essential that a systems approach is used to ensure the links between risk factors and all relevant safety measures are made fully visible.

## 1.2 WORK PACKAGE 3 WITHIN SAFETYCUBE

This deliverable 3.3 is the fourth deliverable produced within Work Package 3. The objective of Work Package 3 is to define the methodological foundations of the road safety Decision Support System. The methodological guidelines developed are applied in Work Packages 4, 5, 6, and 7 to identify and analyse road safety problems and measures addressing road users, road infrastructure, vehicles, and trauma care. A road safety decision support system should help policy makers identify important risk factors and the crashes, injuries and fatalities resulting from them; select measures by estimating their safety effects; and set priorities among measures on the basis of their benefits and costs.

There was a close interplay between Work Package 3 and 8, as WP8 was responsible for implementing the methodology, enforcing a systems approach, and creating the Decision Support System on the basis of the content delivered by WP3.

The WP3 methodology describes
- how studies are selected and prioritised by a systematic and documented literature search
- how studies are entered into a coding template capturing
  - the study design (what was compared to what)
  - the selection of the study population (who was included?)
  - the testing conditions (obstacles to transferability)
  - the results (numerical value and significance of each effect)

- possible biases
- how studies are summarized in a synopsis that gives
  - general information about the measure
  - gives very compact information in the abstract as well as a detailed report
  - conditions under which a countermeasure or risk-factor show their greatest effect.
  - concerns about transferability of the results

A number of topics have been addressed in WP8 and are therefore not discussed in this deliverable. See Deliverable 8.5 for a discussion of the issues below:
- The taxonomy of risk factors and countermeasures
- Links between risk factors and countermeasures
- The structure of the database underlying the DSS
- The user-interface of the DSS

## 1.3 IMPORTANT STEPS IN DEVELOPING A ROAD SAFETY PROGRAM

To put the report into a wider context, the next section gives a brief overview of the road safety policy making cycle indicates the type of information policy makers need at each stage of policy making. Subsequently it is discussed how the Decision Support System (DSS) will support these steps.

To develop an evidence-based road safety programme, policy makers need information to help them (adjusted from Elvik, 2007)
- Identify important risk factors and road safety problems
- Identify potentially effective road safety measures
- Estimate the effects of each measure on crashes or injuries based on current knowledge
- Evaluate the economic efficiency of each measure
- Assign priorities between measures

Each of these supporting functions are briefly explained below.

### Identification of important risk factors (road safety problems)

The term "risk factor" denotes any factor that contributes to crashes or injuries. There are risk factors related to all elements of the road system and the interactions between these elements. The purpose of identifying risk factors is to provide a basis for selecting the most important road safety problems for treatment. The importance of a risk factor can be defined as the size of the contribution it makes to crashes or injuries. The greater the contribution, the more important the risk factor. This concept will be explained in more detail later in the report.

A list of risk factors that contribute to crashes or injuries can be sorted according to many criteria. The importance of a risk factor is just one of these. Not all risk factors are equally amenable to treatment by means of road safety measures; the SafetyCube risk factors are therefore linked to road safety measures. The process underlying this linkage is described in D8.5. For each risk factor a range of road-safety measures is indicated. The measures come from different domains (road user, infrastructure, vehicle, trauma care) because it is important to understand that an effective treatment of a risk factor related to a specific element of the road system may not always be related to that element, but may be related to a different element. As an example, drinking-and-driving is a behavioural problem related to the driver, but the solution may be a technical device, alcohol ignition interlocks.

**Identify potentially effective road safety measures**

The SafetyCube Decision Support System delivers a wide list of potentially effective road safety measures should be conducted. A road safety measure is regarded as potentially effective if either: (a) it evaluates studies that have shown an reducing effect on crash and/or injury outcomes or, (b) it favourably influences risk factors that are known to contribute to crashes or injuries. The DSS also includes measures which are frequently applied even if there is no evidence for a protective effect.

**Estimating the effects of road safety measures based on current knowledge**

This deliverable writes out the methodology for collecting, coding and summarising the current knowledge on the countermeasures included into the system. There are also outlines how to indicate whether the present knowledge is satisfactory and to what extent it is transferable to different situations. The principle suggested here is "in dubio pro reo". Existing estimates should be used unless there are reasons for not using them.

**Economic efficiency evaluation**

To select countermeasures, their effect on different types of crash outcomes has to be considered relative to the costs of each measure. To do so in a benefit-cost analysis, one needs a valuation of crashes and casualties of different severity. The SafetyCube project, therefore, cooperated with the InDev Project to collect crash and casualty costs from all European countries and evaluate them in terms of the methodology used to estimate the costs (Deliverable 3.2). Deliverable 3.4 gives a background on benefit-cost analyses with guidelines of how to estimate the costs of measures.

The SafetyCube project used the collected crash costs to produce a calculator for the economic efficiency evaluation. The $E^3$ calculator produces a benefit-cost ratio and other measures of effectiveness. While the underlying principles for this calculator are described in D3.4, this deliverable gives hands-on instructions how to use the $E^3$ calculator.

**Guidance on priority setting**

From a strictly economic point of view, the priorities between road safety measures should be based on their net benefits only (benefits minus costs). In practice, however, it is rarely possible to set priorities strictly according to such an economic criterion. It is legitimate to set priorities that depart from benefit-cost analysis, but it is good practice to be explicit about this and justify the priorities that are set. These principles are further worked out in D3.5.

## 1.4     STRUCTURE OF THIS DELIVERABLE

This deliverable gives the theoretical framework for the Decisions Support System. It addresses the whole process that has led to generating the content of the DSS. The contents consist of the following elements:
1. Repository of studies
2. Synopses that summarize the studies on a particular topic
3. Economic efficiency evaluation of countermeasures.

While the repository and the synopses concern ***risk factors as well as countermeasures***, the economic efficiency analysis only concerns potential countermeasures.

For the repository and the synopses, the most important challenge for setting up a joint methodology is the wide variety of topics addressed. Research on these topics is based on different principles,

employs different designs, and has different publication policies. It turned out that research on different risk factors showed significant differences and so did research on different countermeasures. In fact, the differences within these two domains were larger than between them, because there is a common core to evaluating risk factors and countermeasures: both are characterized by how they change the probability for a crash or a casualty to occur. Although in the DSS, coded studies and synopses are divided into analyses of risk-factors and evaluation of countermeasures, this is not the case for the methodology. Both are treated in the same general way.

The deliverable consists of 3 Parts. **Part 1** describes the scientific background to features that become important when coding studies for repository or when summarizing the results for a countermeasure in question. In *Chapter 2*, three general approaches taken to identify risks and evaluate countermeasures are described. These are the sequential approach, the epidemiologic approach and the systemic approach. A general idea of the research principles of each approach is given with some examples and it is explained how a good synopsis on any risk factor or measure should combine results from each of the three types of studies. In *Chapter 3*, we are turning to research designs. One of the big challenges was to be able to deal with different research designs. The coding template for the DSS was made very flexible, so that all different kinds of quantitative evaluation studies can be entered, preserving the information about study-design and type of information collected, but also allowing to compare the results. The information coded in the templates is only valuable when it allows the user to know exactly what the figures mean. A good understanding of the research designs applied is therefore a necessary requirement for entering studies into the DSS. In *Chapter 3*, a taxonomy of research designs is introduced and subsequently (*Sections 3.2 – 3.7*) each design is described in more detail with one or two examples.

**Part 2** describes how the literature with respect to a risk factor or countermeasure is reviewed for the decision support system. The studies are selected and prioritised by a systematic and documented literature search (*Chapter 4*), "analysed" in terms of their research design and possible biases, entered into a coding template capturing all relevant information for the DSS users (*Chapter 5*). Because it is not self-evident how to code different types of studies, examples of coded studies are given and explained in *Chapter 6*. Finally, different studies addressing a countermeasure or risk factor are **summarised into a synopsis** using the information contained in the coding template and other information from the literature review (*Chapter 7*). Whenever possible and useful, the synopsis contains a meta-analysis, for which a short instruction is given in *Chapter 8*.

**Part 3** concerns the economic efficiency evaluation ($E^3$). This part details how to use the $E^3$ calculator that has been used for all economic evaluations in the SafetyCube project. *Chapter 10* gives detailed explanations about the input required for the calculator. *Chapter 11* gives the underlying calculations. This chapter is meant as a technical reference. For a more general description of the principles underlying various types of economic efficiency evaluations see D3.5. The outputs of the calculator are presented in *Chapter 12*. Again, see D3.5 for more information about the interpretation of the outputs.

**Part 4** contains the conclusion on the methodological framework presented here and an outlook in its application to continue up-dating and improving the SafetyCube Decision Support System.

# PART 1 – Scientific background

# 2 Three Approaches to Identify Risks and Countermeasures

The approaches to identify risks and risk factors must be coherent with the three crash models defined by Hollnagel (2004): sequential, epidemiologic and systemic, which are standard in the crash investigation and prevention.

In the *sequential approach*, the crash process can be seen as a chronological succession of phases. The output of each phase determines the initial conditions of the next phase. The relevance of the analysis results from the superposition of a spatio-temporal sequence of events and a causal logic conditioning the passage from one phase to the next. Road crashes are considered as consequences of failures of the components of the driving system (driver, vehicle, environment). Road crashes are grouped into scenarios – mostly on the basis of expert judgement but sometimes also by means of statistical clustering – for a better understanding of the different causation mechanisms but also for the evaluation of potential countermeasures.

In the **epidemiological approach**, simple combinations of failures of systems, components or human failures, combined with latent failures (design, maintenance, management...) contribute to the crash by affecting the defence. A risk factor is defined as a characteristic of one component of driving system (driver, vehicle, environment) which is over-represented in crashes compared to its occurrence in normal traffic. Risk factors can be quantified by calculating the relative risk or the odds ratio. In the domain of measure evaluations the reduction of the crash risk is expressed as the CMF (the **crash m**odification **f**actor or function – if the reduction depends on another parameter). Safety performance indicators (risk-factors with a well-known crash risk) can be considered and Biomechanics help to model possible impacts on the human body.

In the **systemic approach**, the hypotheses of decomposition, linearity and simple (sequential) or complex (epidemiological) combinations of failures are no more appropriate. The driving system is considered as complex, non-linear and safety is an emergent property of such a system. Functional variability is the factor that should be monitored and reduced to achieve safety of the system. To gain more precise knowledge of risk factors and fine tune countermeasures, studies can be based on indirect measures of road safety, like driving behaviour in driving simulators and on-road driving tests.

The three approaches to risk analysis and measure evaluation should be considered complementary. Only the epidemiological approach allows estimating increases or reductions in crash outcomes, while the other two approaches serve to inform the design and the fine-tuning of countermeasures.

## 2.1 SEQUENTIAL APPROACHES TO CRASH CAUSATION – ANALYSIS OF RISK FACTORS

The main assumption of this approach is that we can learn from crash investigation by identifying the failures of the components of the system supposed to be decomposable, and by coming up to their single causes which could be seen as root causes. The safety is obtained by the elimination of the root causes.

For the crash causation point of view, the analysis is made in the four phases: the driving, rupture, emergency and collision phases, in order to identify the failures of different order (technical, human or organisational) which leads to the crash and to relate them to the factors which could explain or contribute to these failures.
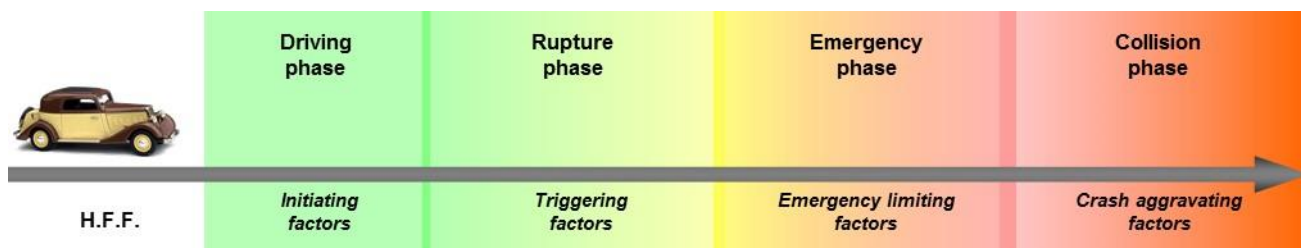


**Figure 2.1** Example of use of the sequential model. Source HFF, Van Elslande et al. (2008)

A number of methodologies have been developed to analyse crash causation for road traffic crashes using a sequential approach:
- Cinematic reconstruction of the crash/ collision, and reconstruction of injury process (bio-mechanics),
- Typology of Crash scenarios
- Accident Causation Analysis System
- Human functional failure (HFF) as special look to the driver (Van Elslande et al., 2008)

For a comprehensive overview of crash causation methodologies relating to road traffic crashes see Hermitte (2012).

To test the effectiveness of a safety device when the safety system is not largely available on the market and sufficiently spread to be able to find a large enough sample of vehicles provided with this system in the databases of injury crashes (for example for new automatic driving aids), we have to use a priori evaluation methods. This requires a case-by-case analysis of the entire data of the individual crashes including the crash reconstruction. When we know the trajectories of the crash involved participants, it is possible to influence one or all the participant's behaviour in order to check the consequences of the influence. This approach has for example a long history in forensic crash analysis, when the question is asked whether or not the participant with right of way would have been able to avoid the crash, for example running with the allowed speed or reacting in time. Due to the increased interest in active vehicle safety systems (systems that reduce the crash risk) this approach is used in a broader way also for the evaluation of countermeasures. Then for each case it has to be decided whether a particular countermeasure could have helped.

### 2.1.1 Cinematic reconstruction of the collision and of injury severity

The objective of the reconstruction of crashes is to estimate the trajectories and velocities of the crash involved participants before the critical phase of the crash. Crash reconstruction is a crucial

task in crash research because it helps to understand the circumstances of the crash and to integrate the crash severity (for example described by the velocity change due to the collision or the deformation energy) when comparing crashes of different severity levels.

For the reconstruction of crashes, one has to consider a relatively large variety of results concerning, among others, the final position of the vehicles, marks at the street, damages and injuries. Often the exact position of the participants when colliding remains unknown and is by itself a result of the crash reconstruction. However, the distance between location of collision and the location of final rest has an important influence on the calculated impact velocities.

When reconstructing a crash, all available information is considered, together with the physical laws such as conservation of energy and conservation of momentum, in order to reduce the possible variation as much as possible. Today computer programs such as PC-Crash are normally used for crash reconstruction. These programs are mainly calculating forces for the pre- and post-crash phase and the momentum for the in-crash phase. Although exact solutions (for example with respect to the final position of rest) can be achieved, this does not necessarily mean that the used assumptions comply with the real crash conditions. However, from the comparison of real crashes and crash tests it can be concluded that experienced crash analysts are able to calculate the real crash conditions within reasonable tolerances (Johannsen, 2013).

Real crash reconstruction is necessary to perform some crash simulation not only need to have an estimation of the vehicles kinematics according to the time (e.g. trajectory, speed, etc.), but because in some cases we need to perform a new simulation of the crash with the implementation of the safety system on the vehicle(s). The tools used to perform a numerical simulation depend on the complexity of the studied system, on the numerical models available, on the man-month effort available (for a case by case analysis), etc. Sometimes the use of Excel is enough, sometimes we need only a reconstruction model such as PC-Crash or V-Crash, sometimes we need to use several numerical models (dynamic vehicle, crash model, sensor model, traffic model, etc.), sometimes we need to create a dedicated tool. Some of these models are able to estimate all the sequences of the crash injury levels included, but most of them are only able to estimate the new collision speed (in cases where the crash is not avoided).

### 2.1.2 Scenarios

In the sequential approach the aggregation of crashes depending on the context (human, vehicle, infrastructure, etc.) can get very complex. In order capture this complexity, one of the tools most frequently used in road safety is the concept of scenarios. A scenario is a cluster of crashes showing similarities from the studied point of view. Scenarios usually propose an (almost) exhaustive and exclusive classification of the studied object.

For the road safety diagnosis, they allow a good overview of the studied problem. Scenarios are important to improve the road safety diagnosis. We can either identify the problems and propose adapted countermeasures (top-down approach) or start with existing safety solutions and identify the residual problems (bottom-up approach). It must be kept in mind though that scenarios are always linked to an initial research question and put forward a classification of the population according to the similarity of the studied characteristics. Every scenario can be also subdivided into sub-scenarios and possibly sub-sub-scenarios.

For the end users, the main interest of scenarios is to allow work on groups of individuals instead of each individual themselves. Thus, the number of sub-levels must be chosen intelligently: enough to describe the problem correctly but not too numerous that the population is too small. The first level of scenarios is generally too generic to be able to distribute the set of the individuals. The sub-

scenarios allow us to take into account other more specific characteristics. Another advantage of this hierarchy is to be able to by-pass the problem of the missing values. In the case of missing information, the individual who cannot be placed in a subclass will belong to the superior class. For a given problem, the relevance of the scenarios depends on one hand on the way the scenarios were developed (level of granularity, interpretation and independence classes, etc.), and on the other hand on the quality and adequacy of the data used to complete the scenarios.

There are two main ways to build scenarios, the one using the statistical tools (data clustering, K-mean, Kohonen, hierarchical ascending classification, etc.), the other one based on the expert method. The outcome of statistical methods depends crucially on the selection of markers (variables), which should be the most relevant to characterise the problem this again can be based either on statistical methods or on expert opinion). The difficulty in the use of the statistical methods in the creation of scenarios is based on the interpretation of the clusters so determined. Some combination of variables, even the modalities of variables, result in a grouping of individuals which is difficult to understand and can complicate the search for adequate countermeasures.

Most scenarios are based on expert judgement and therefore entail a good interpretation of the research question but also on an excellent knowledge of the potential of the available data. The interpretation of each class is easier than in the statistical method because the resemblances are based on known and more concrete characteristics.

Several types of scenarios are used in road safety, most of the time associated with a well-established methodology (e.g. pedestrian scenarios, cyclists scenarios; Voiesur, 2011; CATS project, Human functional failure scenarios, TRACE project).
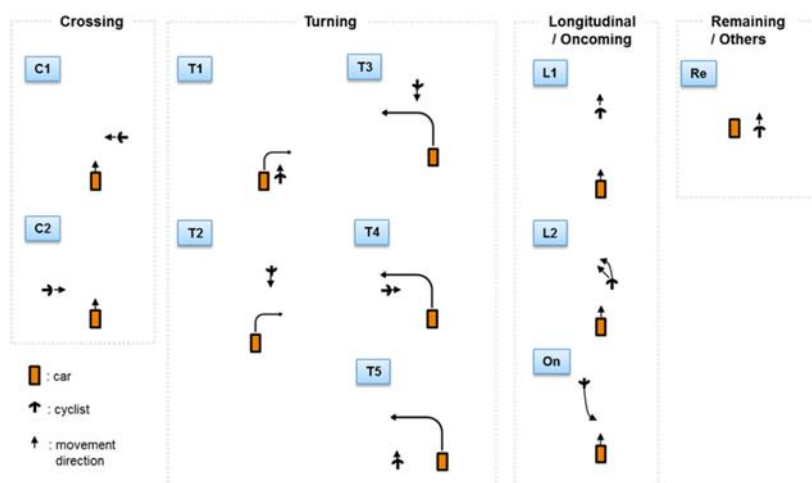


**Figure 2.2** Illustration of the scenarios used in the CATS project [2]

From the assessment point of view, scenarios allow us to select only the relevant injury crashes according to the studied safety system, in order to avoid having to simulate all the crashes. For example if we have to study the AEB City system (Automatic Emergency Braking for low speeds) the relevant scenario is composed by rear-end injury crash in urban area involving a striking vehicle with a collision speed below 50km/h (speed range where the system is active).

Scenarios are used first to quickly quantify the target population. In cases when the target population is very small it allows us to avoid the simulation step. Indeed, the target population gives the maximum number of crashes that could be avoided thanks to the contribution of the safety system

without taking into account all the limits of the use or functioning of the system. Only if this number is statistically significant will the selected scenarios be used for the simulation step.

### 2.1.3    Causation coding

Crash reconstruction is also used to identify factors that have contributed to the crash. To be able to summarize this knowledge across crashes, several coding systems exist. As an example, the Human Functional Failure Analysis will be described. Other systems are the Accident Causation Analysis System (ACAS, Jaensch et al.) which is used by the German In-depth Accident Research teams (GIDAS) and DREAM to which we will come back below.

The aim of the HFF methodology is to be able to clearly define the types of functional failures that humans experience in road collisions. It defines five main stages of the driving task (perception, diagnosis, prognosis, decision-making and taking action) and defines the types of functional failures that can occur at each stage. The HFF method does not use 'failure' to indicate fault, instead it aims to use the failures to identify the limits (physical and mental) of human capacity and therefore be able to understand better the types of countermeasures (i.e. safety systems) that would assist in overcoming these human limitations (Naing et al., 2007; Van Elslande & Fouquet, 2008). Figure 2.3 illustrates the stages of the driving task and the failures that can occur at each stage.
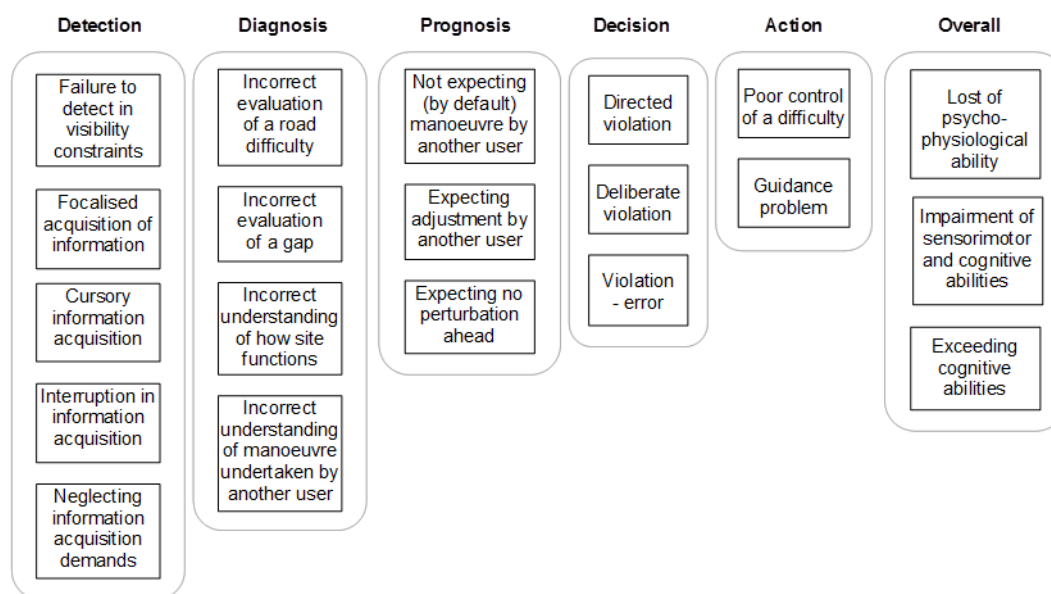


**Figure 2.3** Driving stages and possible human functional failures according to HFF (Van Elslande & Fouquet, 2008)

Next to the functional failure analysis, grids of contributory factors were developed in the HFF framework to determine typical failure generating scenarios. Three categories of factors were addressed:

- User (Human)
- Environment
- Vehicle (Tool)

The User category of factors is described as any factors related to the individual and personal characteristics. This includes any physical and psychological disorders that may be of relevance or any psychosomatic states that the user may have incurred through alcohol or misuse of drugs or emotional/motivational states. The user is defined as any human in charge of a vehicle within the crash (e.g. driver, motorcyclist, cyclist) or any pedestrian injured in the crash, and is described as a 'road user'. Three main subcategories of *user factors* were identified, as follows:

- User State
- Experience
- Behaviour

The Environment category encompasses all aspects related to the users' surroundings (i.e. external to the vehicle and road user). Six categories of **environment-related factors** were defined:
- Road Condition
- Road Geometry
- Traffic Condition
- Visibility Impaired
- Traffic Guidance
- Other Environmental Factors

The Vehicle category involves the equipment or devices the user is interacting with in the task. The subcategories developed to deal with the vast array of tools were:
- Mechanical - Vehicle failures which directly affect vehicle control
- Maintenance - Anticipated vehicle fault, indirectly affects control of vehicle
- Design - Design of vehicle affects safe/efficient operation
- Load - Did a vehicle load affect ability to control vehicle?

### 2.1.4    Countermeasures

Countermeasures can be classified according to Haddon's matrix (1980) taking into account the sequential dimension (before, during, after crash) and the systemic dimension (driver/ vehicle/ environment). Furthermore, Haddon proposes ten strategies to prevent the occurrence and the severity of injuries in a crash using the concept of mechanical energy as the agent of the damages.

| | Pre-Crash | Crash | Post-Crash |
|---|---|---|---|
| **Human Factors** | • Education and licensing<br>• Driver impairment<br>• Crash avoidance maneuvers (braking, turning, etc.) | • Health at time of crash<br>• Sitting properly in restraint<br>• Impairment | • Response to EMS<br>• Severity of injury<br>• Type of injury |
| **Vehicle/ Equipment Factors** | • Crash avoidance equipment and technology (lights, tires, collision avoidance, etc.)<br>• Vehicle design<br>• Vehicle load | • Speed of travel<br>• Functioning of safety equipment (seat belts, air bags, child restraints)<br>• Energy absorption of vehicle | • Ease of extraction from vehicle<br>• Integrity of fuel systems and battery systems |
| **Physical Environment** | • Road hazards<br>• Distractions<br>• Weather conditions | • Roadside features<br>• Guardrails<br>• Type and size of object struck | • Distance of EMS personnel<br>• Notification of EMS personnel<br>• Accessibility to crash victims |
| **Social/ Economic** | • Enforcement activities<br>• Insurance incentives<br>• Social norming<br>• Ability to use safety equipment appropriately | • Laws concerning use of safety equipment | • Trauma system equipment, personnel, training<br>• Information sharing |

**Figure 2.4** An example of Haddon's matrix (Source, NHTSA Safety In Numbers, June 2015)

To summarize, in the sequential approach risk factors and possibly effective countermeasures are identified in the careful analysis of the sequence of events in individual road crashes.

## 2.2 THE EPIDEMIOLOGICAL APPROACH TO THE EVALUATION OF RISK FACTORS AND COUNTERMEASURES

The starting point of the epidemiology of road crashes came with the identification by Haddon (1965) of the mechanical energy as the agent of the damages in case of a crash.
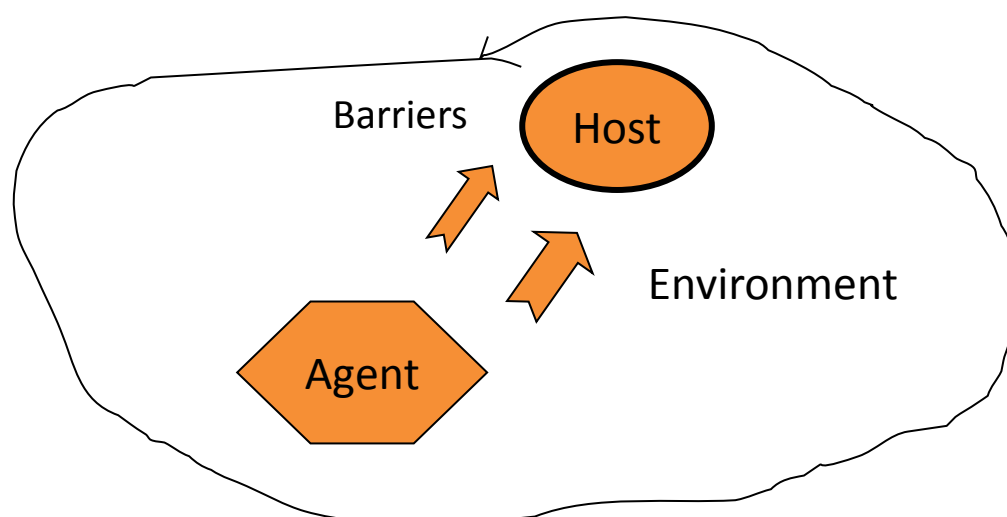
**Figure 2.5** Epidemiological model of disease/crash

In this model the characteristics of vector of the energy, i.e. the motorized vehicle, of the environ-ment, i.e. the infrastructure and the traffic, of the road user can be considered as risk factors.
In addition to these traditional proximal or distal risk factors, we could consider also organizational factors more related to the characteristics of the DVE system Driver/Vehicle/Environment and its interactions. We can use Reason's (2000) Swiss cheese model to make this idea more accessible: (See Figure 2.6) the slices of cheese represent factors in the system aiming to prevent error such as training, engineering or regulations. The holes in the cheese are the risks that exist in the system which are dynamic and are usually mitigated by the defensive layers (cheese). A crash occurs when the holes line up.
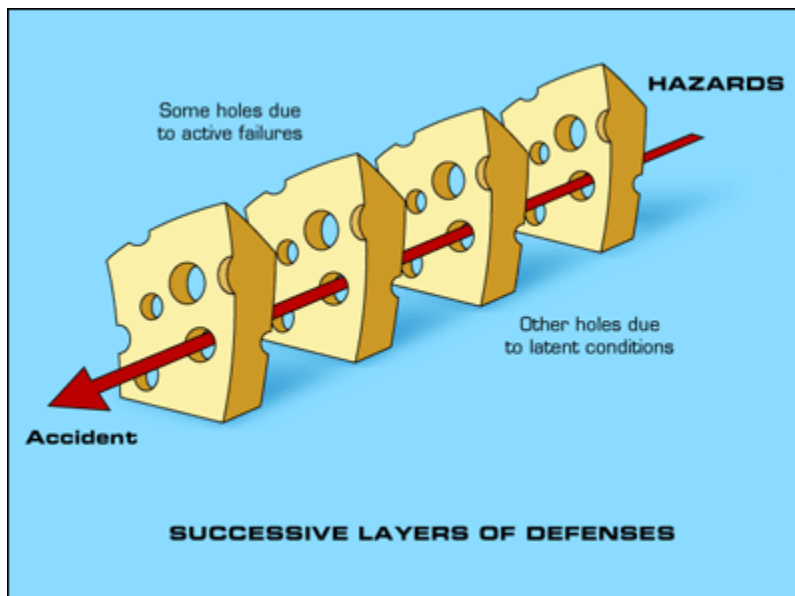


**Figure 2.6** Swiss Cheese Model (Reason, 2000)

### 2.2.1 The importance of a risk factor

By considering road crashes as a sort of "disease", Haddon (1965) opened the way to use epidemio-logical methods to quantify risk factors and the effect of countermeasures. In this approach the causal relation between a risk factor and crash occurrence or severity is not investigated directly, but inferred from the association between the two. This association is quantified as the ***relative risk***. The fact that 38% of the drivers injured in road crashes are under the influence of alcohol does not prove that alcohol had anything to do with the crashes. However, the fact that this proportion is about 10 times higher than in the normal driving population indicates that drunk drivers have a higher crash risk than sober drivers. A relative risk of 1 indicates that the risk with and without the presence of the factor is exactly the same. A relative risk higher than one indicates that the risk is higher when the factor is present. The relative risk can also be calculated for protection factors and countermeasures. A relative risk smaller than one indicates that the risk is smaller when the factor is present.

Formally, the relative risk is the probability to have a crash given the presence of a particular factor relative to the probability to have a crash given its absence.

$$RR = \frac{P(Acc|Factor)}{P\,(Acc|notFactor)}$$

With some rewriting this yields

$$RR = \frac{\frac{P(Factor|Acc)}{P(notFactor|Acc)}}{\frac{P(Factor)}{P(notFactor)}}$$

or in words, the presence vs. absence of a particular factor in the crash population relative to the presence vs. absence of that factor in the total population. For rare events like crashes the relative risk is usually approximated by an Odds Ratio (OR) (see Section 3.1).

### 2.2.2 The effect of a countermeasure

A countermeasure is in principle a protection factor – i.e. a risk factor with a value smaller than 1. One also expects an association between crash occurrence and presence of the measure, but a negative one. One expects crashes to be less frequent among systems (road users, vehicles, roads) in which the measure has been applied than among those systems without the countermeasure. Consequently, relative risk (or the odds ratio as its approximation) can also be used to quantify the effect of measures. While risk factors lead to relative risks larger than 1, successful countermeasures lead to a relative risk smaller than 1.

#### Preventive fraction

A common measure for *effectiveness* is the preventive fraction E (e.g. Fildes et al., 2015; Zangmeister et al., 2007):

$$E = 1 - OR$$

The preventive fraction describes the proportion of crashes that would be avoided among the crashes critical to the measure in question. It has to be multiplied by the proportion of critical crashes to give an estimate of the percentage of the total crash population that could be prevented by the measure.

#### Crash modification factors and functions

While the epidemiological measures discussed so far are based on the comparison of crashes (or other unfavourable outcomes) to non-crash (or other more favourable outcomes), a lot of research also departs from the comparison of groups where measures have been implemented to groups without the measure. Especially in studies concerning road design, there is a long research tradition which has led to the estimation of crash modification factors (CMFs).

A CMF is the expected number of crashes with a countermeasure divided by the number expected without the countermeasure. To estimate these numbers, accident prediction models are often used that relate road characteristics (traffic volume, number of lanes, etc.) to the number of occurring crashes.

While estimating, one crash modification factor has the longest tradition and is still the most common, it is unsatisfactory when there is systematic variation in the effects of a road safety measure (Hauer et al., 2012). Instead of estimating a single crash modification factor for a given countermeasure, it is often more useful or more reliable to develop a crash modification function, which is a

formula or mathematical equation that can be used to compute the CMF for a specific site based on its characteristics (FHWA, 2010).

Variations between CMFactors can have two reasons (OECD-ITF, 2012). The first is related to estimation errors. If data are poor, if the sample size is small, and if bias and confounding factors are not eliminated, the result will be unreliable. As a consequence, the estimates will vary.
The second reason has to do with the transferability of the CMFs, i.e. the fact that the same treatment can have different safety effects in differing circumstances. The only way to reduce this source of variability is to express the CMFactors as a function of the relevant circumstances.

A countermeasure may also have several levels or potential values, and a Crash Modification Function allows the CMF values to change over the range of a variable or combination of variables (OECD-ITF, 2012).

### 2.2.3    Estimation of injury severity and mitigation

Injury mitigation can much simpler be handled by the use of injury risk function. The injury risk curve gives, for a given severity, the probability to be injured according to the parameter modified by the introduction of the safety system. Their construction is based on the following sequential process:

1.  Selection of the sample: the injury risk curves are generally developed for a given population and selected scenarios (for example pedestrians involved in frontal crash with a passenger car). This selection is very important to ensure consistency of the results. Generally, this sample has to include the population impacted by the safety system.
2.  Imputation of the missing data: the injury risk curves allow to determine a level of risk according to the most relevant parameter, generally associated with the speed. In case of unknown values these cases can – up to a certain degree – be imputed. The method chosen should be well documented as it can strongly influence the results.
3.  Weighting process: most of the in-depth databases are not representative of the national injury-crash census, only some aspects of the studied phenomena are present. In order to ensure a good representativeness of the working sample, a weighting process based on the most relevant parameters (most often severity, crash location and type of road users) is carried out.
4.  Choice of the most relevant statistical method for the establishment of the injury risk curve (survival analysis, logistic, cloglog, etc.).

Once the injury risk curve(s) have been created, they can be used to estimate the probability of the risk of being injured on the simulated cases with the presence of the countermeasure. For that, for each case where the crash cannot be avoided, the new simulation (with the presence of the safety system switch on) gives the new collision speed. Then, we have just to estimate the new probability of injury by reading the value for the given collision speed with the associated injury risk curve (Figure 2.7).

The benefit in terms of injury mitigation is calculated by summing all new figures compared to the initial ones.
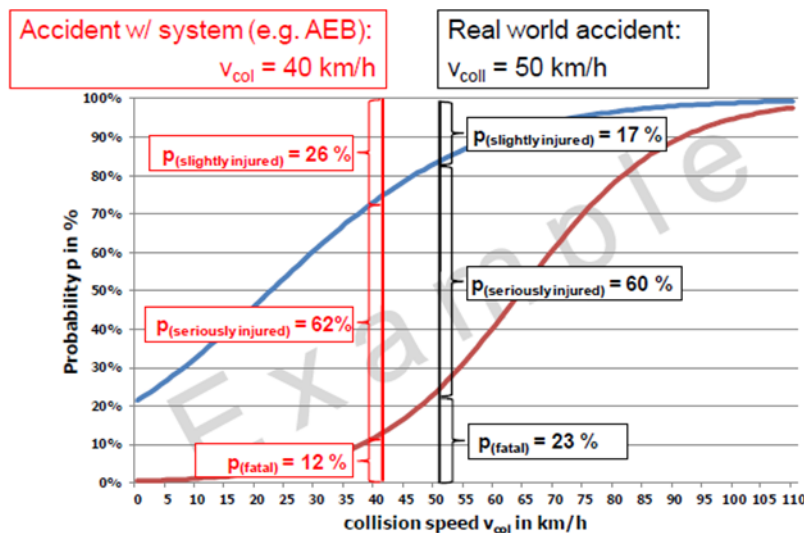
**Figure 2.7** Examples of biomechanics simulation models to estimate injuries (GHBM)

The injury risk functions are statistical measures describing average probabilities for certain injury severity levels. They must not be used for the prediction of a concrete injury severity, for absolute statements in single cases and for special cases that are not in the scope of the calculated functions. These functions are only valid for the type of road users studied and they should not be used for populations with remarkably different characteristics.

### 2.2.4 Organisational risk factors and Performance indicators

Safety performance indicators (SPI) measure factors which are causally related to road crashes and injuries, such as speed levels, drink driving rates or seat belt use. They are an indispensable tool for impact assessment of countermeasures, especially when their effects cannot easily be extracted from the frequencies of crashes or casualties.

Safety performance indicators target various fields:
- Driving behaviour, like speed choice, drink driving or use of seat belts. In addition, attitudes and intentions can be utilised as SPI given that a link between attitudes and behaviour can be established by psychological theory (see also below).
- Infrastructure, e.g. share of median-separated high speed roads or proportion of road section that meet certain safety standards.
- Vehicle, e.g. share of 5-star EuroNCAP cars in a fleet, share of cars equipped with ESP or emergency braking systems.
- Trauma care, e.g. average response time from alarm to treatment; knowledge of first aid.

Behavioural SPIs are frequently used to test impacts – such as of campaigns and enforcement – on attitudes and driving behaviour as intermediate outcomes of these safety interventions.

### 2.2.5 Summary epidemiological methods

By investigating the systematic association of risk factors with particular unsafe outcomes such as crashes, injuries, or unsafe behaviours, the impact of these risk factors is quantified. Road safety measures are considered as protective factors and are tested for systematic association in the other way around: when effective they should be associated with a reduction of unsafe outcomes.

## 2.3    SYSTEMIC APPRAOCHES TO RISK FACTORS AND COUNTERMEASURES

Systemic approaches to crash causation view risk as an emergence within a system made up of humans, technology and organisation (Hollnagel, 1999). Within the context of road traffic crashes, the human is the road user, the technology is the vehicle and related components and organisation are the road environment, weather, communication between the elements of the system as well as factors such as legislation and training. Crashes are the result of failures in the interaction in the system (van Elslande & Fouquet, 2008). Failures or risks can be active or latent in the system. A crash is a result of a failure in the interaction between humans, technology and organisation at a specific point in time and space (active risk or 'sharp end' failure) however other factors may have contributed to this that are more remote in time and space (latent risks or 'blunt end' failures) but add to the risk of the situation (Reason, 1990; Hollnagel, 1999; Hollnagel, 2004). Figure 2.8 illustrates this in the context of road crashes.
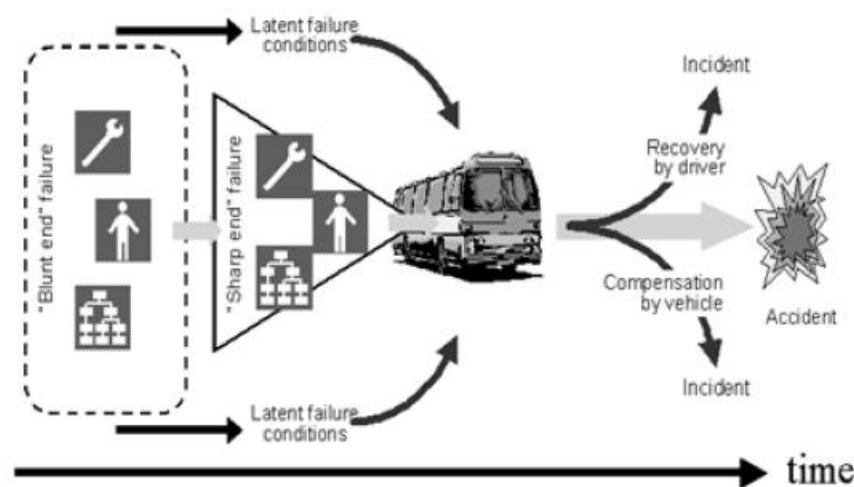


**Figure 2.8** Accident model adapted from Hollnagel (1998), cited in Talbot et al (2013)

The complexity of the regulations and interactions within the driving system requires systemic approaches and models such as the Cognitive Systems Engineering approach (Hollnagel & Woods, 2005). Within the context of road traffic crashes, the human is the road user, the technology is the vehicle and related components and organisation are the road environment, weather, communication between the elements of the system as well as factors such as legislation and training. Crashes are the result of a defective interaction in the system (van Elslande, 2011 see Figure 2.9).

Crash based research has to be augmented by naturalistic driving analyses to gain more precise knowledge of the behaviors of the road users and risk factors to fine tune countermeasures.
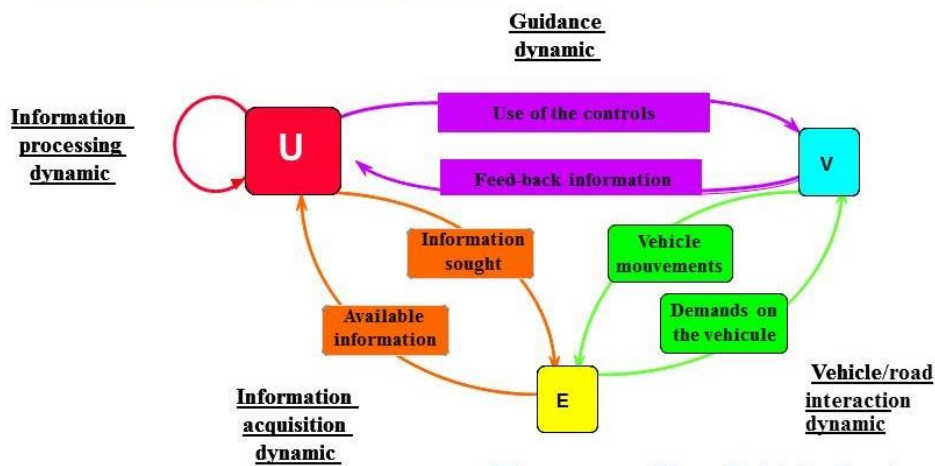
**Figure 2.9 Elementary User-Vehicle-Environment system. Van Elslande, 2001**

### 2.3.1 DREAM Methodology

The Driver Reliability and Error Analysis Method DREAM (Ljung, 2002) originated from CREAM (Cognitive reliability and error analysis method, Hollnagel, 1998) with systemic and human modes of failure (genotypes) as contributive factors. A phenotype is the observable consequence of the failure in the system that leads to the crash and is expressed in terms of time, space or energy. Examples of the DREAM Phenotypes and Genotypes are shown in Table 2.1.

**Table 2.1** Examples of DREAM Phenotypes and Genotypes

| Genotypes | | | Phenotypes |
| --- | --- | --- | --- |
| Organisation/ Infrastructure | Vehicle | Road User | Observable Effects |
| L4 Object on road<br>N1 Time pressure<br>N4 Inadequate training<br>Q1 Inadequate roadside design | G1 Access limitations<br>H1 Permanent Illumination<br>I1 Equipment failure<br>O1 Inadequate vehicle maintenance | B1 Missed Observation<br>C1 Misjudgement of time gaps<br>D1 Priority error<br>E3 Fatigue<br>E4 Under influence of Substance | A1 Timing<br>A2 Speed<br>A3 Distance<br>A4 Direction<br>A5 Force<br>A6 Object |

The DREAM method allows the relationship between causes to be examined as it records the sequence of the prior element is the cause of the next. Subcategories of Phenotypes and Genotypes are assigned to give more detail.

The quantification of a risk factor is based on the frequency of it being identified as causal. In the case of DREAM, the frequencies can be graphically represented by overlaying the individual DREAM charts (See Figure 2.10 for an example).
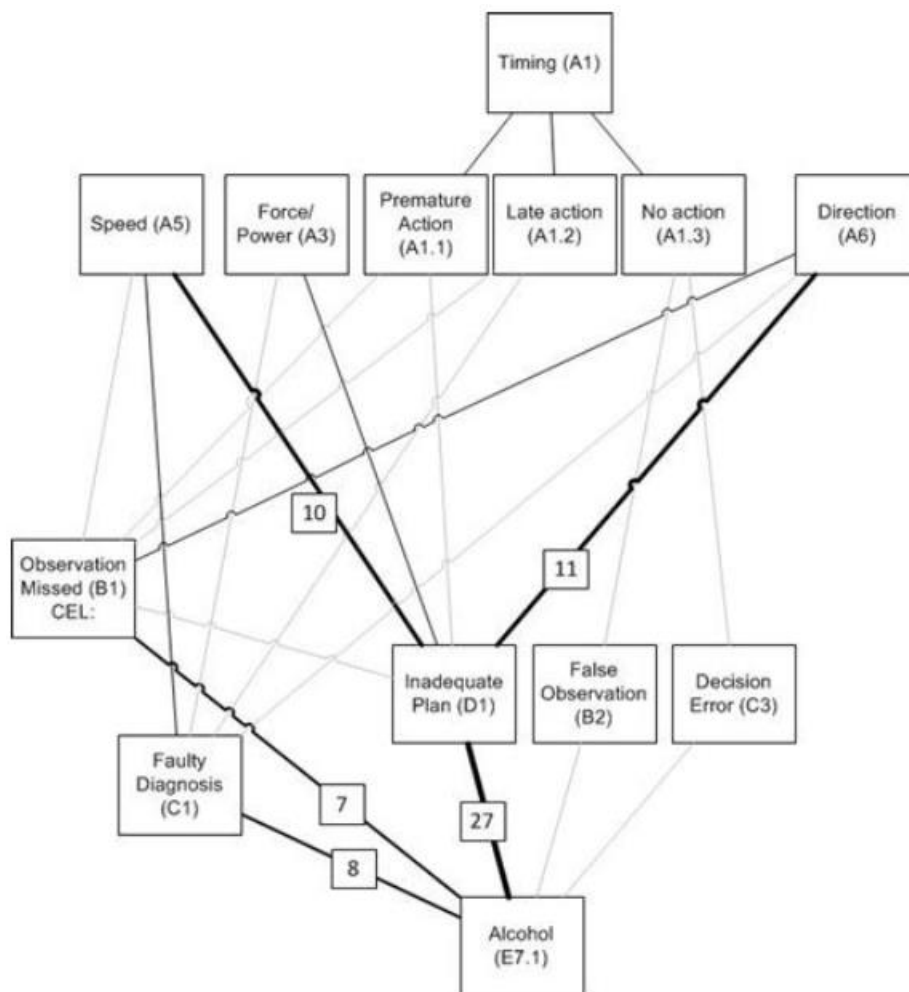
**Figure 2.10** DREAM (version SNACS 2.1) chart for 44 drivers who had a crash whilst being under the influence of alcohol (Talbot and Kirk 2012)

### 2.3.2 Driving behaviors

There is often little or no data available in road crash databases on key risk factors that concern human behaviour. Even when data is available, it is often difficult to extract accurate risk estimates, because there is no data on the related exposure or because it is difficult to identify a pertinent control population. For example, there may be data on the frequency of alcohol or distraction related crashes, but without a measure of the frequency or amount of travel under these factors within the same population, it is difficult to conclude on the risk of driving under the influence of alcohol or under distraction. An alternative to measuring changes in crashes due to some risk factor or countermeasure, it is often easier to measure the effect on the driving behaviour. Driving behaviour can be measured in on-road studies, or – increasingly popular – in driving simulators. There is also a long tradition of research involving psychophysical tasks that measure behaviour that is considered essential for driving.

#### On-road tests

On road tests are conducted by trained experts from multiple disciplines at dedicated sites. In such experimental studies, a vehicle is equipped with instrumentation to take recordings of a variety of aspects of driving (Rizzo et al., 2002). These technologies include GPS, video-cameras, sensors, ac-

celerometers, computers, and radar and video lane tracking systems. On-road experiments attempt to gain greater insights into the factors that contribute to road user crash risk and the associated crash factors at specific conditions; in this sense they may be very similar to naturalistic driving experiments, in which an experimental design is involved for measuring driving parameters under different conditions.

A particular part of on-road experiments concerns pedestrian behavior experiments, in which participants are asked to perform specific walking or crossing tasks at a local level (e.g. an uncontrolled crosswalk (e.g. Chu et al., 2003) or in a small area (e.g. Papadimitriou et al., 2015)).

Another type of on-road test involves the rating of driver behavior by an expert. Examples are studies evaluating a particular type of training in which the driving behavior of the participants is rated subsequently to the training and compared to that of untrained participants (e.g. Boele & De Craen, 2014).

On-road driving studies have the advantage of a higher external validity compared to laboratory experiments and driving simulator experiments. On the other hand, field experiments have been criticized because they do not always collect data over a longer time period and in response to selected interventions. The presence of at least one researcher, to give directions, is also considered potentially bias-inducing (Ball and Ackerman, 2011).

### Driving simulators

Driving simulators have become a very popular tool for carrying out experimental studies in road safety. There are several road safety issues for which it would be unsafe to test them in real traffic. Moreover, the effect of risk factors like alcohol, in-vehicle distraction (e.g. by mobile phones) and response to unexpected incidents are difficult to measure in actual driving conditions, without involving several confounding factors (e.g. weather, traffic etc.) and generating unacceptable risks for the participants. Driving simulators provide a safe environment for the examination of various issues using multiple-vehicle scenarios, where the driver can negotiate very demanding roadway situations. Moreover, greater experimental control can be applied in driving simulators compared with on-road studies, as they allow for the type and difficulty of driving tasks to be precisely specified and any potentially confounding variables to be eliminated or controlled for. Finally, a large range of test conditions (e.g. night and day, different weather conditions, or road environments) can be implemented in the simulator with relative ease, and these conditions can include hazardous or risky driving situations that would be too difficult or dangerous to generate under real driving conditions (Papantoniou et.al., 2015).

Driving simulators, however, vary substantially in their characteristics (i.e. from basic desktop simulators, to quarter- or half-cab simulators - static or dynamic - , and finally to full-cab high fidelity simulators) and this can affect their realism and the validity of the results obtained. Data collected from a driving simulator generally include the effects of learning to use the simulator and may also include the effects of being directly monitored by the experimenter. Furthermore, driving simulators, particularly high-fidelity simulators, can be very expensive to install. Simulator sickness is another problem encountered with simulators and can be is a major source of drop out in simulator experiments. A possible solution is to reduce manoeuvres such as left or right turns.

### Naturalistic driving studies

Naturalistic driving is a relatively new research method for the observation of everyday driving behaviour of road users. For this purpose, systems are installed in participants' own vehicles that register vehicle manoeuvres, driver behaviour (such as eye, head and hand manoeuvres) and external

conditions. In a naturalistic driving study, the participants drive the way they would normally do - and it is expected that after a while they forget that they are being observed - in their own car and without specific instructions or interventions. This may provide information about relationships between driver, road, vehicle, weather and traffic conditions, that are difficult to study by means of traditional research (Regan et.al., 2012), not only under normal driving conditions, but also in the case of incidents or crashes (Van Schagen et al., 2011).

Contrary to driving simulators, where participants are usually tested under different well-defined conditions, this is usually not the case for naturalistic driving studies, where there is no experimental control of the various variables that potentially affect the behaviour of the road user. Risk estimates have therefore to be based on epidemiological study designs, for instance the comparison of critical situations and non-critical ones with respect to the presence of a particular risk factor.

Although it is generally assumed that in a naturalistic study drivers behave as they normally do, this cannot be checked. Moreover, participation is usually based on a voluntary basis, and a self-selection bias in the sense that the volunteers differ in relevant aspects from non-participants cannot be ruled out.

### 2.3.3    Estimation of injury severity with crash tests

More and more we need to go further in the assessment process and consider injury mitigation. For that, it is necessary to use numerical models in order to estimate the new injury level according to the parameter modified by the introduction of the countermeasure. These models can be very sophisticated (such as finite element models) or simple (mathematical equation such as injury risk curves). The finite element models are currently used in biomechanics to simulate body behavior (human or dummy) for crash simulations. From initial conditions given as input data (pulse, limit conditions, deceleration, etc.) the model is able either to directly predict some kinds of injuries, or to calculate predefined criteria that through dedicated risk curves (obtained by real tests) allow the prediction of an injury risk. These models are very time consuming, require a lot of expertise and for most of them the results are not (yet) adjustable to the anthropometry of the real cases.
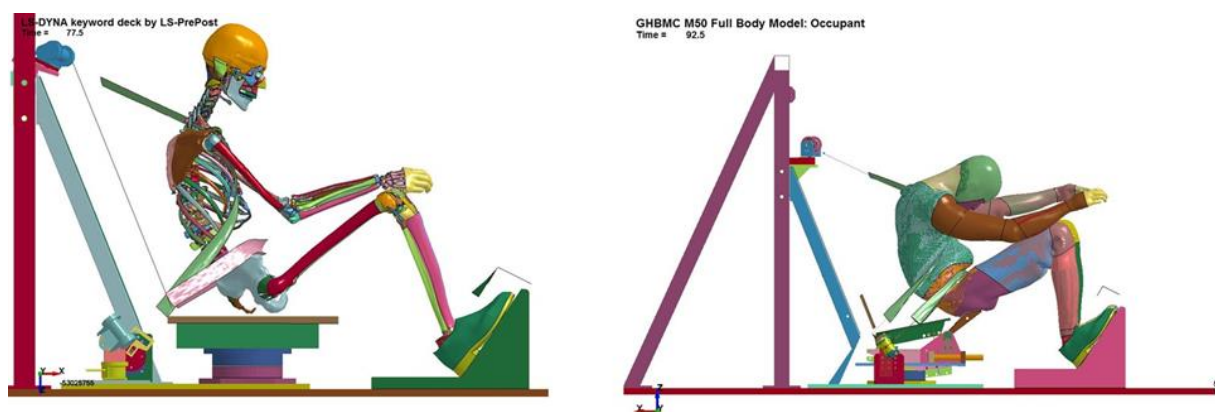


**Figure 2.11** Examples of biomechanics simulation models to estimate injuries (GHBM)

### 2.4    COMPARING THE THREE APPROACHES

Three different approaches to analysing risks and estimating the effects of measures have been discussed. (1) The analysis of crash reconstructions, contributory factors, and scenarios in in-depth crash data to indicate the potential of measures not yet (widely) implemented. (2) In the epidemiological approach, crashes are considered a negative outcome which can be made more less likely by protective layers (the countermeasures) and wholes in these layers (the risks). In the study of risk

factors and countermeasures their presence in the crash population is compared to that in a suitable control group. (3) Systemic models and behavioral studies where the effect of countermeasures is established by investigating second order variables, like driving variables or organizational factors.

Optimally the study of a risk factor of a countermeasure is based on all three approaches. All types of (economic) efficiency evaluation, however, require an estimate of the number of crashes and/or consequences prevented by the measure. Therefore, evaluation methods have to be considered with respect to their ability to produce such an estimate.

From crash reconstructions, such estimates can be produced by means of simulations that introduce the measure on a case-by-case basis. The quality of these simulations depends on the quality of the reconstructions and of the subsequent simulation of the countermeasure's effect. Moreover, the case-by-case work is time intensive. In a simpler version, crash reconstruction gives the target group for a particular countermeasure. Even without indication in what percentage it would have been successful in preventing the crash, knowledge of the target group is indispensable to estimate how many crashes/casualties could be actually prevented. The big advantage of the sequential method is that it offers a way to evaluate vehicle measures that are not yet widely implemented.

In epidemiological research, there are many different ways to arrive at an estimate for crash reduction. The gold-standard is randomised controlled experiments, which are, however, almost impossible to conduct in the field of road safety. As a consequence, practically all methods struggle with the comparability of the cases with countermeasure to the control cases without the countermeasure. There are many statistical ways to deal with this, however the knowledge and use of these methods is not always optimal. The big advantage of the epidemiological method is that it gives the most direct (and therefore reliable) estimates of the proportion of crashes and casualties that can be prevented. These estimates are always related to a particular target group and the number of crashes or casualties prevented needs to take into account the size of that target group as well.

The systemic approach is indispensable for fine-tuning countermeasures. However, it is less suitable for the evaluation of a countermeasure's general effectiveness. To use studies that rely on this approach for a formal effectiveness analysis in terms of casualty reductions, the outcome variables (e.g. a change in speeding behaviour, or the test result from a crash-test dummy) have to be linked to the expected change in the number of crash outcomes. A catalogue of consensual relations between second order outcomes and crash outcomes is therefore the necessary first step, to be able to use these studies in the context of a more general evaluation of measures and risk factors.

# 3 Study Designs in Risk Analysis and Evaluation of Safety Measures

## 3.1 PRINCIPAL EPIDEMIOLOGICAL MEASURES

To estimate the effect that risk factors and countermeasures have on the crash or injury risk, epidemiological methods for the study of injuries or diseases are used. This chapter describes briefly the measures used in epidemiology. For a better understanding there is first a section explaining the principal measures used for studying frequency of road-crash injuries, and a second section about the measures of association used to study risk factors contributing to crashes or injuries or countermeasures preventing them.
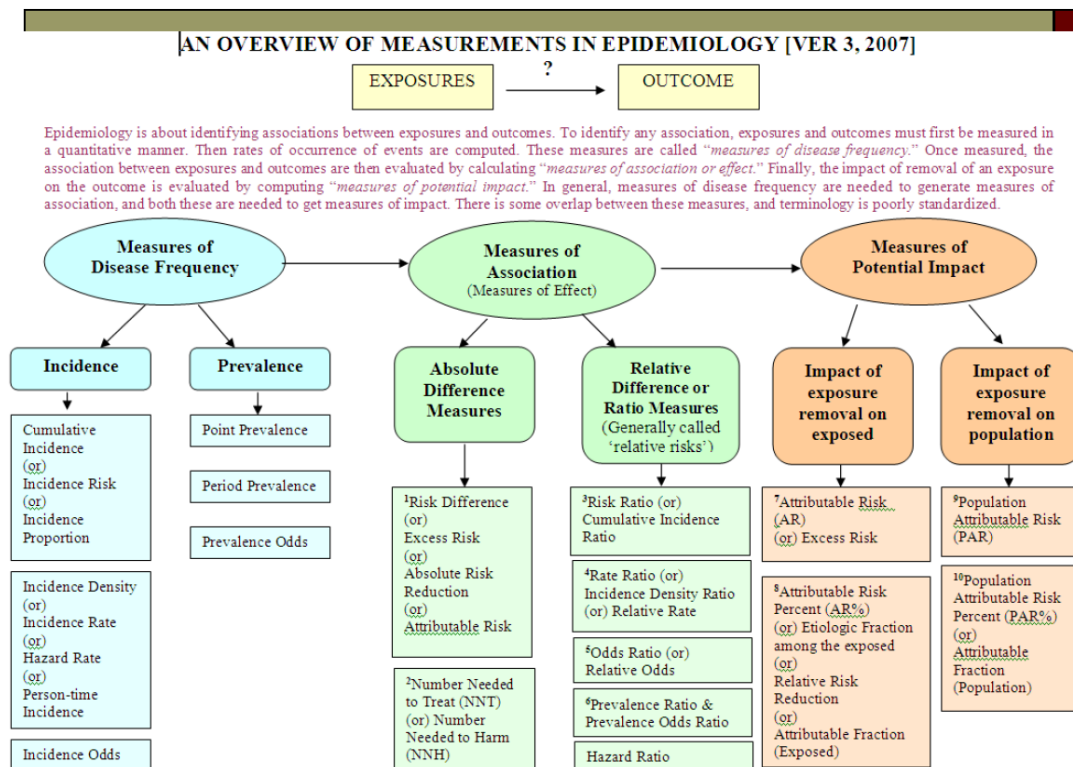


**Figure 3.1** An overview of measurements in epidemiology[1]

### 3.1.1 Principal measures of disease/injury frequency

**Incidence** is defined as the occurrence of **new cases** of a disease/injury arising in a given period in a specified population. Incidence expresses the risk of becoming ill/injured (WHO, 2006). Although it can be expressed in absolute numbers, it is commonly expressed in <u>relative measures</u> such as the **Incidence (I) (also called incidence proportion or cumulative incidence),** which represents the proportion of new events occur in a population in a given time period (e.g. road traffic fatality incidence proportion was 10.3 deaths per 100.000 inhabitants in the European Region in 2013 (WHO 2013)). It measures the denominator only at the beginning of the period, and describes the proportion of individuals that would become ill/injured considering that all of them are susceptible. It expresses the probability that a particular event has occurred after a given time (Moreno-Altamirano et al., 2000).

The **Incidence Rate** (IR) (also called sometimes incidence density) expresses the force/intensity that a disease has to change the health status of a population per unit time. Incidence rate most often used in public health practice is calculated using the average number of persons exposed to risk during this period multiplied by the duration of observation as the denominator (1) (Porta, 2014; WHO, 2006). The denominator is the sum of all the disease/injury-free person-time periods during the period of observation of the population at risk (2).

(1) I = Total number of new cases in the period / Number of people free of disease in the population at risk at the beginning of the period
 (2) IR = Total number of new cases in the period / Average number of persons exposed to risk in that period*duration of observation.

Incidence rate is expressed as the number of crash/injury per person-time, for example 10.3 deaths per 100.000 person*years in the European Region in 2013.

If incident cases are not resolved and continue over time they become prevalent cases (i.e. spinal cord injuries). **Prevalence (p)** is the frequency of existing cases in a defined population at a given point in time or over a given period of time. It is likely used for the study of consequences from road crashes (e.g. 1.1%-1.7% of the U.S. population live with long-term disabilities that result from traumatic brain injuries). For the risk factors that determine the occurrence of road crashes, incident data are more relevant.
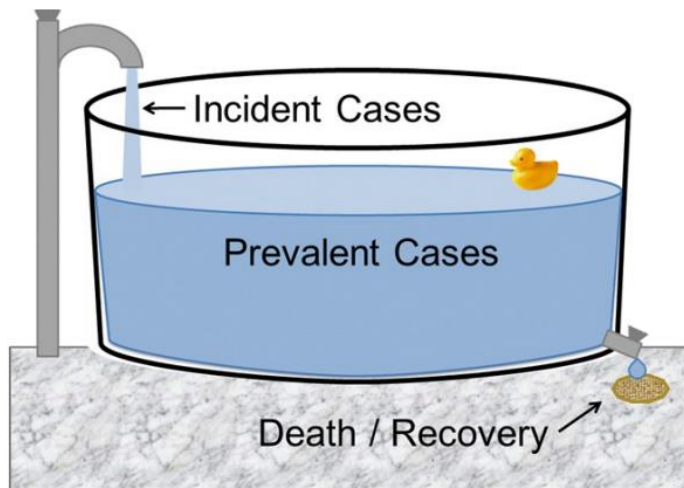
**Figure 3.2** The relationships between incidence, prevalence, risk, and rate. When the tub is the entire population, prevalence is the proportion of the tub filled with water (prevalent cases and incident cases). Risk is the proportion of the tub filled with new, flowing water (incident cases). Incidence rate is a measure of how quickly the water flows into the tub. Prevalent cases leave the prevalence pool by either recovery or death

### 3.1.2    Principal measures of association

In order to study whether prevalence or incidence of the outcome is higher in the presence of any risk factor more than in its absence <u>measures of association or effect</u> are needed. In the same way these measures are necessary to study whether the outcome is related to the *exposure* to a particular countermeasure. Please note that while in the field of road safety research the word "exposure" is often used to describe traffic volume (in the sense that all people who travel on the road are exposed to the risk of a crash), in the present guidelines the word exposure is used in its more general meaning as being exposed to any kind of risk or protection factor.

Epidemiological measures of association aim to compare the incidence or prevalence of an outcome in a group exposed to a concrete factor with the incidence or prevalence in a non-exposed group. The type of association measure used depends on the design of the study. They measure the magnitude of the observed difference. There are two types of association measures: <u>measures of ratio (relative)</u> and <u>measures of difference (absolute).</u>

**Ratio measures**

The ratio measures determine differences between groups with or without the exposure to a risk factor or countermeasure. The main ratios used to determine whether there is an excess risk associated with a given exposure are **Risk Ratio (RR), Exposure Odds Ratio (OR), and Prevalence Ratio (RP)**.

**Risk ratio or relative risk (RR)** is calculated dividing the incidence (risk) in the exposed group by the incidence (risk) in the non-exposed group. It describes how many times more likely the outcome is for exposed individuals relative to those unexposed. It is equal to 1 if the risk is the same in both groups (if the outcome is independent of exposure to a risk factor or countermeasure). The ratio is larger than 1 if exposure is positively associated with having the outcome and smaller than 1 if the exposure can be considered a protection factor from the outcome.

In fact, the incidence is expressed in two ways in cohort study: cumulative incidence (which is a proportion called risk) and incidence rate (or incidence density) called rate (person-time rate = number of events divided by the person-time at risk). If we follow during a period a population exposed to mobility, the incidence rate would be the number of people injured in a crash divided by the total time free of injury of the population (sum of the time until the crash of all persons, plus all the time of those who did not experienced a crash)

The ratio between two cumulative incidences (risk in exposed divided by risk in unexposed) gives the relative risk (or risk ratio). While the ratio between two incidence densities (rate in the exposed divided by rate in the unexposed) gives the incidence rate ratio (IRR or rate ratio).

**The Odds Ratio (OR)**

The calculation of the incidence and the association measures calculated using the incidence (RR) are the most reliable measures in epidemiology. Unfortunately, only cohort (follow-up) studies enable its calculation, and these studies are not usually feasible. In case-control and cross-sectional studies, where only prevalence data are available, only the calculation of ORs are possible.

An odds ratio (OR) is a measure of association between an exposure to a risk factor or countermeasure and an outcome. The OR represents the odds that an outcome will occur given an exposure, compared to the odds of the outcome occurring in the absence of that exposure. It is obtained when the design of the study is case-control or cross-sectional, when no incidence is available due to impossibility to carry out a cohort or follow-up study as is usually the case for road safety studies, because road crashes are relatively rare events (Schiaffino, 2003). It can be mathematically shown that OR is approximately the same as RR when the prevalence of disease/injury is low, say less than 5%.

## 3.2 A TAXONOMY OF STUDY DESIGNS

A compact overview of the most common study designs is given and their typical biases are briefly mentioned. Each study design is characterised by a number of principles (addressing exposure to risk/measure; experimental vs. observational; presence of control group; time dimension) which are summarised in the figure below.

Figure 3.3 gives an overview of the categorisation of studies as discussed below. The dimensions according to which study designs can be characterised are described below along with brief information concerning their principal application. While this section is mainly focused on differentiating between different study designs, a more detailed description with attention to good practices and typical biases is given in the following sections.
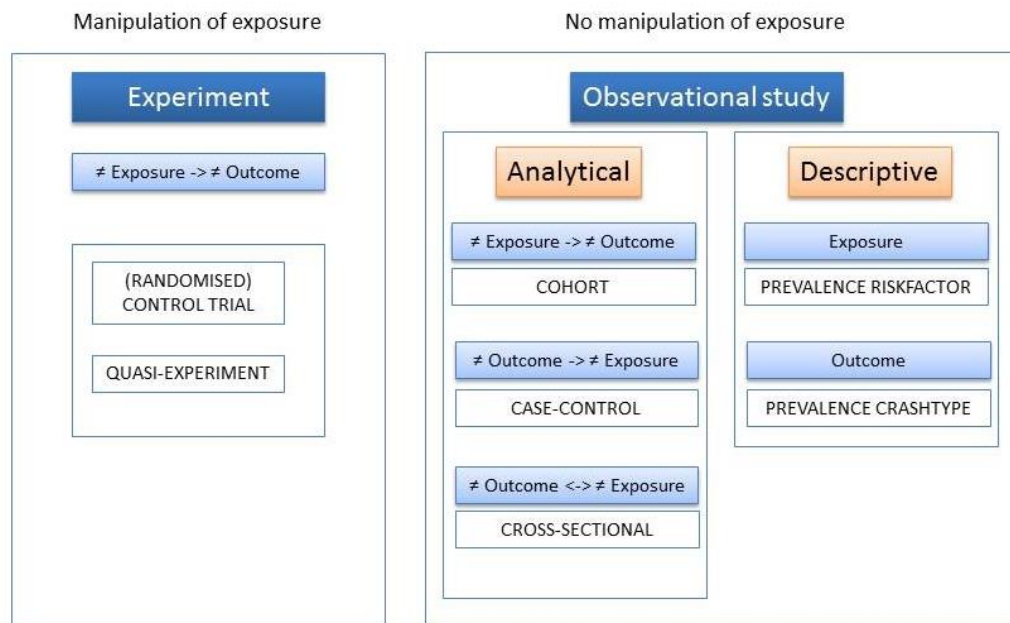
**Figure 3.3** Study designs in risk analysis and evaluation of countermeasures

### 3.2.1   Outcomes versus exposures

Study designs in road safety are closely related to those in epidemiology. As in epidemiology, the bulk of research is concerned with establishing the relationship between exposures to a certain risk factor or countermeasure and outcomes. **Outcomes** typically concern crashes or injuries and in particular, their (absolute/relative) numbers, types and severities. Apart from such direct indicators of road safety, variables like driving skills (e.g. expert rating), attitudes towards safe behaviour (e.g. willingness to drink and drive) or even physiological (e.g. eye-movements, electro encephalogram) and physical measures (e.g. km/h) can also be considered as outcomes, since they are known or can reasonably be assumed to influence crashes or injuries (numbers/types/severities). **Exposure**, in the context of road safety, either refers to exposure to risk factors or exposure to countermeasures. In the latter case, it might sound more natural to speak of "implementation of countermeasures" (e.g. roundabouts) or "use of countermeasures" (e.g. helmets), but using "exposure" helps to see commonalities with designs in studies on risk factors and the epidemiological literature. Depending on the study design, both outcome and exposure variables can take on:

- binary values (e.g. crash: yes/no; roundabout: exposed/non-exposed),
- unordered categorical values (e.g. injury nature; signalization types),
- ordered categorical values (e.g. injury: AIS1..-AIS5[2]; sleepiness scale values),
- counts (e.g. crash count; number of cars equipped with feature x)
- or other continuous values (e.g. crash rates, lateral deviation; blood alcohol concentration).

### 3.2.2   Experimental versus Observational studies

As in epidemiology, we can make a principal distinction between **experimental** and **observational** studies in road safety. **_Experimental studies_** are concerned with the effect of a particular manipulation of the level of exposure to a risk factor or countermeasure has on outcomes. Hence, the starting

---

[2] AIS= Abbreviated Injury Scale

point is always a number of different exposure levels that are actively imposed within the study or have been imposed by another party. In the clinical domain the word "experiment" is often reserved to the purest type of experiment, namely randomised controlled trial experiment (see below). Here we use the broader sense of all studies where a researcher has tried to create comparable conditions in which a risk factor or countermeasure is either applied or not applied (or applied to different degrees).

In **observational studies**, on the other hand, there is no intervention whatsoever, neither by researchers nor by any other party. The natural occurrence (distribution) of exposure to a risk factor or countermeasure and outcome is studied.

### 3.2.3    Descriptive observational studies

As mentioned above, most studies in road safety look at the relationship between different exposures to risk factors or countermeasures and different outcomes. This is always true for experiments. However, not all observational studies have this property. If they do, they are called "***analytical***" observational studies. These are to be contrasted with "**descriptive**" observational studies. Descriptive observational studies typically involve risk factors, rather than countermeasures, and merely describe the presence (or distribution) of exposure to risk factors in *either* a crash/injury *or* no-crash/injury population. Hence, there is no comparison of different outcomes, merely an *observation* of exposures. Road-side surveys on drinking and driving and in-depth crash analyses reporting on the frequency of occurrence for contributing factors are typical examples of this. Although descriptive observational studies do not allow us to quantify the relation between exposure to a risk factor or countermeasure and outcomes, they still provide valuable information for road safety management. For established risks, such as drinking and driving, for instance, studies that merely monitor exposure to these risks in the general population are clearly important for prioritising road safety campaigns.

### 3.2.4    Experimental versus Analytical observational studies

When the impact of risk factors or countermeasures on road safety is either unknown or needs quantification, an analytical approach is needed, i.e., an analysis of the relationship between different outcomes and different exposures to risk factors or countermeasures. As mentioned already, experiments are always "analytical" in this sense. ***The difference between experiments and analytical observational studies is that observational studies respect the natural occurrence of exposure in the population, whereas experiments always involve some kind of manipulation of exposure levels.***

In theory, both analytical observational and experimental designs can be applied in the context of risk factors and countermeasures, but in practice, risk factors are more often studied using an observational approach, whereas countermeasures are often evaluated using (quasi-)experiments. The reason is that imposing risks in an experimental context is usually unethical, but this is not a general rule (e.g. imposing sleep deprivation is not necessarily unethical in a driving simulator study). Countermeasures, on the contrary, often lend themselves well for experiments or sometimes even demand an experimental approach. Some countermeasures are particularly easy and non-expensive to manipulate (e.g. taking a 15 minutes nap). Sometimes, experiments are also preferred to study risk factors or countermeasures because it is difficult to observe the natural occurrence in a non-experimental population (e.g. phone use or sleep deprivation in simulators or different vehicle designs in crash tests). Finally, more so than for risk factors in road safety, there can be a clear onset of the exposure to a countermeasure (e.g. the implementation of a roundabout; a law that is passed). Such a situation allows for a specific kind of experimental design where outcomes are compared for the same units of analysis (e.g. intersections) *before* and *after* the onset of exposure to a countermeasure (e.g. conversion to a roundabout; see below). Again, experiments are not mandatory in the

context of countermeasures. For instance, an analytical-observational approach might be preferred for clearly observable countermeasures, such as seatbelt and helmet wearing, because it is feasible to collect exposure to these countermeasures and outcome data on a large scale (e.g. crash reports linked to hospital records).

### 3.2.5 Experimental studies: Controlled versus quasi-experiments and randomisation

More so than other domains, road safety research often deals with *quasi-experiments*. This is generally the case when investigators have no control over the assignment of the different exposure levels that are imposed. Typical examples are studies that look at the reduction of crashes on road segments where certain infrastructural countermeasures have been implemented. Usually, investigators are not involved in the selection of segments where infrastructural changes are made, since this is governed by local authorities. Another example is a study where drivers are recruited on the basis of whether or not they completed an advanced driving course and the investigator compares driving skills between exposed and non-exposed drivers. Again, if the assignment of the exposure (completion of driving course) was not controlled by the investigator we are dealing with a quasi-experiment.

The critical aspect of quasi-experiments is that they inherently lack randomisation, more specifically, there is no random assignment of exposure levels. Hence, they require special care for potential sources of selection bias. In the case of infrastructural countermeasures, for instance, it is often the case that the selection of road segments is biased towards sites with particularly high crash rates. This raises two problems. First, if an effect on crash rates is observed, this result might be poorly generalizable, more specifically, when there are systematic differences in the characteristics of segments with high and low initial crash rates which interact with the effect of the countermeasure. The second problem is known as "regression to the mean". When high crash rates are observed, there is always the possibility that these happen to be extreme values drawn from an underlying continuous probability distribution with a mean that is in fact much lower. In that case, the reduction of crashes following infrastructural changes might simply represent a more typical draw from the same underlying distribution as in the before-period. Also, in the example of drivers recruited on the basis of whether or not they completed an advanced driving course, selection bias can be a substantial problem especially if the initial choice to participate in such training was made on a voluntary basis. All these biases form violations of the assumption that the two conditions that are compared differ *only* with respect to the exposure to the risk factor or countermeasure, while everything else is the same (ceteris paribus).

We speak of *controlled experiments* whenever investigators are in control of the assignment of exposure levels. Importantly, this does not automatically mean that assignments occur randomly. While so-called "**randomised control trials**" are generally considered as the gold standard, it is sometimes impractical or even impossible to achieve full randomisation of exposure levels. Moreover for small group-sizes, *matching* might lead to better results than random assignment.

Experiments, either controlled or quasi-experiments, can further be distinguished on the basis of whether or not the effects of different levels of exposure to a risk factor or countermeasure on the outcomes are measured across different analysis units. For instance, when driving test results are compared between drivers who drank a solution containing alcohol and other drivers who received a placebo, one can either give one group alcohol and another group the placebo (-> between group), or test every driver twice, once with alcohol and once with placebo (-> repeated measures). For between group experiments it is highly important to ascertain that there are no other systematic differences across the exposure groups that could explain observed differences in outcome variables (e.g. age, sex, etc.). Random assignment to the different exposure groups offers protection against such confounds, but especially in smaller samples random assignment does not guarantee that

groups are comparable with respect to all relevant non-experimental variables. For repeated measures experiments the order of testing conditions must be carefully considered.

### 3.2.6    Repeated measures: Before and after versus cross-over

In repeated measures experiments, different levels of exposure to a risk factor or countermeasure are sequentially imposed to the same units of analysis. Here we can distinguish two types: "before and after studies" and "cross-over studies". As noted above, **before and after** designs are useful when there is a clear onset in the exposure, as is the case for infrastructural countermeasures, legislation, campaigns, etc. Here, it is a natural choice to follow the same analysis units (e.g. intersections, driver population) before and after they were exposed to the risk factor or countermeasure. The critical property of before and after studies, is that the order of the repeated measurements is fixed, i.e., outcomes are always measured first without and then with exposure. This implies that the experimenter needs to ascertain that any observed changes in outcomes are not due to other variables whose values are confounded with the before-after time windows. For infrastructural countermeasures, for instance, it is important to check whether a reduction in crashes can be explained by a reduction in the traffic volume. Note that this potential source of bias is different from those associated with the typical quasi-experimental nature of these studies discussed above.

To deal with order bias (and selection bias), before and after results are often compared with some form of control measurements. These can be measurements of outcomes for similar, but "untreated" analysis units (e.g. intersections) during the same before and after time windows. Other studies use a Bayesian approach, wherein measurements of outcomes at untreated sites are used to build a probabilistic model. The observed changes in outcomes at treated sites are then evaluated in the light of the predictions of this model. These control techniques are also known as "difference-in-differences".

Repeated measure experiments in which the order of measurements is not fixed are called ***cross-over designs***. The number of different exposure levels can be considerably larger than two. For instance, if sleep deprivation conditions are tested in a simulator study, the levels can be expressed in number of hours of sleep (1, 2, 3, 4...). One always needs to be attentive of potential spill-over effects from one exposure level to another. Randomisation of the specific sequence of exposures across analysis units is highly recommended to avoid systematic order effects, but does not eliminate spill-over effects, which might cause a considerable amount of variability in the outcomes that would not be present in a between-group design.

### 3.2.7    Analytical observational studies

As defined above, in analytical observational studies, the relationship between exposures to risk factors and countermeasures on the one hand and outcomes on the other hand is investigated while respecting the natural occurrence of exposure levels. The investigator does not intervene to impose exposure levels, neither does he/she act as if this was the case in a quasi-experiment. The difference is not always clear, because exposed and un-exposed cases in quasi experiments are sometimes selected from a database. However, in an experimental design, the researcher first demonstrates that two comparable groups (exposed, non-exposed) have been selected from the database and in a second step compares the outcomes of these two groups. In an observational design, the relation between exposure and outcome is expressed in a measure of association (e.g. odds-ratio or correlation).

Analytical-observational designs are often applied to risk factors when it is unethical and/or impossible to impose exposure levels (e.g. drug addiction, socio-economic status). For countermeasures, analytical-observational designs are useful when exposure is variable/transient, i.e., not strictly de-

fined in space and time (e.g. use of cycling helmets). In correspondence with the epidemiological literature, three families of analytical-observational designs can be distinguished: cross-sectional, cohort and case-control. At a general level, these designs can be characterised by a different "directionality". In cohort studies investigators start from different a priori levels of exposure to a risk factor or countermeasure in the population and monitor differences in outcomes (exposure -> outcome). In the case-control design, one starts from different outcomes in the population and studies differences with respect to the distribution of exposure levels (outcome -> exposure). In cross-sectional designs, finally, the distribution of exposure and outcome is considered simultaneously (exposure <-> outcome).

### Cross-sectional

In a cross-sectional design the investigator "cuts through" a target population at a specific moment in time and looks at the level of exposure to a risk factor or countermeasure and the outcome for each sampled member. Typical examples are in-depth crash databases containing information on outcomes (e.g. injuries) but also the exposure to risk factors (e.g. road conditions, sobriety, pre-crash speed…) and measures (seat-belt use, ABS…). Another example is a road-side survey where cyclists are interviewed about their crash history and where at the same time the prevalence of electrical bikes is registered among those interviewed. Whereas cross-sectional designs are very useful as descriptive tools (to determine the prevalence of exposure and outcome), using them for analytical purposes it is often challenging. Indeed, when the analysis shows a significant association between exposure and outcome it is often difficult to infer causality since exposure to a risk factor or countermeasure typically correlate with many other uncontrolled characteristics in the sample.

### Cohort studies

Cohort studies start with the identification of a target population which, at a given initial point in time, is *not* associated with a certain negative road safety outcome (e.g. not injured in a traffic crash). This population (or a sample; "panel") is then followed over time while monitoring the occurrence of the outcome of interest. The relationship between exposure to a risk factor or countermeasure and outcome can be established in two ways. In the first approach, different cohorts are compared that have been selected with respect to known a priori exposure levels (e.g. children living in an urban environment and children living in a rural environment). These are referred to as cohort studies with external controls. In the second approach, a single cohort is selected with variable a priori exposure levels. Eventually, the distribution of these levels is compared between panel members that suffered "unsafe" outcomes and those with "safe" outcomes. These cohort studies thus use internal controls. The main strength of a cohort design is that the association between exposure to a risk factor or countermeasure and outcome is monitored through time and hence, stronger claims can be made about causality. However, regardless of the type of controls, there is always a risk that exposure levels are correlated with other panel characteristics which are in fact the true cause of the observed distribution of outcomes. In road safety research, cohort studies are often performed retrospectively, i.e., on the basis of existing data, since the outcomes of interest are rare and therefore long periods of data collection would be needed in a prospective design. A prospective cohort study of cyclist safety was reported by Poulos et al. (2015).

### Case-control studies

In a case-control design, the investigator identifies two populations: one with an outcome of interest ("cases") and one without the outcome of interest ("controls"). In each population for which exposures to a risk factor or countermeasure are measured the association between exposure and outcomes is determined. An example is a study where blood alcohol concentrations are compared across a population of drivers that was admitted to a hospital after a crash and a population of driv-

ers that was not involved in a crash. The fact that outcomes are defined as grouping variables is a critical distinct feature of the case-control design and is especially advantageous when the natural occurrence of the targeted outcomes is rare. In controlled/quasi-experiments and cohort studies with external controls, grouping is instead based on exposure levels. When an "unsafe" outcome is rare (e.g. head injury) the data volume that is eventually obtained can then turn out to be very small as opposed to "safe" outcomes. Hence, statistical power might be too low to detect a real association with exposure to a risk factor or countermeasure . The main quality of case-control designs is that for rare outcomes they allow the collection of much more data on the exposure to a risk factor or countermeasure . Hence, they offer the potential to provide a deeper insight in exposure-outcome associations than other analytical-observational techniques.

The quality of the conclusions in a case-control design relies heavily on whether cases and controls are comparable with respect to all relevant characteristics, other than exposure to the risk factor or countermeasure under study. This is often very difficult to achieve in practice. In a "matched case-control" study, the investigator makes assumptions about a number of relevant secondary characteristics (age, sex, etc.) and equates cases and controls with respect to these variables. This can be done on a one-to-one/one-to-many basis or at the group level. This is not always feasible in practice, however. The success of such matching also relies on the initial intuition of the investigator about relevant secondary characteristics. Hence, an "unmatched design" is a more popular choice. The lack of matching can be compensated by increasing the sample size and/or incorporating secondary characteristics (along with exposure levels) in an explanatory model of the binary outcomes (1=case; 0=control). For example, data from a large-scale hospital survey where cases are screened is coupled to data from a large-scale road-side survey. A case-control design can also take the form of a "case-crossover design". Here, cases function as their own controls. In longitudinal data, for instance, one can retrospectively select, for each case, a time window where the outcome of interest did not occur and compare exposures with the "case window". Although this approach has the advantage of keeping inherent characteristics of analysis units constant, it is only useful when exposure to a risk factor or countermeasure is transient and also not chronically recurring or habitual. These conditions are rarely met in road safety research. Cell-phone use while driving, for instance, is a transient risk factor, but is usually habitual. Hence, cell-phone use before a crash occurred will be correlated with cell-phone use in a non-crash situation for the same individual, resulting in an underestimation of the risk of cell-phone use while driving in a case-crossover design.

### 3.2.8 Ecological designs

Ecological designs aim to map the effects of specific risk factors or countermeasures onto predefined spatial and/or temporal units. The general purpose is thus to take into account local dependencies in exposures and outcomes. The design can be used for descriptive or analytical purposes and can in theory be applied within either a cross-sectional, cohort, case-control or even quasi-experimental approach. An example of the latter is a study where the effect of a national campaign on drinking and driving is evaluated using a before-after design with respect to different regions.

### 3.3    EXPERIMENTAL STUDY DESIGNS

In experimental studies the exposure to a risk factor or a countermeasure is assigned to the units under study and the outcome with respect to some road safety risk factor or measure is then compared between exposed and non-exposed subjects. In randomised control trial experiments the units or individuals are assigned to the conditions at random and in quasi-experiments this assignment is not under the control of the researcher. The main threat to experiments is their validity (experimental setting, tasks) and generalizability (sample selection). The most frequent type of quasi-experiments are before-and-after studies. The main

biases are regression-to-the-mean, long term trends, and exogenous changes (e.g. traffic volume). Bayesian or Empirical Bayes methods help to correct for first cause of bias. A control group is the minimum requirement to control at least for long term trends. Regressions are useful to control the influence of exogenous variables such as traffic volume.

### 3.3.1    Randomised control trial experiments

In **the experimental studies** the investigator is the one who assigns the exposure to a risk factor or countermeasure, randomly or non-randomly. **Randomised control trials** are the gold standard of epidemiological studies; however they are unethical when the exposure is harmful such as injuries. Consequently, although they are the studies which provided the best evidence on causality, they are rarely used for road-traffic injuries studies. Even so, they can be used for the study of treatments after road injuries occur. Other experimental designs in road safety are laboratory experiments, driving studies (simulator or road), and field studies.

#### Experimental setting

*Laboratory experiments* in road safety refer to studies in which participants are asked either to perform specific tasks, or verbally indicate their opinions, intentions or actions in given scenarios, within a controlled (and often virtual) environment of a research laboratory or similar facility. Scenarios may be presented to participants in printed form (e.g. lists, sketches, pictures etc.), however in the recent years it is very common to use more or less sophisticated audiovisual and interactive means, such as videos. Effects in laboratory experiments are typically measured as counts of the "safe choice" versus a less safe choice or in terms of response times to the stimuli presented.

*Driving studies* are experiments where the driving behaviour is measured in response to the exposure to a particular risk factor (e.g. distraction, inexperience, or old age) or a particular countermeasure (e.g. a training or a short break). As described before, driving simulators have become a very popular tool for carrying out experimental studies in road safety, because they offer a good compromise between a realistic driving task and a well-controlled environment. In comparison to simulated driving, on-road examination of driving behaviour in response to different conditions (e.g. training vs. no training) is much more costly and difficult to standardise. While the resulting variables in on-road driving tests are mostly subjective scores by one expert judge, driving simulators offer several objective measures like lane-deviation, variation in speed, gap-acceptance, response times to incidents, etc.

*Field experiments* are studies in which the implementation of a particular treatment is done "for real" but under the regime of an experimental design. An example would be to select crossings *at random* for a particular treatment (e.g. change into a roundabout) and then compare them to those crossings that were not treated. The important feature is that the decision which crossing would be treated and which would not had *nothing* to do with the characteristics of the crossing (i.e. was taken at random). The effect size in such a study is given by a CMFactor or an odds ratio exactly as they are in before-after studies in which the allocation has not happened at random (which are in fact quasi-experiments).

Within the context of naturalistic driving studies, experiments are in principle possible. A safety system can be tested by selecting some of the participants for which to install the system and these participants are then compared to those that do not have the system (e.g. on the number of critical incidents). Again, the critical characteristic is that it is decided at random in which cars the system is installed and in which it is not. If the decision is based on the convenience to implement the system or on the choice of the participants, or on any other characteristic that could be related with the outcome, the study is not an experiment anymore, but a quasi-experiment.

Experiments can be classified in two groups: Randomised and non-randomised experiments (quasi-experiments). In fact, the first principle of an experimental design is **randomisation**, which is a random process of assigning treatments to the experimental units. Every experiment relies upon selecting subjects (persons, vehicles, crossings, etc...) and placing them into groups, with the objective to form groups that are equal with respect to all characteristics except for the one under investigation. A researcher may fail to take into account all of the potentially confounding variables. The random process implies that every possible allotment of treatments has the same probability. The purpose of randomisation is to remove selection bias and other sources of extraneous variation, which are not controllable (Boyle, 2011).

For example, in a randomised experiment with two conditions A & B, the units (participants) will be randomly assigned to the conditions by means of a procedure often referred to as "repeated fair coin-tossing". The main drawback is the possibility of imbalanced group sizes in small experiments (e.g. less than 200 subjects). Consequently, other randomisation procedures (restrictive or adaptive) may be opted for. The most common alternative is **block randomisation**, in which a "block size" and "allocation ratio" (number of participants in one group versus the other group) are specified, and participants are allocated randomly within each block. A special case of block randomisation is **random allocation**, in which the entire sample is treated as one block (Schulz & Grimes, 2002; Lachin et al., 1988); these procedures are however not without limitations (e.g. selection bias).

Experiments are also often classified in **blinded (or masked) or unblinded** experiments, referring to the procedures that may prevent participants, researchers or outcome evaluators from knowing which conditions were experienced by each participant. This type of experiment is recommended in cases where the experiment outcomes cannot be measured objectively (e.g. an expert judging the driving behaviour) but less important when experiment outcomes are assessed on the basis of clearly defined measures. Another influence could be **unintended side effects** on the participants (e.g. when they become aware that they are in the "risk" condition, this might influence their behaviour). Unintended side effects are all effects that are not intended by the treatment, but that may arise as a result of behavioural adaptation among subjects. Such effects were observed for the first time in industrial experiments designed to enhance productivity and were named "Hawthorne" effects. In medical trials, a well-known unintended effect is the placebo effect. The safest way to avoid such side-effects is to keep the participant in the dark of which condition he is assigned to (e.g. by giving a placebo to those not treated). An experiment in which neither the participant nor the experimenter who measures the outcome is aware which condition is tested, is called a **double-blind** experiment. However, blinded experiments are not always possible, e.g. when the active involvement of the participant is required, which is often the case in road safety experiments (e.g. in driving simulator experiments, in field trials, etc.).

Next to the allocation of the participants, experiment designs can be broadly classified according to two main design characteristics (see Figure 3.4 for examples):
- **Within- or between-subjects variables**: Experimental research always compares the same type of outcome under different conditions of a particular factor (e.g. darkness vs. light; alcohol intake vs. no alcohol intake; ESC vs. no ESC[i], etc.). If all conditions are tested on the same subject (participant), this is called a within-subject factor. If different subjects are assigned to each condition, it is called a between-subject factor. Often between-subject and within-subject factors are combined in one experiment, because that there are variables which are by nature between-subject (e.g. gender, as a participant can be either male or female) while others can be within-subject (e.g. driving with distraction or without distraction – a condition that can be tested for all subjects). A mixed design includes both within-subjects and between-subjects factors.

- **Full factorial or fractional factorial design**: many experiments are based on a combination of levels of different variables of interest. The complete combination of all levels of the variables of interest results is a full factorial design. In several cases, however, a fractional factorial design may be opted for, by eliminating some of the combinations of levels of the variables examined, on the basis of appropriate criteria(McLean and Anderson, 1984), especially when the number of variables is high, resulting in an unmanageable full factorial design. More specifically, a fractional factorial design is most often based on a full factorial design of some key variables of interest, complemented with selected combinations of these variables with other variables of interest (Montgomery, 2000).

Each type of design has its own advantages and limitations, as well as different requirements in terms of sample power, analysis methods etc. A detailed presentation of the features and methodologies related to each type of design is beyond the scope of this report (e.g. see Boyle, 2011). However, the design principles concerning each type of design should be explicitly taken into account for the design of any experiment.

*Within-subject experiment design:*

| Within-subject variables | Subjects: males, <25 years old |
|---|---|
| Undistracted driving | √ |
| Driving while using mobile phone | √ |
| Driving while talking to passenger | √ |

*Between-subject experiment design:*

| | Between-subject variables | | |
|---|---|---|---|
| Driving task: | <25 years old | 25-55 years old | >55 years old |
| Undistracted driving | √ | √ | √ |

*Mixed design, full factorial (2 variables×3 levels, $3^2$ combinations):*

| | Between-subject | | |
|---|---|---|---|
| **Within-subject** | <25 years old | 25-55 years old | >55 years old |
| Undistracted driving | √ | √ | √ |
| Driving while using mobile phone | √ | √ | √ |
| Driving while talking to passenger | √ | √ | √ |

*Mixed design, fractional factorial ($3^2$-2 combinations)*

| | Between-subject variables | | |
|---|---|---|---|
| **Within-subject** | <25 years old | 25-55 years old | >55 years old |
| Undistracted driving | √ | √ | √ |
| Driving while using mobile phone | √ | √ | |
| Driving while talking to passenger | √ | √ | |

**Figure 3.4** Experiment designs at the example of distracted driving: between- or within-subject, full or fractional factorial

**Threats to validity**

The main threat (Caird and Horrey, 2011) is the presence of **carryover or order effect,** when one experimental condition influences the results in another. Whenever subjects are tested in more than one condition (as they do in within-subject designs) there is a possibility of carryover effects. For example, in a driving simulator study, participants may be tested in three conditions: undistracted

driving, driving while conversing with a passenger and driving while using a mobile phone. In this case, it is possible that, driving the "undistracted" trial may make it easier for participants to subsequently drive the "conversing with passenger" trial, and consequently their driving performance while conversing with passengers may not be measured as impaired, i.e. a learning effect is involved). On the other hand, when driving the third "using a mobile phone" trial, a further learning effect may be involved or - on the contrary - a fatigue effect may be involved; in the latter case, participants may be overwhelmed by the duration of the tasks and exhibit impaired driving behaviour not due to the complexity of the task itself. These carryover effects are often difficult to identify, measure and isolate.

In some designs, order may be the focus of the analysis (e.g. skill acquisition at the effects of practice) and therefore carryover effects are not a concern. In most designs, a between-subject design will be the only way to eliminate carryover effects, but this will not always serve the purposes of the analysis. Consequently, a typical design strategy to deal with carryover effects is **counterbalancing**.

The simplest type of counterbalanced measures design is used when there are two possible conditions, A and B. As with the standard repeated measures design, every subject will be tested for both conditions. Subjects will be divided into two groups and one group will be tested with condition A, followed by condition B, and the other will be tested with condition B followed by condition A. If there are three conditions, the subjects will be divided into 6 groups, treated as orders ABC, ACB, BAC, BCA, CAB and CBA. Overall, for n conditions, the number of combinations (levels) are n!, and the number of participants should be a multiple of n! (so that each level may be assigned an equal number of participants). It can be easily understood that in complex experiments with more conditions the levels quickly multiply and make the experiment impractical or even unfeasible.

For this reason, **incomplete counterbalanced** designs may be opted for, such as the very common **balanced Latin Square** design. The complete presentation of these designs is beyond the scope of this report (see e.g. Montgomery, 2000). As a general recommendation, if carryover effects are moderate or unknown, the typical design strategy depends on the number of levels of the within-subjects factor of interest:
- If the number of levels is small (e.g. 2, 3, or 4 perhaps), all levels can be presented (full counterbalancing).
- If there are more levels (e.g. > 4), a Latin squares approach or a randomised ordering may be adopted.

On the basis of the above, the general criteria for the **quality of experimental studies** can be outlined as follows:
- Adequate screening of participants: The presence of concrete inclusion or exclusion criteria, with respect to the objectives of the study will be an advantage for validity of the experiment;
- Generalization issues: Tasks, population samples and environments are not similar to whom or what one wishes to generalize. It is crucial to qualify results according to generalizability limitations. Researchers should include similar tasks, samples and environments to desired generalizations;
- Sample representativity and power: In several cases, sample representativity limitations are involved in experiment design, especially in experiments with monetary benefits for participants. Selection bias mentioned above may be also involved. Moreover, sample size should be large enough for the number of factors investigated, allowing for the desirable factorial design to be analysed.
- Randomisation of participants or treatments (counterbalancing): Although the existing techniques may not fully eliminate biases due to the selection and allocation of participants, and the order of trials, researchers should make an effort in accounting for or isolating as much of these

biases as possible. Full randomisation may not always be necessary, and full counterbalancing may not always be possible. However, a study should clearly demonstrate the methods used to handle selection bias and carryover effects.

- Drop out: It is likely that participants may drop out before the experiment tasks are completed, due to sickness, fatigue or discomfort from the experiment activities, loss of interest or motivation for remaining in the study (especially in longitudinal experiments or studies requiring more than one visit in the laboratory or site) etc. The drop-out rate of each study should be mentioned, in addition to the ways it was handled.

### 3.3.2    Quasi-experimental designs

Quasi-experimental designs imitate experimental designs by having a control group in which a measure is not introduced or a risk factor is not present. As mentioned before, the difference is that the control group is chosen on the basis of external circumstances (e.g. whether a local politician had decided to build a roundabout or not); there is no random assignment of units or individuals to it.

**Difference between groups**

Outcomes from an experimental group are compared with outcomes from a control group. There are two ways to form a control group.

1.) Concurrent control: exposure/treatment and control group participants are matched at the group level based on demographic and other characteristics, and receive different treatment/exposure conditions (with/without) at the same time.

2.) Historical control: investigators compare outcomes among a group of participants who are receiving the treatment (experimental group) with outcomes among participants who received standard treatment in a previous period (control group).

**Beforeafter studies without and with control group**

The most common quasi-experiments in road safety research are before-and-after studies which establish the number or the percentage of crash reduction following treatment (Hauer, 1997). There are confounding factors which influence the number of road crashes and casualties and, therefore, should be accounted for in the selection of sites for treatment and in the estimation of a real safety effect of the treatment.

The threats to validity are (Hauer, 1997):

- **Selection bias**: Road crashes have a random behaviour. Hence, in some periods, the values measured on given points of the network can be greater (or less) than the average values expected for those points. If a group with extreme measurements are selected (e.g. the crossings with the highest crash number), a selection bias occurs because it is more likely that the measurement for the chosen cases had been abnormally high than abnormally low. In the measurements made after the treatments, an effect of decrease of crashes is registered (also known as "*regression to the mean*"), independent of the treatments.
- **Uncontrolled environment**: Road crashes occur in a setting, which, unlike a laboratory, is not 'controlled'. Therefore, for some types of road crashes, some medium-*long term trends* can be observed due to various safety features of vehicles or a change in driver habits. If a decreasing road crashes trend took place in the previous years, the reduction of road crashes after a treatment would probably have occurred even without the treatment.
- **Other external factors**: These can also affect the number of road crashes where a treatment took place; for instance, a *reduction or an increase in traffic flows* may bring about a variation in the number of road crashes, independent of the treatment. Other co-incident factors mostly

concern the **parallel implementation of other measures** like lowering the speed limit or increasing control activities.

In order to properly quantify the effects of a treatment, a simple before-and-after comparison is not sufficient, as it is necessary to compare the situation with the treatment (**"after"**) to the situation that would have existed if the treatment were not applied. The latter presents a corrected value of a previously observed (**"before"**) situation. Determining what would have occurred in a site without the treatment is a critical part of the entire process and is performed in two steps: first, the determination of the correct "before" value (of the effect), which accounts for the selection bias, and second, the determination of the correct "after" value without the treatment, accounting for the uncontrolled environment.

Traditional methods involve a simple before–after comparison of crash counts or rates, with or without a comparison or control group. This typically involves a process in which sites are selected for possible treatment on the basis of their safety record and then randomly allocated to either a treatment or a control group — a classical experimental design. This would create similar crash frequency distributions in the two groups, allowing for regression-to-the-mean effects to be controlled for. In practice, this method of project selection is problematic due to a number of reasons (Persaud & Lyon, 2007).

To avoid these issues in using a control group, a quasi-experimental design is commonly used in which an untreated "comparison" group of sites similar to the treated ones is selected separately from the treatment site selection process. A comparison group can account for general trends but will not account for regression-to-the-mean unless sites are precisely matched on the basis of crash occurrence in addition to all the factors that affect crash occurrence (Persaud & Lyon, 2007). This is very difficult to achieve in practice.

The Empirical Bayes method constitutes one possible and effective instrument for this step. Some mixed generalised linear models could also be used. The hypothesis is that the crash risk, defined as the mean of a Poisson distribution, is not constant over the units but varies randomly among them, usually as a gamma distribution. Finally, the number of crashes of each unit follows a negative binomial distribution. A correction of "before" safety effects is performed with the help of reference group statistics, for each site in the treatment group. The empirical Bayes (EB) method for road safety estimation is a non-parametric method and utilises two sources of data regarding safety to develop estimates that are site-specific and thus account for the site-specific characteristics that influence the number of crashes. The two sources of data are:

1. A model-based estimate of the number of crashes expected to occur on a site with known values for all independent variables included in the accident prediction model.
2. The number of crashes recorded on a site during the same period as used to develop the accident prediction model.

The logic of the method is shown in Figure 3.5. A number of factors are entered into a multivariable accident prediction model and their relationship to crashes estimated. Local risk factors cannot be included in such a model, but will influence the recorded number of crashes at a specific site.
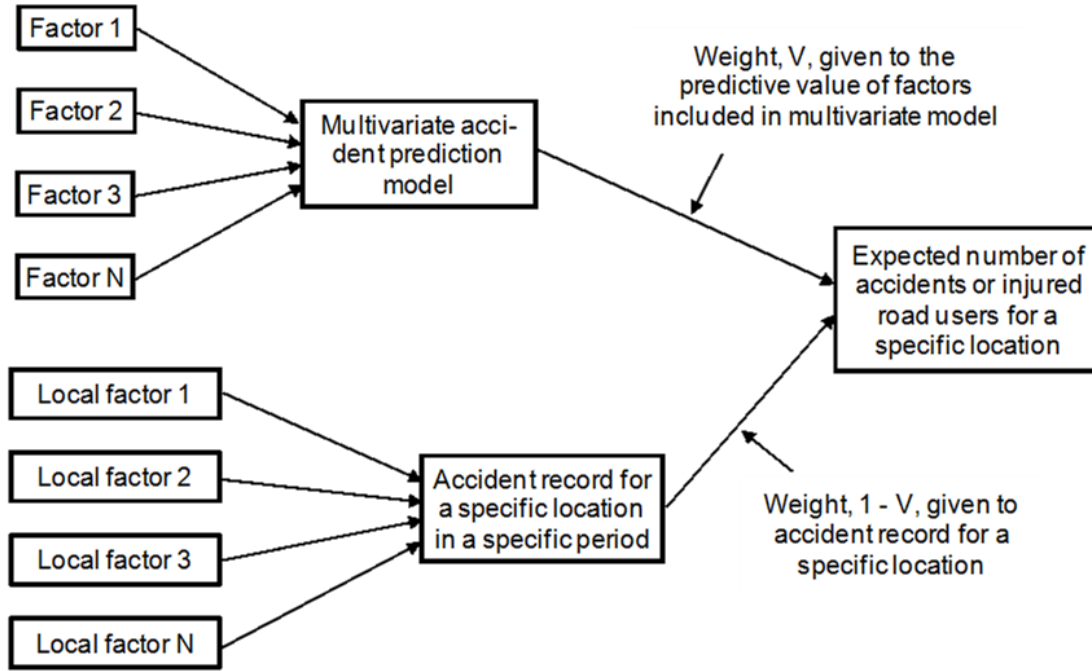
**Figure 3.5** The logic of the Empirical Bayes method for road safety estimation

In the Empirical Bayes (EB) evaluation of the effect of a treatment, the change in safety at a treated location is given by:

$$E_A - E_B \qquad\qquad (3.1)$$

where $E_B$ is the expected number of crashes that would have occurred in the "after" period without applying the treatment and $E_A$ is the number of reported crashes that occurred in the after period.

Hauer (1997) and Persaud and Lyon (2007) suggest that because of changes in safety that may result from changes in traffic volume, from regression-to-the-mean and other exogenous factors, the number of crashes before a treatment by itself may not be a good estimate of $E_B$. More specifically, the authors suggest that the $E_B$ should be estimated from an EB procedure in which a safety performance function (SPF) is used to first estimate the number of crashes that would have been expected in each year of the "before" period at locations with similar characteristics geometric and traffic parameters to a treatment site being examined.

The sum of the annual SPF estimates (P) is then combined with the number of crashes (n) in the before period at the treatment site to obtain an estimate of the expected number of crashes (m) before the treatment. In this case, m equals:

$$m = w_1 * x + w_2 * P \qquad\qquad (3.2)$$

where,

$$w_1 = \frac{P}{P + \frac{1}{k}} \qquad\qquad (3.3)$$

$$w_2 = \frac{1}{k * (P + \frac{1}{k})} \qquad\qquad (3.4)$$

and k is the dispersion parameter of the negative binomial distribution that is assumed for the number of crashes in estimating the safety performance function.

According to Persaud and Lyon (2007), a correcting factor is then applied to Eq. (2) account for the length of the after period as well as differences in traffic volumes, weather, geometry and other related factors and measures between the before and after periods. This correcting factor is the sum of the annual SPF predictions for the after period divided by P, the sum of these predictions for the before period. By doing so, the obtained result, is an estimate of $E_B$.

Finally, the estimate of $E_B$ is then summed over all road sections/segments in a treatment group of interest ($E_{Bsum}$ is then obtained) and compared with the number of crashes during the after period in that group ($E_{Asum}$ is then obtained). It is noted that the variance of $E_B$ is also summed over all sections in the group of interest.

Thus, the overall crash reduction (θ) or in other words the "index of effectiveness" is the following:

$$\theta = \frac{\frac{E_{Asum}}{E_{Bsum}}}{1+[\frac{Var(E_{Bsum})}{E_{Bsum}^2}]}$$

The percent change in crashes is in fact 100*(1−θ). For instance, if θ = 0.8 a 20% reduction in crashes is observed. The standard deviation of θ is estimated by the next equation:

$$Stdev(\theta) = \sqrt{\frac{\theta^2 * \{[\frac{Var(E_{Asum})}{E_{Asum}^2}]+[\frac{Var(E_{Bsum})}{E_{Bsum}^2}]\}}{\{1+[\frac{Var(E_{Bsum})}{E_{Bsum}^2}]\}^2}}$$

Two basic approaches are possible for estimating the global effect on the sample of sites (Yannis et al, 2008): (1) using a comparison group or (2) using a multivariable model.
(1) Using a **comparison group**, assumes that changes in the safety effect in the comparison group forecast accurately the changes that would have occurred at the treatment sites in the absence of treatment. The evaluation of the treatment effect is performed by means of the Odds-ratio, where for the "before" period the "corrected" effects numbers (from the first evaluation step, i.e. the EB method) are applied.

In this case, the safety effect of the treatment at site i is estimated as:

$$\text{Estimated effect } (\theta_i) = [r_A/E_B]/[C_A/C_B]$$

Where:
$r_A$ - the number of road crashes observed at the treatment area in the "after" period
$E_B$ - the "corrected" (EB) number of road crashes at the treatment area in the "before" period
$C_A$ - the number of road crashes observed at the control group area in the "after" period
$C_B$ - the "corrected" number of road crashes at the control group area in the "before" period

The statistical weight of the estimate is then:

$$w_i = \frac{1}{\frac{1}{A^i}+\frac{1}{B^i}+\frac{1}{C^i}+\frac{1}{D^i}}$$

Where A, B, C, D are the four numbers of the odds-ratio calculation.
The global effect is estimated by the weighted mean:

$$\text{Weighted mean effect (WME)} = \exp\left(\frac{\sum_i w_i \ln(\theta_i)}{\sum_i w_i}\right)$$

with 95% confidence interval for the weighed effect estimated as follows:

$$\text{WME} \exp\left(-\frac{z_{\frac{\alpha}{2}}}{\sqrt{\sum_i w_i}}\right), \text{WME} \exp\left(\frac{z_{1-\frac{\alpha}{2}}}{\sqrt{\sum_i w_i}}\right)$$

The applicable value of the safety effect, i.e. the best estimate of crash reduction associated with the treatment (in percent), is calculated as (1-WME)*100.

(2) When there is **no comparison group**, the **Empirical Bayes estimate** can be based on a multivariable model. Having estimated the sum of the expected number of crashes in the "before" period $E_B$, the second step is to calculate the expected number of crashes for the after period $E_A$. A factor (multiplier) is developed to account for the differences in the period length and traffic volume between the before period and the after period. This multiplier is the ratio between the predicted crashes $\lambda_A$ for the after period and the predicted crashes for the before period $\lambda_B$. The expected number of crashes for the after period can be calculated by applying this multiplier to the expected number of crashes for the before period.

The third step is to calculate the overall odds ratio of collision reduction (θ) and its standard error (Hauer, 1997; Hauer et al. 2002) as follows (for each site):

$$\theta = \frac{\frac{r_A}{E_A}}{1 + \frac{Var E_A}{(E_A)^2}}$$

with Var $E_A = \left(\frac{r_A}{r_B}\right)^2 E_B (1-\alpha)$

and standard error $SE = \dfrac{\dfrac{r_A^2}{E_A}\left[\dfrac{1}{r_A} + \dfrac{Var E_A}{(E_A)^2}\right]}{1 + \dfrac{Var E_A}{(E_A)^2}}$

It is important to note that even the most sophisticated before-and-after study designs are susceptible to unintended side effects, most notably crash migration. Crash migration denotes a tendency for crashes to "migrate" to neighbouring sites close to treated sites. As an example, drivers might find newly implemented roundabouts or speed bumps bothersome and chose an alternative route. As a consequence, the number of crashes may decrease at the treated side but increase at non-treated sites nearby.

To summarise, the main potentially sources of bias in before-after studies are:

- Regression-to-the-mean
- Long term trends in crashes
- Exogenous changes in traffic volume

A control group is the minimum requirement to control at least for long term trends. The Empirical Bayes method helps to correct for all three biases and can be applied to correct the "before" measurements either on the basis of a control group or on the basis of multivariable modelling of a larger set of reference sites.

**Interrupted time-series analysis**

Another way to gauge the true importance of an observed change in outcomes before and after an exposure is to consider the change in a much broader time frame. Time-series analysis techniques then allow to verify whether the observed change is perhaps part of natural variations over time or any other secular trend. In such cases we speak of an "interrupted time-series design" or time series with interventions. The analysis of the monthly road traffic fatalities in Great Britain have become the classic example of the evaluation of a road safety measure using time series analysis (Harvey & Durbin, 1986). In 1983 the seat-belt law took effect and led to a significant reduction of the number of killed and seriously injured car occupants. In the time-series analysis of killed and seriously injured car drivers, this can be detected by a significant level-shift in February 1982 (see Figure 3.6 , panel d), which means that the number of fatalities dropped at that moment without returning to the original level afterwards. The fact that the level shift was significant implies that it was a larger shift than observed at other moments in time.
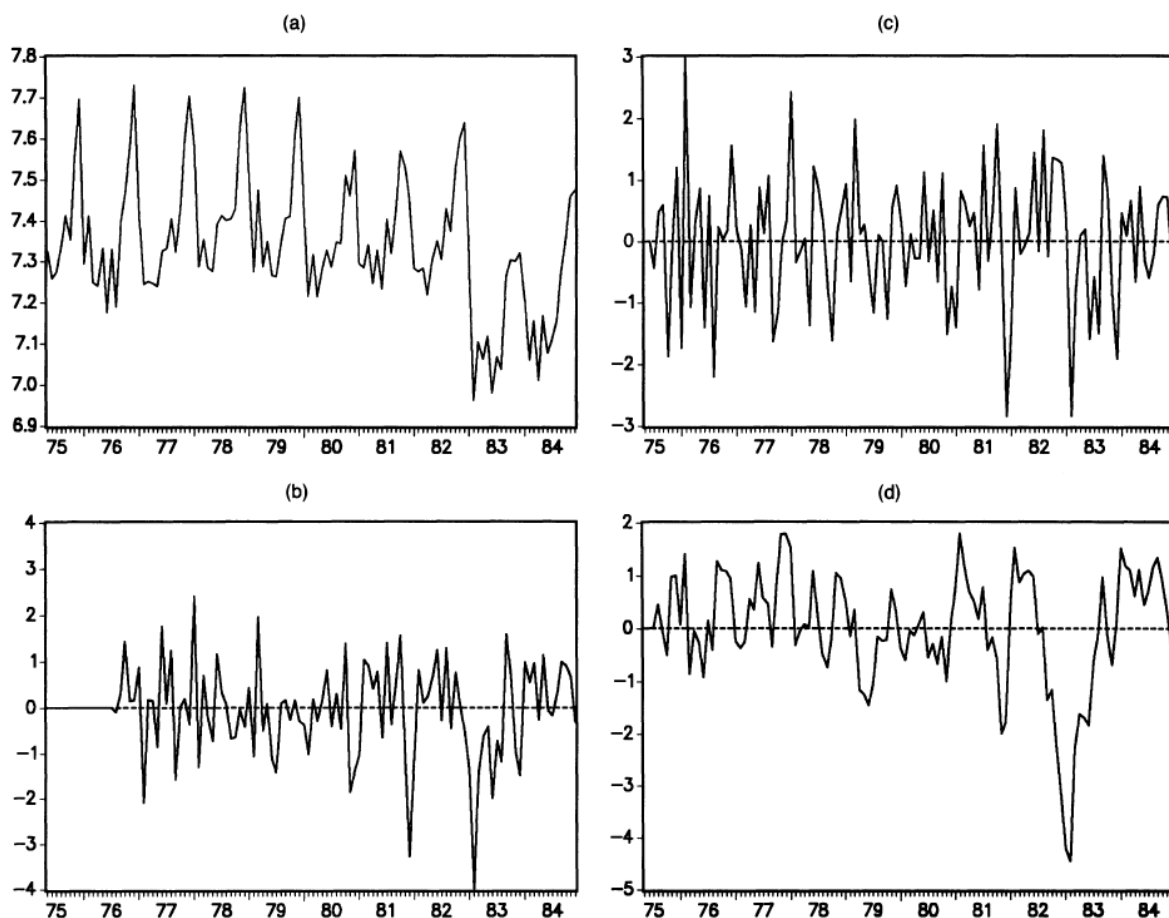


**Figure 3.6** Monthly (log) number of killed and seriously injured car drivers, Great Britain 1975 - 1984 a) observations b) innovations c) irregular d) level residual

In practice it is very difficult to detect the change due to one measure in the development of the national victim total. It is more promising to analyse the development among the specific subgroups of crashes or victims to which the measure applies.

## 3.4 ANALYTICAL OBSERVATIONAL STUDY DESIGNS

### 3.4.1 Cohort studies

Cohort studies follow a sample over a long period and allow to observe how the incidence of road crashes and their consequences unfolds over time. In road safety research cohort studies are not so common, except if we consider longitudinal studies on a sample of units such as sites or drivers along time, which can be studied with panel technics: micro panel if the number of points in time is low, macro panel based on time-series if it is high.

There are comparatively few examples of cohort studies in road safety. In epidemiology, cohort studies, in particular prospective cohort studies are regarded as a strong design, less exposed to bias than, for example, case control studies. The potential sources of bias in cohort studies are a mix of those in experiments and those in case-control studies. More specifically, based on Jarde et al. (2013), the main potential sources of bias in cohort studies are:

- The use of convenience samples or self-selected samples
- Differential attrition
- Poor data on exposure to a risk factor or countermeasure
- Simultaneous exposure to several highly correlated risk factors
- Not adequate control for confounding factors
- Mixing levels of crash- or injury severity

If a cohort is not sampled from a known sampling frame, such as the population cohort of all individuals born in a certain year, it may not be clear to what population the results of a cohort study can be generalised. Generalisation of the results of cohort studies based on convenience samples or self-selected samples must be based on replications. That means that if cohort studies based on convenience samples are repeated and made in different locations, but still get the same main findings, one may conclude that the findings reflect general patterns.

In a cohort study, study participants are usually followed up with during a fairly long-time period. One of the earliest cohort studies, the well-known Framingham heart study (named after the town where it took place), started in 1948 and went on for several decades. In birth cohort studies, individuals may be followed up their entire lives. Another famous cohort study was the study of smoking and lung cancer among British physicians that Richard Doll started in the early nineteen-fifties and continued for several decades.

In any study that collects data for such a long period, differential attrition can be a problem. Differential attrition denotes systematic dropping out from a system, for example that those who are exposed to a particular risk factor are more likely to withdraw from a study than those who are not exposed to that risk factor. An estimate of the risk associated with the factor may then, of course, become biased.

Some cohort studies rely on self-reported data. The accuracy of such data is always a concern. Instructions for data collection must be very precise to ensure that all study participants have the same understanding of what is meant by a trip, when it is dark, when it is slippery, and so on.

Simultaneous exposure to several risk factors or countermeasures can be a problem and makes it difficult to estimate the unique contributions of each factor. Whenever one wants to estimate the

effect of a specific factor, it is important to have a clear idea about what the potentially confounding risk factors or countermeasures are and try to control for these in analysis.

Finally, some risk factors may have different effects on different levels of crash- or injury severity. In such cases, it is important to avoid mixing different levels of crash- or injury severity.

A matched design may be used in cohort studies to help control for confounding by extraneous factors. For cohort data, matched-pairs are displayed as follows:

| | Non exposed pair member | | |
|---|---|---|---|
| Exposed pair member | Case | Non case | Total |
| Case | t | U | n1 |
| Non case | v | W | n2 |
| Total | m1 | M2 | N |

**Time series analysis**

Time-series models have a long history in road safety studies. The first applications included studies evaluating the effects of measures to control drinking and driving (Ross, 1982). Techniques for time-series analysis have since developed considerably, in particular with the development of state-space time-series models (also known as structural time-series models) (Commandeur and Koopman, 2007). State-space models allow for considerable flexibility in parameter estimation. In addition to time, other independent variables can be included in the models. Time-series models can be used both to assess risk factors, in particular those that vary over time (weather, daylight), and to evaluate the effects of road safety measures.

### 3.4.2 Case-control design

Case control studies compare cases with a particular outcome (usually crash or injury) to controls without that outcome with respect to the distribution of a potential risk factor or countermeasure in both groups. An over-representation among the cases is taken as evidence for an increased risk and vice versa. In classic case-control studies crash data are compared to large scale studies on exposure to a risk factor or countermeasure (e.g. road side surveys, travel survey's or odometer readings). *Induced exposure* means that control-cases that can be considered representative for the general population are sampled from the crash population. The validity of induced exposure studies depends on the definition of the control group.

Case-control studies have been widely used to estimate the effects of risk factors on crash occurrence or injury severity. They are less often used to evaluate the effects of road safety measures. Case-control studies have a long tradition in epidemiology. They compare a sample of cases, usually defined as a sample of road users who were involved in crashes, to a control group, which could be road users not involved in crashes. Case-control studies have been used to estimate, for example, the crash risk associated with drinking and driving, by comparing the prevalence of drunk drivers in a road-side survey to the prevalence among a crash sample (Hels et al., 2012).

Not all risk factors are easy to observe. Poor data on the exposure to a risk factor is therefore one of the main obstacles to conducting case-control studies in road safety. Many studies of the risk associated with fatigue, distraction, and various medicinal or illicit drugs have relied on self-reports

among a crash population, which is then compared to a non-crash population. Information obtained this way is very likely to be imprecise and biased, and any estimate of risk based on such data may have a large errors-in-variables bias. Unfortunately, without access to more accurate sources of data, the size of this bias cannot be determined.

For case-control studies in general, the most important threats to the validity of a case control study are:
• Poor data on exposure to a risk factor or countermeasure
• Cases and controls may have different prior probabilities of becoming exposed to the risk factor of primary interest
• Differences between cases and control with respect to prognostic factors(factors influencing the probability of survival, given an injury)
• There may be simultaneous exposure to several highly correlated risk factors or countermeasures
• Studies may not adequately control for confounding factors, in particular differences in driving distance
• Mixing levels of crash- or injury severity

As mentioned above, it is often difficult to obtain reliable data about the **exposure** to a risk factor or countermeasure. In particular, the data for cases and controls are often derived from different sources (e.g. crash data vs. road side survey) which can introduce a bias in the estimation of the relative risk.

In a case control design the risk associated with a particular factor is derived from the fact that crash cases are more exposed to it than non-crash cases. As a consequence, any factor that would be more frequent among the crash group for some *other* reason, would lead to an artificially increased risk estimate.

The most common reason for this to happen is the association with another risk factor. As an example, very young drivers drive more at night than others (they often do not own a car and borrow that of their parents when it is not needed for commuting. They also go out at night more than older people). Young people also have a higher crash risk (as they are inexperienced and reckless). So young people will be over-represented in crash populations and due to this also "night rides" will be. Without correcting for driver experience one might therefore overestimate the risk associated with driving at night.

More generally speaking, **high correlations** in the exposure to different risk factors is a problem. If not controlled, it is likely to result in inflated estimates of risk. Since risk factors tend to be positively correlated, they can greatly confound a study. Including other risk factors in a multivariable model can help (e.g. Elvik, 2012, gives some examples).

While the difference between singlevariable (crude) estimates and those that come from a multivariable model (adjusted estimates) varies, adjusted estimates are almost always smaller than crude estimates. Hence, *the poorer the control for potentially confounding factors in a study, the more likely the study is to exaggerate the risk associated with the risk factors studied.*

### 3.4.3    Induced exposure studies

In the context of crash investigations, we often have a good estimate of the prevalence of risk factors in the crash population. However, the estimation of the prevalence of a particular factor in the general population would require costly collection of additional data. In the sections below we will therefore discuss **induced exposure** techniques which aim at finding a group of control cases in the

crash data that can be considered representative for the total population. The prevalence of a risk factor or countermeasure in that group is then used as a proxy for the prevalence of the factor in question in the general population.

The most common practices in this context are culpability studies in the context of investigating risk factors and the consideration of system neutral crashes in the context of the evaluation of measures, and injury mitigation analyses.

### Culpability studies (quasi induced exposure)

This design is based on the idea that some participants of crashes have become innocently involved in the crash. Just by being at the wrong time in the wrong place they happened to be involved in a crash that was caused by the error of another road user. This group of innocently involved drivers is not different from other drivers who were just luckier. In a culpability study the participants are therefore divided into drivers at fault (culpable drivers) and those non-at-fault. The idea is that only the drivers at fault have a structurally increased crash risk, while the second group's risk is just as high as those of other non-involved drivers.

This method was first suggested by Thorpe (1967) and improved by Haight (1970) who named it a quasi-induced exposure. The core assumption is that non-at-fault drivers are a random sample of the driving population at or around the crash scene and can therefore function as a control group for the cases – the culpable drivers.

Next to the practical advantage of being able to derive an exposure indicator from the same data-set as the crashes, the quasi-induced exposure design has the advantage that cases and controls are naturally matched with respect to place and time of the crashes, which cannot be guaranteed when using large scale measurements of travel behaviour (e.g. vehicle km) as a measure of exposure.

There are however, also a number of issues that should be critically considered:
- The concepts "culpable" and "non-at-fault" drivers relate to an outdated view of crashes being caused by the error of one road-user. In modern systems views (see Section 2.2) crash causation is rather considered a set of circumstances that all contributed to the crash. In most cases contributing factors can be determined for all crash participants. Drivers who could have done absolutely nothing to prevent the crash are rather exceptional.
- Culpability is most often defined on the basis of police officers judgements and/or on the registration of driving offenses (e.g. drunk driving, speeding, etc.) (for an overview see Jiang et al., 2014). Jiang et al. (2010) compared the police's culpability judgments with those of experts. They showed that intoxicated, young, and male drivers have a higher probability of being considered guilty by the police even if experts' judgement indicated them as the less culpable driver. Conversely the severely injured or killed victim as well as the remaining driver at the scene in a hit and run crash had a lower probability of being considered guilty by the police. As a consequence, the relative risk of all variables related to these characteristics will be distorted by this bias.
- Data cleaning and the selection of crashes where the roles are clearly divided between an at-fault driver a non-at-fault one leads routinely to the elimination of more than half of the originally analysed crashes. It is unclear how this affects the risk estimates (Jiang, Lyles, & Guo, 2014). Single vehicle crashes and crashes with vulnerable road users are usually excluded from the studies. Consequently, the risk estimates only apply to a limited range of crashes.
- Drivers and vehicles with a higher crash avoidance capacity (e.g. newer vehicles, more experienced drivers) and have a lower probability of ending up in a crash – even as an "innocent" victim. Their share among the control group will consequently be underestimated. Similarly, physi-

cally vulnerable drivers have a higher chance of ending up in a crash that is registered by the police (as crashes without injury are often not registered) (Mendez & Izquierdo, 2010).

As an example for a culpability study, Gladegbeku and Amoros (2007) compared the share of drivers under the influence of alcohol and other narcotic substances among drivers considered responsible for the crash and those not considered responsible.

The "responsibility" was determined on the basis of an algorithm proposed by Robertson and Drummer (1992). This method consists in computing a responsibility score, based on information from 8 groups of characteristics: information on driving offences, road conditions, traffic conditions, vehicle conditions, crash type, complexity of the driving task, complexity of traffic regulation, tiredness of the driver, and witnesses comments. The information about possible alcohol offenses was disregarded in determining the responsible driver. This was important because otherwise the categorisation of the driver would artificially increase the share of drivers under the influence in the "responsible" group. For a subgroup the categorisation was evaluated by comparing to a responsibility rating of a group of crash experts. The resulting relation (kappa=.67) was considered satisfactory. The odds ratios - adjusted for age and gender - were calculated by means of a multiple regression analysis. The resulting Odds Ratio for alcohol was 8.39, meaning that the odds for at fault drivers to be under the influence of alcohol was more than 8 times higher than for non-at fault drivers.

Given the problems indicated above, results from culpability studies should be interpreted with care. To check the reliability and validity of the results consider the following questions.
- Is the data cleaning procedure documented and conducted in a responsible way? Does the author consider the possible effect of systematically discarding particular types of crashes?
- Is the assignment of responsibility for the crash documented? Is it done systematically? Do the authors make sure that the criteria for the assignment of responsibility are not related to the risk that should be measured? *As an example, Gadegbeku and Amoros (2006) determined the responsibility risk of alcohol intoxicated drivers relative to those who were not intoxicated. They correctly dropped the item "drink driving" from a list of criteria that were used to determine the drivers' responsibility.*
- Is the validity of the assumption that non-at-fault driver are representative of the general population evaluated (e.g. by means of comparing them to population)? *As an example, Stamatiadis and Deacon (1997) compared the percentages of cars, straight-trucks, and combination-trucks among the non-culpable drivers to those from road-side counts for different types of roads. They found a good agreement, which suggests that – at least with respect to this aspect – the non-culpbable drivers are indeed comparable to the general driving population.*

### Neutral crash conditions

For the evaluation of safety features (SF) it is more useful to differentiate between those crashes to which a safety feature was relevant and those for which no protective function can be expected – the neutral crashes. A detailed description of this method can be found in Zangmeister, Kreiss, & Schüler (2007). The main idea is to compare the percentage of equipped vehicles (equipped with a certain safety configuration of interest) within the observed crashes of interest (i.e. the percentage of ESC-equipped vehicles within loss of control or skidding crashes) with the percentage of equipped vehicles on the roads. Since this last frequency typically is not observed, the idea is to estimate this percentage by the percentage of equipped vehicles within a category of neutral crashes.

The validity of this approach depends on how well the observed population within the neutral crash/vehicle category resembles the population on the roads. Consequently, the selection of the category of neutral crashes is most essential in the quantification of the effectiveness of safety functions or safety configurations based on real-world crashes. The effectiveness, calculated by subtract-

ing the odds-ratio from 1 (E=1-OR), quantifies the expected reduction *among the crashes that are considered relevant to the system*. See Zangmeister et al. (2007) for the necessary corrections to estimate the reduction achievable among *all* crashes.

An example for this approach is Fildes et al. (2015) who investigated rear-end collisions and compared the rate of vehicles with autonomous emergency braking (AEB) among the striking vehicle (the relevant crash situations that could be helped by AEB) and the struck vehicles (the neutral crash situations that could *not* be helped by AEB). The calculated effectiveness was therefore:

$$eff_A(AEB) = 1 - OR = 1 - \frac{\frac{nStriking\ with\ AEB}{nStriking\ without\ AEB}}{\frac{nStruck\ with\ AEB}{nStruck\ without\ AEB}}$$

The effectiveness was calculated in 6 countries. 5 of the countries showed a positive but not significant effectiveness. The measures of effectiveness per country were summarized as weighted average, which resulted in a significant safety effect, indicating that the odds for AEB equipped vehicles to be the striking car in a rear-end crash are reduced by 38%.

The effectiveness of a safety function in a specific crash highly depends on which of the involved vehicles is considered for the evaluation. There may be several vehicles involved in a single crash and it is crucial that each of them is considered separately (e.g. the hitting and the hit vehicle in a rear-end crash). It is crucial that the considered safety features by their technical description do not at all influence crash situations which may lead to a crash belonging to the neutral category. It is often not easy to find such a crash type. In the example of Fildes et al. (2015), this is problematic, because a car that braked hard to avoid striking the vehicle ahead has actually a higher chance of being struck from behind themselves by a following car. Therefore the "neutral" condition (struck car) is not really neutral but more likely for cars with AEB, which might inflate the estimated effect.

Relevant and neutral crash conditions could differ on variables that have nothing to do with the functioning of the safety system, but do affect its distribution. For example: young drivers tend to be over-represented in single vehicle crashes. As many safety features make a vehicle more expensive and therefore more often bought by middle-aged drivers these systems might be under-represented in single vehicle crashes. If either the neutral or the relevant crash conditions contain mainly single vehicle crashes, this relation could affect the estimation of the safety effect.

To check the reliability and validity of the results consider the following questions:
- Is the assignment of relevant and neutral crashes documented and conducted in a responsible way? Does the author consider the possible effect of systematically discarding particular types of crashes?
- Are the neutral crashes really neutral? I.e., is there no way how the occurrence of the crashes in the neutral category could be influenced by the safety system under investigation?
- Is the safety-effect (or odds-ratio) adjusted to give the expected effect on all crashes?

Induced exposure studies are case control studies. However, while in "classic" case – control studies, the controls did not have a crash, in induced exposure studies one tries to find a group of vehicles or drivers that is representative of the whole population with respect to the characteristics relevant to the measure or risk factor under investigation.

The great advantage of this technique is the possibility to base cases and controls on the same database and automatically correct for all specificities of the crash data in question. The great problem

of this technique is that the controls must be representative for a population of (mostly) non-crash individuals, which is somehow contradictory to the fact that they are recruited from a crash database.

The quality of an induced exposure study mainly depends on the definition of the neutral group or the non-at-fault group. Whether this group can indeed be considered a fair representation of the population depends on the underlying mechanisms of risk-factors and safety systems. There are no standards for this and generally, one should apply the rule "not representative until proven otherwise". This means the authors of each study have to make a convincing case that their control group although found in the crash database can indeed be considered neutral to the countermeasure or risk under investigation.

### 3.4.4 Matched case-control and case crossover studies

In a matched study, we enroll controls based upon some characteristic(s) of the case. For example, we might match the sex of the control to the sex of the case. The idea in matching (*one-to-n matching)* is to match upon a potential confounding variable in order to remove the confounding effect. In an analysis of a matched study design, only discordant pairs are used. A discordant pair occurs when the exposure to a risk factor or countermeasure of case is different than the exposure of the control. Analytic methods for matched case control studies include conditional logistic regression, conditioned upon the matching.

For case-control data, matched-pairs are displayed as follows:

|  | Control pair member |  |  |
| --- | --- | --- | --- |
| Case pair member | Exposed | Non exposed | Total |
| Exposed | t | u | $n_1$ |
| Non-exposed | v | w | $n_2$ |
| Total | $m_1$ | $M_2$ | N |

The case crossover design is useful when the risk factor/countermeasure is transient. For example, cell phone use or sleep disturbances are transitory occurrences. Each case serves as its own control, i.e. the study is self-matched. For each person, there is a 'case window', the period of time during which the person was a case, and a 'control window', a period time associated with not being a case. Risk exposure during the case window is compared to risk exposure during the control window.

### 3.4.5 Cross-sectional studies

Cross-sectional studies are widely used both to identify risk factors and to evaluate the effects of road safety measures at one point of the time, usually over one year. Unless corrected by multivariable (accident prediction) models, they are subject to many confounding factors. The most important variable to correct for is the distance driven as well as gender and age. An important threat to the validity is the endogeneity bias, the tendency for measures to be applied to units with extreme outcomes. This bias can lead to the reversion of the observed outcome.

Cross-sectional studies are widely used both to identify risk factors and to evaluate the effects of road safety measures. Usually, studies analyse a cross-sectional data set by means of a multivariable model.

### 3.4.6 Examples of confounding factors

The most common estimator of safety in cross-sectional studies is the crash rate, that is the number of crashes per million kilometres of driving. The crash rates may refer to drivers, road sections, junctions or vehicles. The effect of a risk factor or a safety measure is stated in terms of the crash rate ratio, which is the crash rate for one value of a certain variable divided by the crash rate for a different value of the variable. Variables of interest are often discrete, rather than continuous, meaning that they take on only a small number of values.

The main potentially confounding factors in cross-sectional studies are:
- Self-selection to a risk factor or a safety treatment
- Endogeneity of a risk factor or safety treatment
- Differences in traffic volume or annual driving distance
- Differences in traffic composition or environment of driving
- Differences in other relevant risk factors

Examples will be given of some of these confounding factors. Endogeneity is an issue in multivariable accident prediction models as well and will be discussed below.

Annual driving distance has been an issue in studies of the relationship between driver age and crash rate. Are older drivers more often involved in crashes than middle-aged drivers? Crude crash rates suggest that the answer is yes. The crash rate is, however, not independent of annual driving distance. Older drivers drive a shorter annual distance than middle-aged drivers. Does the difference in crash rate remain when one controls for differences in annual driving distance?

Elvik (2012) reviewed studies comparing driver crash rates between different age groups and compared their findings with and without controlling for annual driving distance. Figure 3.7 reports the results.



**Crude and adjusted accident rate ratios for older drivers compared to the safest age group of drivers in six studies**
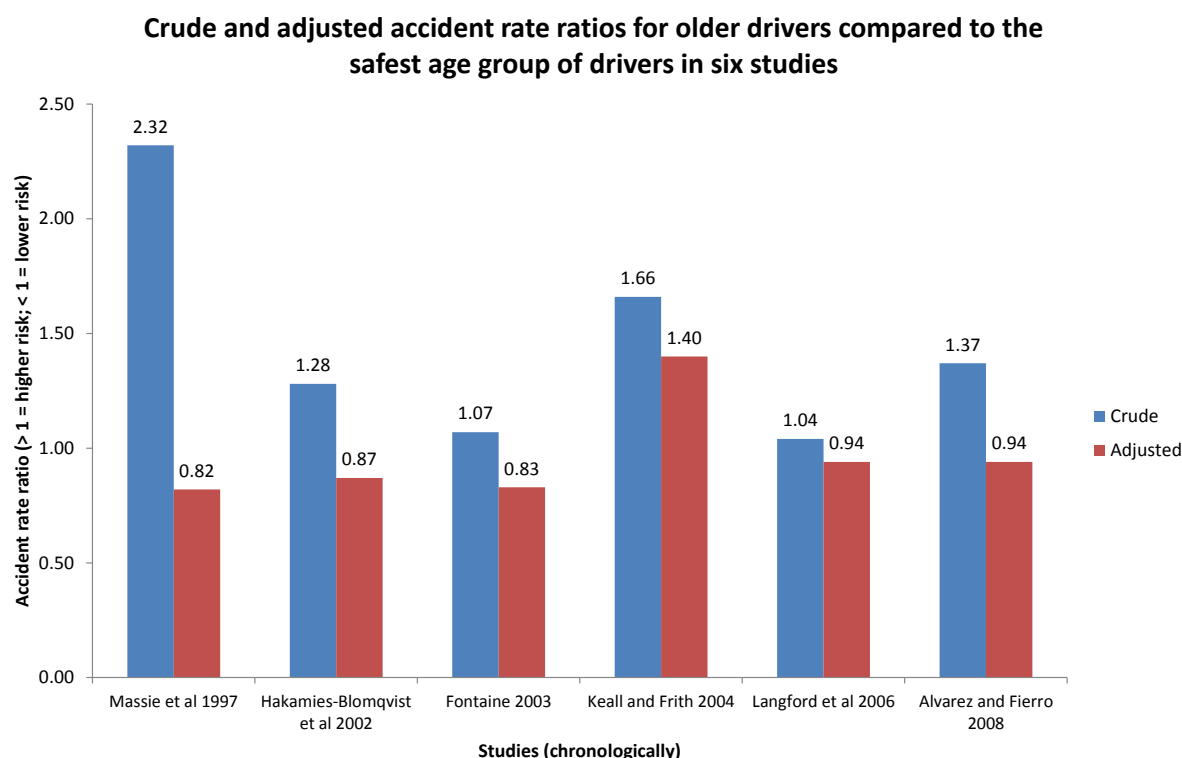
**Figure 3.7** Crash rate ratio for older drivers compared to the safest group of driver with and without control for annual driving distance

It is seen that all studies found a higher crude crash rate for older drivers than for the safest age group of drivers. When estimates of crash rate were adjusted for differences in annual driving distance, older drivers were no longer found to have a higher crash rate than the drivers in the safest age group (middle aged drivers).

In a similar vein, the importance of differences in traffic volume for road-related risk factors can be shown by a Nordic study. Figure 3.8 presents the relationship between paved road width and injury crash rate for national roads in Norway (Nordtyp-projektgruppen, 1980). In the report, the thick curve in the middle of the figure is highlighted and presented as the main result. It shows a consistent decline in crash rate as road width increases.
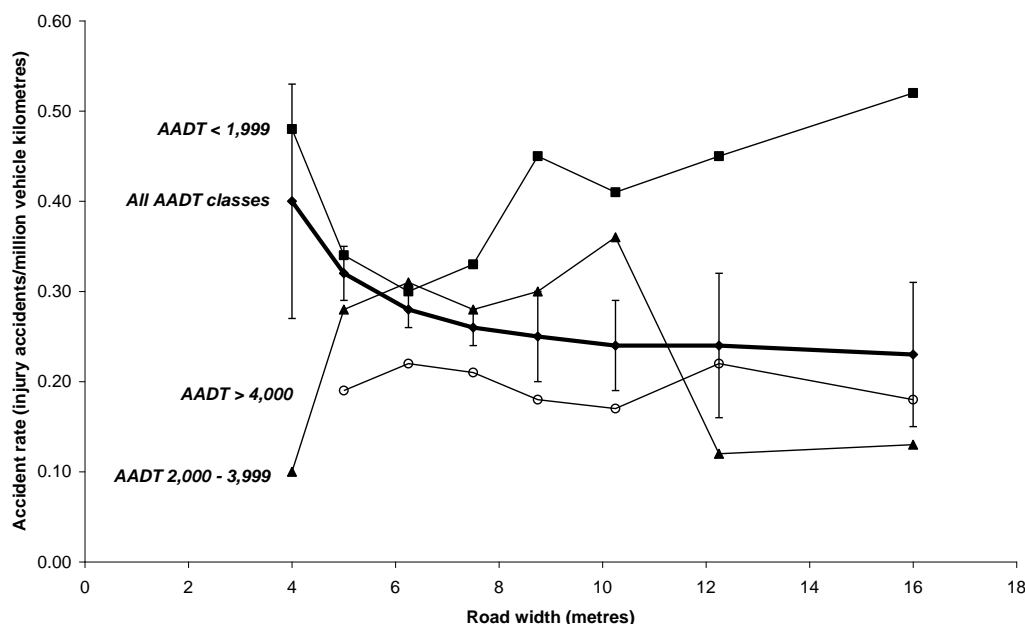


**Figure 3.8** Relationship between paved road width and injury crash rate. Based on Nordtyp-projektgruppen (1980)

The thick curve shows the simple bivariate relationship between road width and crash rate. This relationship is confounded by numerous confounding factors. The three other curves shown in Figure 3.8 refer to roads with AADT less than 2,000, roads with AADT between 2,000 and 3,999 and roads with AADT 4,000 or more. As can be seen, even the very crude control for traffic volume introduced by these three classes makes the relationship between road width and crash rate vanish completely. The three curves representing roads with different traffic volume in Figure 3.8 show no consistent relationship between road width and crash rate. These curves fluctuate erratically around a flat line. In other words, the relationship indicated by the thick curve is entirely spurious and can be fully explained in terms of the correlation between traffic volume and road width and the fact that the number of crashes does not increase in strict proportion to traffic volume.

### 3.4.7 Multivariable crash regression models

Multivariable statistical models are increasingly applied both to estimate the effects of risk factors and to evaluate the effects of road safety measures. The typical estimator of effect in such models is a regression coefficient. For continuous variables a coefficient indicates the *slope* (i.e. the change of the dependent variable for one unit of the independent variable). For discrete variables this coefficient indicates the *difference* between test and reference condition. Whether these changes can be given a causal interpretation is a very difficult and controversial topic.

The principal argument for using multivariable models is that these models make it possible to simultaneously control for a large number of potentially confounding factors. Thus, each regression coefficient shows the effect of the variable it applies to, controlling for all other variables included in the model. There are, however, huge problems in developing good accident prediction models.

**Endogenous variables** refer to independent variables in a model that are influenced by the dependent variable. This will typically be the case for safety treatments. The introduction of a safety treatment is often influenced by crash history. Sites that are treated tend to have worse safety records than sites that are not treated. Even if treatment improves safety, the treated sites may continue to have a higher rate of crashes than untreated sites. A multivariable model may then find that treatment is associated with an increased number of crashes, when in fact the opposite is true. Endogeneity bias can seriously distort the results of multivariable models. It essentially reverses the direction of causality. Very often this will also reverse the sign of the treatment effect.

**Wrong model form:** The most prototypical multivariable analysis is a linear regression. This is based on the assumption that the dependent variable is normally distributed. In road safety research we are often modelling crash occurrences. These are often counts which should be modelled in a poisson or quasi-poisson regression or yes/no data which are most often modelled in logistic regressions.

*Omitted variable bias* denotes bias that occurs because a variable not included in a model is statistically associated both with a variable which is included and the dependent variable in the model. The effect of the variables included in the model can in fact be due to their relation to the underlying (but omitted) variable. No multivariable accident prediction model will contain all variables that influence safety. This will not necessarily introduce any bias in the model. How can we know whether a relevant variable has been omitted from a model or not? There will almost never be a well-established theory that tells us which variables to include or not include in a model, but if several other studies have shown that it makes an important difference to include a variable (e.g. the traffic volume), results of a study that failed to include it should be treated with care.

*Collinearity:* When two (or more) correlated predictors are included into the model, the predictors become very unstable and should not be interpreted. In the extreme case, two variables that measure almost the same end up with high coefficients in opposite directions. A good study therefore reports a correlation matrix for all (potential) predictors and indicates how the predictors in the final model have been selected to deal with this problem. Note, that collinearity is a problem for the interpretation of the coefficients, not for the predictive power of the model.

*Misspecification of systematic variation:* Systematic variation can be introduced if the data are sampled from a particular structure. Examples for such structures are data from a spatial or a temporal structure or data that form clusters within the sample (e.g. measured at different sites). This structure will influence the variation observed in the sample (cases close in space or time will be more similar to each other than more remote ones) which distorts significance tests and other results of the analysis. As a solution, multilevel or random effects models have to be used for clustered data, and time-series and spatial analyses techniques have to be used for analysing temporally or spatially distributed data.

### 3.4.8 Injury severity regression models

Physical vulnerability is a risk factor which starts to play a role when a collision occurs. The vulnerability can be measured by a probability function of the chance, when involved in a crash, to be injured more or less severely – from no injury to death. It is the domain of biomechanics, and the mitigation of the severity of the crash can be obtained through the use of protection devices such as seat belt or helmets. The resulting injury probabilities can be used in the evaluation of injury reduc-

tion potential of active safety measures (which are not widely implemented yet). Simulations on the basis of crash reconstructions can lead to an indication of how many crashes could have been prevented by a particular measure, but also how many crashes would have taken place with a reduced impact speed. The analyses described in this section allow an estimation of the severity of crashes at a particular impact speed.

The severity of injuries is measured by a scale applied to different parts of the body. This could either be a scale, like the abbreviated injury scale AIS (Gennarelli (2008) going from AIS0 (no injury) to AIS 6 (death)) or we could restrict the scale to a binary choice. Examples are fatal/non-fatal injury, or MAIS2 or less vs. MAIS3 or more. To estimate the effect of independent variables such as age, use of seat belt, collision type, occupant position, mass of the car, $\Delta V$... on the probability of injury we can use a logistic regression model such as:

$$G(E(y)) = X\beta + \varepsilon$$

It is possible to transform this model into a mixed logistic model by introducing a second error term $\eta$ in the regression to take into account some unobserved heterogeneity (Savolainen et al., 2011). To illustrate, we can take the case of a pedestrian hit by a vehicle. There are empirical studies on the relationship between the conditional probability of death according to the impact speed from samples of pedestrian crashes classified by impact speed classes of 10 km/h. Rosen and Sander (2009) used GIDAS German In-Depth Accident Study data to estimate a logistic function for pedestrians over 15 years old (490 in the sample) with weighting to control the over-representation of severe crashes in the database.

$$log \frac{p(death)}{1 - p(death)} = -6{,}91 + 0{,}095 Speed + 0{,}04 age$$

Tefft (2011) used the same method with the Pedestrian crash data study in the US for pedestrians struck by a single car or light truck model year 1989-1999. Some sensitivity analysis is made to check the influence of the quality of the impact speed measurement. Usually the effect of age affects the constant of the logistic model and not the coefficient of the impact speed.
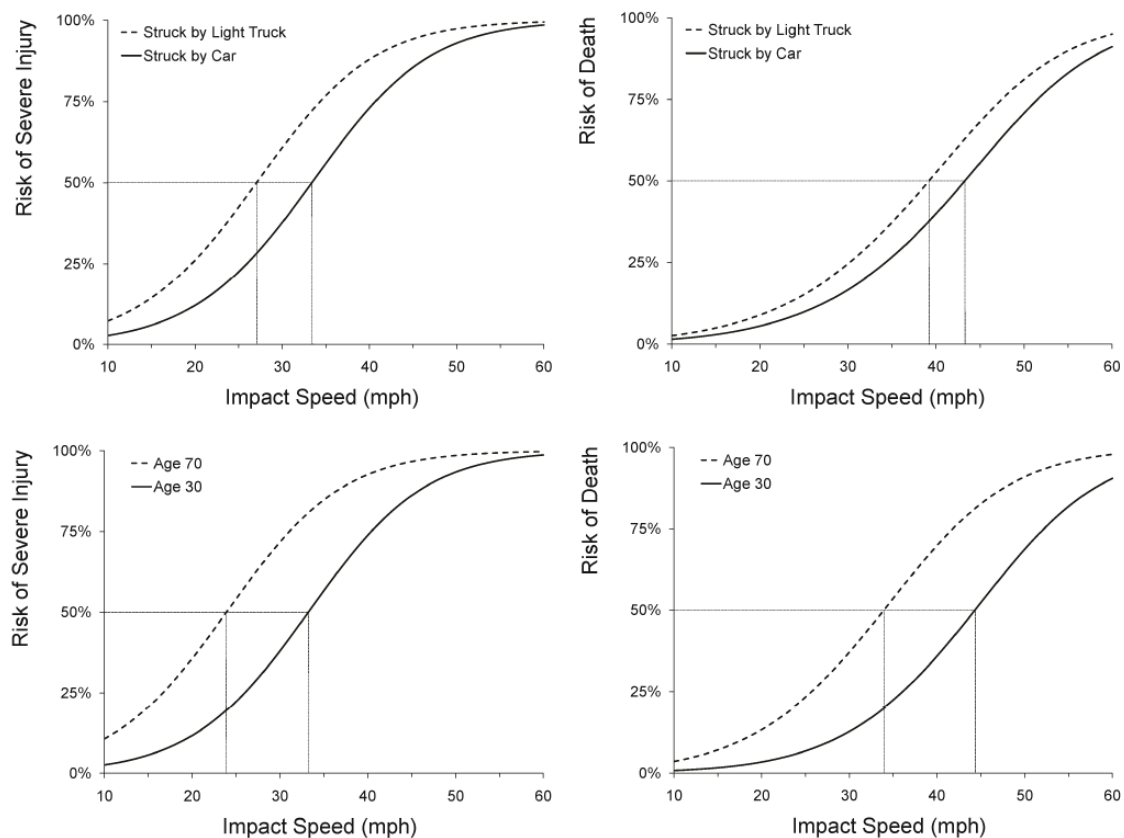
**Figure 2.** Risk of severe injury (left) and death (right) in relation to impact speed in a sample of 422 pedestrians aged 15+ years struck by a single forward-moving car or light truck model year 1989–1999, United States, 1994–1998. Risks are adjusted for pedestrian age, height, weight, body mass index, and type of striking vehicle. Top panel shows average risk for pedestrians struck by cars vs. light trucks, standardized to the age distribution of pedestrians struck in the United States in years 2007–2009. Bottom panel shows average risk for pedestrians ages 30 vs. 70, standardized to the distribution of type of striking vehicle for pedestrians struck in the United States in years 2007–2009. Serious injury is defined as AIS score of 4 or greater and includes death irrespective of AIS score.

**Figure 3.9** Risk of severe injury and death in relation to impact speed, Pedestrian crash study, Tefft, 2011

A comparison among models (Tefft, 2011) showed remarkably consistent results, even if the limitations described in each study were numerous due to selection problems in the sample and the weighting and normalization procedures are subject to discussion.

**Table 3.1** Results from Tefft, 2001: Impact speed at which estimated fatality risk reaches 10% 25%, 50%, 75%, and 90%, and odds ratio for change in odds of death given 5 mph increase in impact speed

| Study | Data | Confounders adjusted | Odds Ratio (95% CI) | Risk of death (%) | | | | |
|-------|------|----------------------|---------------------|-------------------|---|---|---|---|
| | | | | 10 | 25 | 50 | 75 | 90 |
| | | | | Impact speed (mph) | | | | |
| Davis (2001) | United Kingdom, 1966-1969 & 1973-1979 | None | 2.78 (2.09-3.69) | 33 | 38 | 43 | 49 | 54 |
| Rosen & Sander (2009) | Germany, 1999-2007, pedestrians ages 15+ years struck by front of car | None | 2.06 (1.60-2.66) | 32 | 40 | 48 | 55 | 63 |
| | | Age | 2.15 (1.67-2.76) | 33[a] | 41 | 48 | 55 | 63 |
| Richards (2010) | United Kingdom, 2000-2009, pedestrians struck by front of car | None | 2.41 (1.79-3.24)[b] | 33 | 38 | 45 | 51 | 62 |
| Current Study | United States, 1994-1998, pedestrians ages 15+ years struck by forward-moving car or light truck | None | 1.99 (1.60-2.49) | 28[c] | 36 | 44 | 52 | 60 |
| | | Age, height, weight, BMI, vehicle type | 2.18 (1.78-2.67) | 30[d] | 38 | 46 | 53 | 61 |
| | | | | 23[e] | 32 | 42 | 50 | 58 |

a.  Adjusted risks for age = 45 years (mean age in sample).
b.  Richards did not report standard error of odds ratio; standard error estimated using information in Table A2.2 of Richards (2010).
c.  Unadjusted model, pedestrians struck by cars only, pedestrians struck by light trucks excluded.
d.  Average marginal prediction for pedestrian aged 45 struck by cars, adjusted for age, height, weight, BMI, and vehicle type.
e.  Average marginal predictions for pedestrians struck by cars and light trucks, standardized the distribution of age and type of striking vehicle for pedestrians struck in the United States in years 2007–2009, adjusted for age, height, weight, BMI, and vehicle type.

The vulnerability will depend on
- The characteristics of the person (age, physical condition…)
- The crashworthiness of the vehicle (structure, mass…)
- Types of collision (frontal, lateral, rear-end…)
- In-vehicle condition (position relative to impact, safety devices such as seat belt…)
- Mechanical energy released during the collision (e.g. $\Delta V$).

In a logistic regression model, these variables can be included – either to investigate their effect or to control for the effect even if it is already known.

## 3.5    META-ANALYSIS

Meta-analysis is the statistical analysis of a set of numerical research results for the purpose of developing a weighted mean result and identifying sources of systematic variation in individual results. There are several techniques for meta-analysis. The most commonly applied technique in road safety is the inverse variance technique. Each estimate of the effect of a risk factor or a safety measure is then assigned statistical weight which is inversely proportional to its sampling variance. For more information see Chapter 8.

### 3.5.1    Principles

Meta-analyses are normally part of systematic literature reviews. A systematic literature review is a review which is performed according to rules that are intended to make it replicable. A review is

replicable if two or more researchers working independently, following the same rules, get identical results, i.e. identify the same studies dealing with a topic. The results of meta-analyses are normally reported in terms of one or more summary estimates of effect, i.e. weighted mean estimates using the inverse of sampling variance as weight. See also Chapter 8 for a more detailed description on the execution of a meta-analysis.

### 3.5.2 Potential biases in meta-analysis

There are a number of general issues in meta-analysis to consider when coding such analyses. These issues may serve as a checklist and as a crude instrument for assessing the quality of meta-analyses. The points identified below should be listed on the list of potential sources of bias on the coding template.

1. How were primary studies obtained? Does the analysis contain a description of the search for relevant studies?

If the search for studies is not described, one should code as a potential source of bias (code: maybe a problem): *No systematic literature search reported.*

2. What were the criteria for study inclusion? Are excluded studies listed together with the reason(s) for excluding them?

Ideally speaking, all studies should be included if they reported estimates of effect and the standard errors of those estimates. If additional criteria for study inclusion have been defined, they should be stated. Studies not included in the meta-analysis should be listed and the reason for excluding them stated. If excluded studies are not listed, this should be coded as a potential source of bias: *Excluded studies not listed.*

3. Which variables were coded for each study?

Any meta-analysis should, as a minimum, identify the country (or countries) and year in which each study was reported. It is often possible to code a number of additional variables. A good meta-analysis will list the variables that were coded for each study and explain how these variables were utilised in analysis. It is particularly important to code potential moderator variables, i.e. variables that may influence the size of an effect. A potential source of bias is: *Information on potential moderator variables missing in many primary studies.*

4. What type of outcome variable(s) was used in the study? In what metric was the summary estimate stated?

In studies of risk factors, the outcome (dependent) variable will typically be the odds ratio or relative risk. While the odds ratio is often used as an approximation to relative risk, the two metrics are not identical and should, ideally speaking, not be mixed up. That is, if a set of studies contain both studies estimating the odds ratio and studies estimating relative risk, separate summary estimates should be developed for the two groups of studies. The odds ratio is also a very common estimator of effect in studies evaluating the effects of road safety measures. However, a number of different estimators of effect tend to be used in road safety evaluation studies: crash rate ratio, simple odds (after/before), odds ratios, ratios of odds ratios, relative risk ratios, and regression coefficients. These estimators cannot be converted to the same metric. It may still be possible to do a meta-analysis, but then results based on the different estimators of effect should form subgroups in the meta-analysis. Thus, a potential source of bias in a meta-analysis is: *Different estimators of effect were mixed or not converted to a common metric.*

5. Was an exploratory analysis performed? Were the results of this analysis reported?

Meta-analysis makes sense if the distribution of the individual estimates of effect is "well-behaved". If individual estimates are "all over the place", i.e. display a large and apparently non-systematic variation, a meta-analysis may not make sense. The purpose of exploratory meta-analysis is to as-

sess whether the distribution of estimates is sufficiently well-behaved to proceed to a main analysis. There are many tools for exploratory meta-analysis, but a very useful tool is the preparation of funnel plots. These show at a glance how estimates are distributed. A funnel plot may also serve as the basis for more formal analyses, for example to detect outlying data points or publication bias. Therefore, a potential source of bias in a meta-analysis is: *Exploratory analysis was not made or not reported.*

6. What was the principal model applied in the main analysis? Were results based on different models compared?

The most common model of meta-analysis is the random-effects model. This model is used whenever there is systematic variation in estimates of effect. It is, however, not the only option. One may try to account for heterogeneity (systematic variation) in estimates of effect by means of meta-regression. This is a weighted regression analysis in which moderator variables are used as independent variables. Meta-regression is applied to identify sources of variation in summary estimates of effect. It should not be mixed up with using meta-regression to develop a summary estimate of effect (Stanley and Doucouliagos, 2015). Potential sources of bias at the main stage of meta-analysis include: *No testing for systematic variation in estimates of effect; Subgroups in analyses not defined in advance.*

7. Was a sensitivity analysis made? What did the analysis include? What were the principal findings?

It is good practice to include a sensitivity analysis in every meta-analysis. This should address issues such as outlying data points, publication bias, study quality and, if relevant, estimator of effect. If, for example, outlying data points are detected, there should be a description of how they were dealt with (included or excluded). There should always be a test for publication bias and the results of it reported. Furthermore, if study quality varied, there should be an analysis of whether this influenced the summary estimate, or estimates of effect. If possible to test, any bias introduced by study inclusion criteria should be assessed. Key potential sources of bias include: *No test for outlying data points; No test for publication bias; No assessment of study quality.*

## 3.6 CONCLUSIONS STUDY DESIGN

Important differentiations between research designs are those between experiments (-> researcher manipulates exposure to a risk factor or countermeasure) and observational studies (-> natural distribution of risk factor or countermeasure). Experiments can be divided into (randomised) control trial experiments and quasi-experiments where the researcher does not have (full) control over which subjects are exposed to the countermeasure and which are not. Typical experimental designs are: between-group comparisons, and before-after studies (which is a within-group comparison).

Observational studies can be analytical – when different outcomes are linked to different exposures to a risk-factor or a countermeasure – or they can be descriptive giving the prevalence of either road safety outcomes or the prevalence of a risk factor. The most important observational analytical designs in road safety are case-control studies (comparing crash cases to non-crash controls) and cross-sectional studies, where outcomes and exposures are linked to each other by means of a (mostly multivariable) statistical model.

We have given a detailed description of different study designs, describing how effects of countermeasures or risk factors are estimated and listing typical biases and quality criteria are given for each design. To summarise the Table 3.2 lists the study designs that were discussed and the potential sources of bias listed for each study.

**Table 3.2** Study design and potential sources of error

| Study Design | Most common estimators of effect | Potential sources of bias |
|---|---|---|
| **Experiments** | Absolute difference/real difference | Pre-trial non-equivalence |
| | Slopes/correlation coefficients | Differential attrition |
| | Odds ratio | Diffusion of treatment |
| | | Unintended side-effects |
| **Before-after studies** | Odds ratio | Regression to the mean |
| | Percentage crash reduction | Long term trends |
| | Crash modification factor | Changes in traffic volume |
| | Relative difference | Co-incident events |
| | | Use of several measures |
| | | Crash migration |
| **Cross-sectional studies** | Crash rate ratio | Self-selection bias |
| | Correlation coefficients | Endogeneity of risk factor or treatment |
| | | Differences in exposure |
| | | Differences in traffic environment |
| | | Differences in other risk factors |
| **Multivariable models** | Regression coefficients: | Wrong dependent variable |
| | - Odds ratio | Endogenous variables |
| | - Slopes | Wrong functional form |
| | - Relative difference | Omitted variable bias |
| | Marginal effects | Co-linearity |
| | Elasticities | Misspecification of systematic variation |
| | | Mixing levels of crash severity |
| | | Wrong model form |
| **Case-control studies** | Odds ratio | A priori differences in cases and controls |
| | | Poor data on exposure to risk factor or countermeasure |
| | | Exposure to several risk factors |
| | | Poor control for confounding factors |
| | | Differences in prognostic factors |
| **Induced exposure** | Odds ratio | Inadequate definition of neutral crashes |
| | | Selected cases not representative |
| | | Inadequate documentation of case selection |

# PART 2 – Risk analysis and assessment of measures within SafetyCube

# 4   Selecting and Prioritising Studies

The core of the Decision Support System (DSS) will consist of a repository of studies on risk factors and countermeasures in road safety. The aim is to include key studies from the European perspective. Therefore the selection and prioritising procedure describes how to first identify all relevant studies and then to prioritising them with the European Road Safety Observatory in mind as the main client.

## 4.1   IDENTIFYING RELEVANT STUDIES

Systematic searches are to be conducted in relevant databases. The databases used will likely differ between Work Packages and will be influenced by which databases partners can access. Google scholar and Scopus are recommended. It may be appropriate to specifically target institute reports and conferences, especially if not enough results are identified from the database search. This will be decided on at a WP level. When searching prioritise papers published in 1990 or later. The search terms used need to identify papers relevant to the topic of interest and to road safety. Below are some suggested search terms for road safety with the example of fatigue (search each combination of each top row word with each bottom row term). The total number of "hits" in each database should be recorded.

| Fatigue | "fatigue*" OR "Sleep*" OR "Tired*" OR "drowsy" OR "drowsiness" OR "alert*" OR "monotony" OR "time on task" |
|---------|------------------------------------------------------------------------------------------------------------|
| AND | |
| Road Safety | "road safety" OR "driv*" OR "road" OR "transport" OR "crash" OR "accident" OR "incident" OR "traffic" OR "collision" OR "traffic safety" OR "risk" OR "measure OR "Road Casualties" OR "Road Fatalities" |

**Minimum requirements:**

- Record the search terms used.
- Record the databases/institute reports/conferences searched.
- Record the number of search hits for your topic in each database. This will give an indication of the size of the topic area, although will be biased by the search terms.
- Record the date the search was conducted.

## 4.2   SCREENING

All duplicate studies are to be removed. The total number of unique studies should be recorded (this includes meta-analyses). The abstracts and titles of unique studies to be assessed based on inclusion/exclusion criteria is recorded to identify those related to the topic (risk/measure).
Any papers where it is not clear if they should be included or excluded should be retained for step 3.

**Minimum requirements:**

- Record the inclusion and exclusion criteria for screening abstracts and titles.

## 4.3 ELIGIBILITY

Obtain the full text of all remaining papers. Review any papers where inclusion/exclusion could not be determined from the abstract and title. If few papers are identified as being eligible, examine the reference list of all eligible papers to identify any additional relevant papers.

**Minimum requirements:**

- Record the end number of eligible papers.
- Record how many full-text could be obtained.
- Record whether the reference lists were examined or not.

## 4.4 PRIORITISING CODING

Ideally an attempt should be made to code all eligible papers. However, if there are many this will not be possible. Prioritise the order papers should be coded (most important first). Some suggestions for prioritising:

- Start with most recent meta-analysis papers. Check with Ingeborg whether TOI has a more recent meta-analysis than the one in the handbook of road safety measures (these will be in Norwegian and will be coded by Ingeborg). Don't code papers included in the meta-analysis separately. Include meta-analysis not published in English if possible.
- If there is a recent meta-analysis, next prioritise studies published more recently than the meta-analysis. Followed by other relevant studies conducted before the meta-analysis but not included in it. Check the aims of the meta-analysis to see if all papers of interest would have been included.
- General points for prioritising:
  - Outcome measures: crashes before incidents before observed safety performance indicators before self -reports.
  - Country of origin: Europe before US/Australia/Canada before other countries.
  - Most recently published (older studies of particular relevance can be included).
  - Importance: number of citations.
  - Language: Studies published in English (relevant other language papers can be included).
  - Source: Peer reviewed journals (non-peer review sources can be included).
- Each paper to be reviewed in turn in descending priority.
  - Papers meeting all inclusion criteria and are suitable for coding should be coded.
  - Papers not suitable for coding should be noted with the reason coding is not possible.

**Minimum requirements:**

- Record the steps used to prioritise coding.
- Record which papers were coded and which were not (and why).

## 4.5 CODING

Don't code lots of studies for the sake of coding lots. The number of studies coded should reflect the size of the topic area. Confirm with WP Leader how many papers are sufficient. The WP Leader may advise that time is better spent by expanding the number of topics coded rather than increasing the number of papers coded within a topic.

**Minimum requirements:**

- Meta-analysis papers' references should be added to the relevant WP coding control sheet (so that they are not double-coded).
- Excel coding sheets will be saved for each included paper.

**Which studies to code:**

In principle only data concerning the relevant risk factor will be coded. In cases where there is no data available which can be coded and the study complies with the search criteria – only the abstract is coded. This way, studies included in the review, which are important for the topic, can still be included in the DSS.

**Which effects to code:**

Some studies can include many estimates of effects – often several different estimates for the same effect. Always include the best estimate only (corrected effects, not uncorrected effects; optimal model form when several have been evaluated).

In the case of many detailed results consider only reporting the super-categories (e.g. just main scales, but not subscales). Try to report the most general results without "hiding" crucial differences (if the effect of sub-categories differ substantially it would be wrong to only report the main effect).

Code only your topic, but inform other partners, who code the other topics (see taxonomy). (Exception: you find the result worthwhile to include into DSS, but it does not belong to any other risk-factor in the taxonomy either).

## 4.6   QUALITY CONTROL

Each Work Package Leader is responsible for the quality of coding within their WP. The following approaches for quality control are suggested:

- Coded studies be cross checked e.g. by a second partner or a second person in the same partner organisation.
- The WP Leader to check at least 1 coding example from each partner, to confirm consistency between partners.
- If partners have specific difficulties with particular studies they should contact their WP Leader.

WP Leaders to circulate some coded examples and summaries of common study methodologies/topics within their area for partners to refer to. These key examples will then be consolidated into WP3 and can be discussed between WPs.

# 5 Coding Studies for the Repository

This chapter explains how to code studies in the repository of studies of risk factors and safety measures. It needs to be studied closely to ensure consistent coding across partners and Work Packages.

## 5.1 OBJECTIVE

One of the main objectives of the SafetyCube project is to create a repository of estimates of risk factors and safety effects. While there are already a number repositories of safety effects around (CMF clearinghouse; Australian Clearinghouse), these are tailored to infrastructural measures. Here we want to apply a much broader scope, comparable e.g. to the Handbook of Road Safety Measures (Elvik, 2009), where measures directed towards infrastructure, vehicles, and human behaviours are evaluated.

The collection of different types of studies, based on different underlying theories of crash causation, using different designs, analysis methods and variables constitutes a big challenge for the creation of a joint database for all these studies which is on the one hand flexible enough to capture important details of different types of studies but on the other hand allows to compare studies even across domains. The coding template was developed with the goal of creating such a database.

## 5.2 GENERAL STRUCTURE OF THE TEMPLATE

The template consists of an Excel-file with several sheets:
- Core info
- Results
- Summary
- Flexible info
- Custom info
- $exposure
- $outcome

Each of the three "info" sheets lists a number of variables in the rows. The corresponding values can be specified in the columns.

The variables contained in the "Core info" sheet are core variables that should be considered for every study. The "Flexible info" sheet contains flexible variables that should only be used when they are relevant for coding the specific study at hand. The "Custom info" sheet is intended either for proposing new variables and their possible values or for proposing new values/levels for existing variables.

To see an overview of all variables in Core, Flexible, and Custom info, right mouse-click on tab-names, select "unhide". Select $codes. On this sheet, here you see a complete list of all variables and their possible values. You can copy from this sheet, but not edit it.

The "Results" sheet is used to provide the numerical and statistical details of *effects* that are reported in a given study. These effects always quantify a particular association between **exposure** (either to a risk factor or a countermeasure) and a road safety **outcome**.

The different types of exposure and outcome that are considered in a study are specified in two dedicated sheets "$exposure" and "$outcome", as explained in detail below.

---

The "Results" sheet *can* only be completed *AFTER* the complete design has been specified, using the "Core info", "$exposure", "$outcome" and possibly also "Flexible info" and "Custom info" sheets. As explained in more detail below, the "Results" sheet consists of a dynamic table that will automatically shape itself according to the design specifications that are made in the info sheets. When additional variables are included in the design, all lines in the result table will move down. Information that had already been entered will therefore be shifted away from the correct place.

---

The summary sheet is intended for a synthesis of the design and the conclusions.

Across the different sheets, <span style="color:blue">**variable names**</span> occur in the rows with a blue background. These cannot be changed by the user. The values/levels for each variable should be specified in the columns. There are 3 different value-formats, indicated by 3 different background colours:
Pink: <span style="color:magenta">**Free text input**</span>
Green: <span style="color:green">**A fixed list of possible input-values**</span>
Brown: <span style="color:brown">**Numerical input**</span>
For some variables, only one input is required (e.g. title, publication year, WP, etc.). For other variables more than one value can be specified when needed by using multiple columns. This difference is made clear by the cell colouring and borders.

The following sections provide a detailed description of the content and logic of the sheets and how the user should proceed when completing the template.

---

*ALWAYS START FROM A BLANK TEMPLATE, and not from earlier documents, as this can/will generate input errors.*

---

For numerical values, make sure they are entered **consistently with the numerical format** (in particular the decimal separator) in your Excel version. In Anglosaxion articles the dot is used as decimal separator. For Excel versions that use a comma for that, there are two options:
**1.** Change in Excel the default setting to the English version (dot as decimal separator):
File / Options / Advanced / Use system separation -- here you can change "." and ",".
**2.** Alternatively use the settings of your own version. Excel will automatically change all decimal commas to dots when opened in an English version. However, be sure that you are consistent with your versions settings and don't type over the figures with dots from an article you are coding.

## 5.3 "CORE INFO" SHEET

### 5.3.1 Coder

The variables in this section (*name*, *institution* and *date*) serve to identify the person that was responsible for coding a certain study. This information is important to trace and solve potential problems during the automatic transfer of the coding sheets to the database or when questions are raised about the content.

---

*After submitting a study to the database, coders might be contacted for clarifications/corrections.*

---

### 5.3.2 Reference

Give the full journal reference. It is essential that end-users of the repository can return to the original study. Please make sure that all information is available in this section to allow that. It is advisable to use copy-paste functions as much as possible to avoid typo's in **title, author names (separated by semicolons)**, **source (e.g. journal name and issue)** and **URL (doi)**. However…



*… when COPY-PASTING information into or within the template, ONLY PASTE VALUES. Excel provides several ways to do this.*
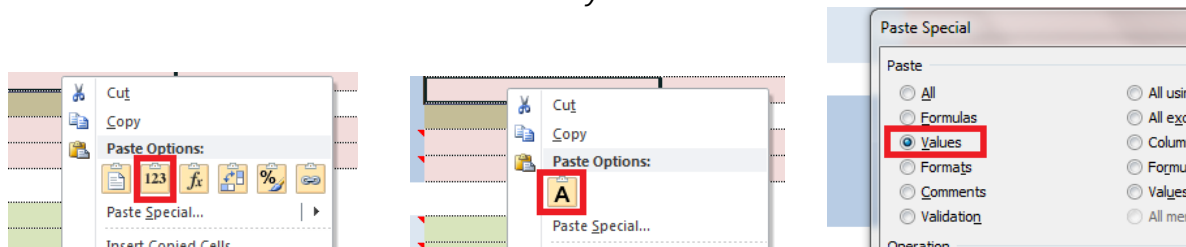
NEVER USE <u>CUT</u>-AND-PASTE. If you need to remove values, USE DELETE instead.

### 5.3.3 Topic

The variables that can be selected here are based on the SafetyCube taxonomy which forms the backbone of the DSS-database. The taxonomy allows users to find studies in the DSS and to link risks to measures. It is therefore essential that the taxonomy level to which a study belongs is correctly specified.

- After completing the first two fields (*Risk factor or Countermeasure* and *WP*), the corresponding taxonomy fields will appear with the corresponding list items (in alphabetical order).

- Coders always need to select a main level in the taxonomy. However, they can choose not to link the study with more specific levels in the taxonomy by selecting "NA". If NA is chosen for a given level, all subsequent levels also need to be "NA".

- VALID:

| Topic | | Risk factor or Countermeasure? | Risk factor |
|---|---|---|---|
| | ☐ WP | | WP5 |
| | ☐ Header 5 - Infrastructure element | | Alignment-junctions |
| | ☐ Header 6 - Risk factor | | NA |
| | ☐ Header 7 - Specific risk factor | | NA |

- INVALID:

| Topic | | Risk factor or Countermeasure? | Risk factor |
|---|---|---|---|
| | ☐ WP | | WP5 |
| | ☐ Header 5 - Infrastructure element | | Alignment-junctions |
| | ☐ Header 6 - Risk factor | | NA |
| | ☐ Header 7 - Specific risk factor | | insufficient ramp length |

- It is possible to select multiple WPs and/or taxonomy levels and check the corresponding rows as design variables (see below for more explanations on checkboxes for design variables). This is to accommodate studies where effects are reported that belong to different "leafs" in the taxonomy "tree" or even to different WPs.

- If the article contains information for different taxonomy fields or even different WPs, coders should code what is important for their topic and send the article to the WP-leader responsible for the other topic(s). However, there is an exception to that rule if the other topic is not included in the taxonomy but nevertheless found worthwhile to include in the DSS: the effects should be included in the present coding template. The Work Package and possibly high-level taxonomy fields should be selected according to the coders' understanding and set to NA for the other taxonomy levels.

| Topic | Risk factor or Countermeasure? | Risk factor | |
|---|---|---|---|
| | ☐ WP | WP4 | |
| | ☐ Topic | Age | Diseases and disorders |
| | ☐ Subtopic | Elderly (65+) | Diabetes |
| | ☐ Specific risk factor | Elderly (65+) | Typ B |
| | ☐ | | |
| | ☐ | | |

The design check-box allows to include different taxonomy levels in the result sheet.

| ☐ Differences between effects | Effect 1 | Effect 2 |
|---|---|---|
| Topic | Age | Diseases and disorders |
| Subtopic | Elderly (65+) | Diabetes |
| Specific risk factor | | Typ B |
| | | |
| Measure of effect/association | | |
| Specifications | | |

The last two variables concern the abstract and keywords. Providing an **abstract** will help users to grasp the specific context of a study. If no abstract is available, please consider writing a brief description of the study yourself. Providing a list of **keywords** is very helpful, as it will allow the clustering of studies and improve the database search options. Separate keywords by semicolons.

As an alternative to typing in a list of keywords, asterisks can be used to highlight *keywords* directly in the abstract.

### 5.3.4 Sampling Frame

These variables are used to specify the *scope* of the study and are also used to define the *design*. Each variable can take one or more values. If the information does not apply to a study (e.g. crash severities in a simulator study) or if it is not specified leave it empty. Multiple values are simply specified using different columns. The columns are used freely, i.e., there is no need to align values across different rows.

---

*Please always fill-in columns continuously; avoid empty cells.*

***RIGHT***

| ☐ Road user profile - Modes | Pedestrian | Cyclist | PTW | | |
|---|---|---|---|---|---|

***WRONG***

| ☐ Road user profile - Modes | | Pedestrian | | Cyclist | PTW | |
|---|---|---|---|---|---|---|

---

Like any numerical variable, the variable *Road user profile - Age* can be used to specify multiple age categories. To achieve this, use a hyphen "-" for closed categories and ">" or "<" for open categories.

| ☐ *Road user profile - Age* | <21 | 21-40 | 40< |
|---|---|---|---|

*Note that these rules apply to all numerical fields (i.e., cells with a brown background, e.g. Speed limit: 21-30, >50, <90, 121-140).*

At this point it is important to explain the concept of ***design variables***. Perhaps the most critical part of the coding of studies is to provide a correct definition of the study design. All variables in the sampling frame are preceded by a checkbox: ☑. When the study differentiates the results, i.e., effects, with respect to a given variable, this checkbox should be checked and the different levels of the variable that are concerned should be specified. For example, if a study examines the gain in road safety by daytime running lights and does so separately for cars and motorcycles, the checkbox in front of *Road user profile – Modes* should be checked. Additionally, the values *car* and *PTW* should be provided in the columns. If the study would instead make a distinction between (a) cars and vans and (b) motorcycles, the values *car*, *LGV* and *PTW* should be provided. (See information "custom info sheet" and "flexible info sheet" below for missing variables or categories).

*If a CHECKBOX is checked for a variable, this means that the variable qualifies as a DESIGN VARIABLE, i.e., the different values for that variable define different effects in the Results sheet. If a study includes different values for a given variable in the sampling frame, but the results are not differentiated with respect to these values, the variable should not be checked as a design variable. However, it is still important to provide the different values for the purpose of determining the scope of a study (e.g. all vehicle types included in a study).*

### 5.3.5 Design

The section starts with a field where the basic **features of the design** can be given as keywords. These keywords correspond to the terms described in the chapter on study designs of the guidelines. Several features can be selected.

The following fields are critical for completing the correct specification of the design and the correct rendering of the Results table in particular.

The "Direction" field is concerned with the direction in which effects are specified. There are two options: **≠ Exposure -> ≠ Outcome** or **≠ Outcome -> ≠ Exposure**.
When the first option is selected (**≠ Exposure -> ≠ Outcome**), effects denote a change in an outcome variable with respect to different exposure levels to risk factor or countermeasure (test versus reference). The result table will then show for each effect the conditions of exposure (e.g. exposed/not exposed). The outcome is only presented if different outcomes are measured. This should be chosen for all experimental and quasi-experimental designs, where the investigator (or another party) manipulates exposure to a risk factor or countermeasure (e.g. a road safety campaign) and measures the effect this generates on (**->**) outcomes (e.g. attitudes towards road safety). In case of doubt it can also be selected for cross-sectional studies.

In case-control or case-control-like designs, investigators instead start by identifying different outcomes and study differences between these outcomes in the extent of exposure to the risk-factor or countermeasure in question. Effects thus denote difference in exposure to a risk factor or countermeasure with respect to different outcome levels, namely, cases and controls. Hence, one needs to choose the second option: **≠ Outcome -> ≠ Exposure.**

For observational study-designs it is difficult to decide the direction. As indicated in Table 3.3, there is not real direction in the analysis. Many studies process data that can be put in a 2 X 2 table:

| | | Outcome (e.g. Injury) | |
|---|---|---|---|
| | | Yes | No |
| Measure/risk-factor | Exposed | n1 | n2 |
| | Not exposed | n3 | n4 |

and the measure of effect is often an *odds-ratio*, which is (n1/n2)/(n3/n4) or (n1/n3)/(n2/n4).

In this specific case both directions are correct. Formally both ways of entering the results are equivalent (because (n1/n2)/(n3/n4)=(n1/n3)/n2/n4)). Depending on the direction selected, the result tables in the excel-template will look different. This has, however, no effect on the information that is contained in the database.

If one selects *Exposure->Outcome*. The result table specifies which exposures to a risk factor or countermeasure have been compared. The measure of effect (odds ratio), can for clarity, be labelled as the odds to become injured.

| ☐ Differences between effects | Effect 1 |
|---|---|
| MyRiskFactor - Test group | Exposed |
| MyRiskFactor - Reference group | Non-exposed |
| | |
| Measure of effect/association | Odds ratio |
| Specifications | Odds of injury |
| | |
| Estimate | |
| Standard error of estimate | |

If one selects *Outcome -> Exposure* the result table specifies which outcomes were compared and the odds ratio should be labelled as the odds to be exposed to the risk/measure (e.g. odds of wearing a helmet).

| ☐ Differences between effects | Effect 1 |
|---|---|
| Injury - Cases | Yes |
| Injury - Controls | No |
| | |
| Measure of effect/association | Odds ratio |
| Specifications | Odds of exposure to risk-factor |
| | |
| Estimate | |
| Standard error of estimate | |

---

*Choosing the right option for the "Direction" field is critical, as it determines whether in the "Results" sheet effects denote differences in outcomes (≠ **Exposure -> ≠ Outcome**) or exposures (≠ **Outcome -> ≠ Exposure**).*

---

The following two fields cannot be edited directly. The coder needs to follow the links to the "$exposure" and "$outcome" sheets. They provide a dedicated interface to specify, respectively, the exposure and outcome variables in the study. Please consider that correct input in these fields is perhaps the most critical part of the coding activity.

Once coders have identified which variables function as exposure variables (i.e., variables that quantify or qualify the exposure to a risk factor or a countermeasure), and which as outcome variables (i.e., variables that quantify or qualify the outcome of risk factors or countermeasures for road safety), they can proceed to complete the dedicated sheets. In both sheets there is room for one or more variables, as it is often the case that exposure/outcome is measured/operationalized in different ways.

The final field in the "Design" section simply requires the coder to specify the total number of effects for which details will be provided in the "Results" sheet.

### $exposure

For exposure variable specifies the conditions of the risk factor or countermeasure. One needs to specify a name, a description/specification, the data type and levels where needed. The **name** is chosen freely and should give a good impression of what was measured as it is the label in the output tables. More details about a particular exposure variable can be given in the "**Specifications**" field.

It is critical to reflect on and to correctly specify the **data type** of each exposure variable.

In (quasi-)experiments (≠ Exposure -> ≠ Outcome), it will often be the case that exposure is dichotomous (cf. Exposed/Non-exposed) or categorical. In these cases, level names need to be provided. In the dichotomous case, the default names are "Exposed" and "Non-exposed", but this can be changed, again, following work-package-specific conventions as much as possible. For the "Categories" data type, all the level names need to be specified.

When exposure is compared across different levels of outcome (i.e., all case-control-like designs; ≠ Outcome -> ≠ Exposure), the exposure identified among cases and controls could be dichotomous (e.g. alcohol: yes/no), categorical (blood alcohol: <.05, >.05<.08, >.08), or continuous (e.g. blood alcohol concentration). For continuous variables the coder can choose between "Counts" and "General numerical". Of course, no level names are required in these cases.

Although in Exposure -> Response designs the exposure is mostly categorical, there are also analytical observational studies where outcomes are presented as a linear or non-linear *function* of a numerical exposure variable (e.g. % fatalities as a function of impact speed). The measures of effect are then slope estimates or estimates of non-linear function parameters. We return to this issue when discussing the "Results" sheet.

### $outcome

Here you have to specify the type of data that were collected to measure "road safety" (most often crash or injury data, but they can also be reaction times, speed driven, etc.). It is important to differentiate between these outcomes (data) and the measures of effects that are calculated with them (e.g. odds-ratios, mean-difference, etc.). The $outcome sheet follows the same principles as the $exposure sheet, but there are a number of important differences.

The coder is again free to choose a **name** (in accordance with the Work Package conventions), but here one also needs to select a **type of outcome** variable from a fixed list.

The specifications, data type and level fields need to be used in the same way as for exposure variables. Other than for exposure variables, dichotomous variables are indicated by selecting "Yes/No" as the data type. There is also an extra data to be used when studies make use of offset variables when analysing count data, i.e., *Rate (Count per …)*. The specific nature of the offset variables should be described in the specifications field.

> *Importantly, when the type of outcome is crash or injury and the data type is Yes/No or Categorical, coders should specify the level names by selecting the correct codes form the predefined lists that is provided*

> *Injury/crash characteristics might appear as design variables and/or as outcome levels. If they only function as outcome levels, these levels may also be specified in the core/flexible part, but they should not be checked as design variables.*

### 5.3.6   Limitations / Potential sources of bias

The final section in the core sheet is concerned with methodological shortcomings of the study. In the left-most column, several types of limitations or potential sources of bias can be selected. For each item, the coder should give a general indication of the extent to which it presents a problem for the interpretation of the study results (i.e., "Maybe a problem" or "Definitely a problem") as well as written motivation. Please consider this carefully and be aware that authors of a study are potential users of the repository and, in any case, deserve an adequate explanation for each shortcoming that coders identify. For a discussion on potential sources of bias in relation to specific design features, we refer to the guidelines document. Note that there is also an "Other" category, which allows the coder to report a particular problem that is not found in the list that is provided. However, this option should be considered carefully as it inevitably adds complexity to the coding system. Please refer to Table 3.2 in Section 4.6 for an overview of the most common biases for each study design.

## 5.4   "FLEXIBLE INFO" SHEET

The flexible info serves to code the sampling frame of the study with respect to variables that are not (or at least less) generally applicable than the variables in the "Core info" sheet. The sheet thus simply provides an extension to the sampling frame section in the "Core info" sheet and should be used in the same way.

Recall that multiple values are specified using different columns and, if these values define different effects, the variables should be identified as design variables using the checkbox on the left.

Within the flexible coding sheet there is the possibility to give more details on the selection criteria for the crashes that have been included into a study. In the following the meaning of these variables and in particular their application to opponent vehicles will be explained.

**Figure 5.1** Variables describing the crash characteristics level

- *Collisions* – describes different types of collisions (e.g. rear-end, frontal side, etc.).
- *Position in the vehicle* - It details all the main seating position available in a vehicle.
- *Accident type* – describes the accident from the perspective of the driving situation prior to the accident.
- *Accident - Vehicle/user type\** - The values are the same as in the variable *Road user profile – Modes* in the core info sheet. **Use this variable only if you want to enter separate values for each opponent**.
- *Accident - Injury severities\** - The values are the same as in the variable *injury severities* in the core info sheet. **Use this variable only if you want to enter separate values for each opponent.**
- ***Accident - Type of impact\*- it defines the vehicle impact area***.
- *Accident – EES\** - Energy Equivalent Speed - equivalent speed at which a particular vehicle would need to contact any fixed rigid object in order to dissipate the deformation energy corresponding to the observed vehicle residual crush
- *Accident – DeltaV\** - vector difference between impact velocity and separation velocity
- *Accident - Relative speed* - vector difference between impact velocity and velocity of the centre of gravity of a vehicle/object struck immediately before impact
- *Accident - Collision speed\** - velocity of the centre of gravity (CG) of an accident -involved vehicle immediately prior to impact
- *Accident - Initial/driving speed\** - speed of an accident-involved vehicle before any accident-related events. Accident-related events could be avoidance manœuvres or an unstable situation.
- *Accident - Seat belt use\** – This variable defines the road user seat belt status
- *Accident - Airbag availability\** – The variable defines the activation status of the airbag relevant for the studied crash type.
- *Accident - Vehicle model year\** - The model year of a product is a number used to describe approximately when a product was produced, and it usually indicates the coinciding base specification (design revision number) of that product.

- *Accident - Year of first registration\** – it states the year of the first registration of a motor vehicle with a government authority, either compulsory or otherwise.
- *Accident - Overlap (%)\** - The overlap is defined by the percentage of the deformed area in contact with collision partner at the time of maximum exchange of the collision forces.
- *Accident - Collision angle* - The collision angle is measured from the positive longitudinal axis of the concerned vehicle towards the positive longitudinal axis of the collision opponent vehicle.
- *Accident CDC: Directions of force\** - the Principal Direction of Force (PDOF) during impact is the direction of the force that caused the crush and sheet metal displacement on the damaged vehicle. The PDOF is determined by the vector result of forces normal and tangential to the surface of the vehicle in the area of deformation. The PDOF is designated by hour sectors on a conventional clockface and referred to as the clock direction. 12 O'clock indicates a head-on impact, 06 O'clock refers to a rear end impact, 03 and 09 O'clock refer to perpendicular impact to the right and left sides respectively.
- *Accident - CDC: Areas of deformation\** - the Deformation Location Code specifies the general area of the vehicle which sustained damage – either the Front, Right, Back or Left side. The Top and Underside of the vehicle may also be damaged.
- *Accident - CDC: Types of damage distribution\** - Some entries violate the assumptions of analysis, notably Sideswipe, Rollover and multiple impact (K). Narrow impact is specified, the width of the damage profile is set to 16in (410mm); otherwise Wide impact is assumed and the width of the damage is set based on tabular data in conjunction with the Specific Longitudinal or Lateral Location of Deformation
- *Accident - Emergency manœuvre\** – It describes the manoeuver realized by the road user to avoid the crash
- *Accident - Guidance problems\** – It describes the lane departure mechanism

For all variables marked with a star*, it is possible to give separate criteria for the vehicle with the system and for its opponent. The starred variables give the coders the possibility to combine existing levels using "<>" (no space between < and >).

The most important points to take into consideration:
- The vehicle defined on the left part of "<>" is **the vehicle with countermeasure**/risk factor.
  - As an example: "Truck <> Van", means that the vehicle with the system were all trucks that collided with a van. "Van <> Truck" means that the vehicles with the system were vans.
- If you have **several values** for the vehicle with the system and/or the opposing vehicle, there are two ways to code this: A list of combinations or a combination of lists. As an example, consider a study in which blind-spot systems in trucks and vans are tested in crashes with pedestrians, cyclists, or PTW.
  - **List of combinations:** "Truck <> Pedestrian"; "Truck <> Cyclist"; "Truck <> PTW"; "Van <> Pedestrian"; "Van <> Cyclist"; "Van <> PTW". You can mark this variable as design variable and report effects for each combination separately.
  - **Combination of lists:** "Truck; Van <> Pedestrian; Cyclist; PTW". In this case you can only report effects for all combinations together.
- It is possible to define both sides of "<>" or only the left side or the only the right side.
  - "<> Frontal impact" means that there is no impact criteria for the vehicle with countermeasure/risk factor. And the opponent vehicle sustained a frontal impact.

In practice, coding opponent values means that you have to **type in** (part of the) **values yourself** rather than just selecting them from the drop-down list. Please regard the following rules:

- When using "<>" in the starred variables, please *use the list of possible input-values* in order to guaranty coding homogeneity.
- All spaces are ignored (except for space between < and >):
  - "Truck <>Van" = "Truck<>Van "= "Truck<> Van" =" Truck <> Van"
  - ≠ "Truck < > Van"**!**
- Use the upper part of excel to define an opponent characteristic (see figure below)
- When defining an opponent characteristic, a warning message appears. Just click "yes".
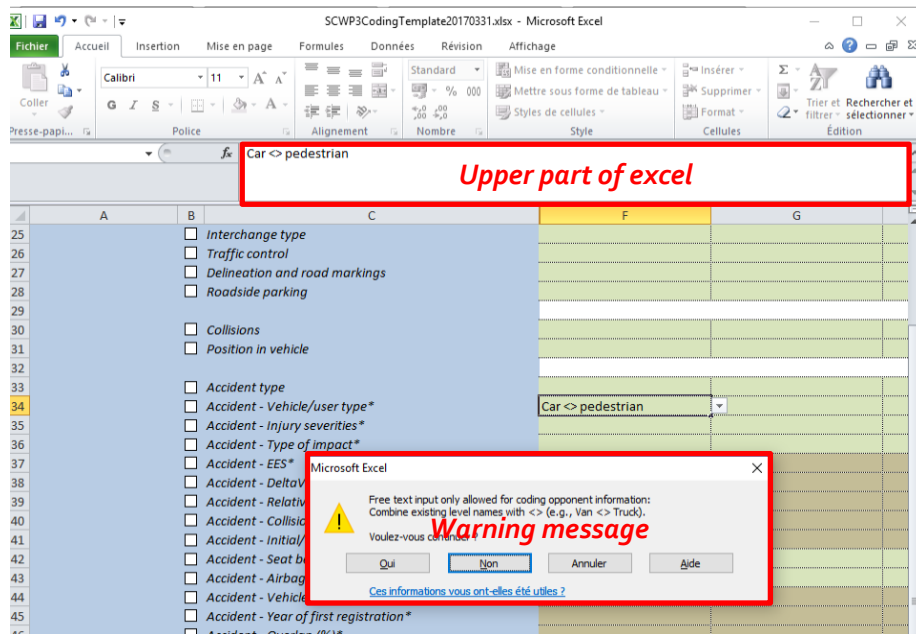


## 5.5 "CUSTOM INFO" SHEET

Especially at the beginning of the coding activities in the project, it might sometimes occur that variables or values/levels that are needed for a correct representation of the study and its results are not available yet, neither in the core nor flexible part of the sampling frame.

In those cases, the "Custom info" sheet can be used to formulate a proposal. After the proposal has been formulated following the procedure described below, the entries are to be used in the same way as those in the "Flexible info" sheet.

> *Please make sure to only formulate proposals that are absolutely necessary. It is critical to consider whether a new variable is really needed. The coder should analyse whether the issue that is encountered can be solved by combining the levels of existing variables. If the issue can be solved by simply proposing new values/levels to existing variables, this also overrides the need for proposing a new variable. In general, each coder shares the responsibility to keep the level of redundancy in the repository as low as possible, as redundant variables/levels compromise the usability of the repository.*

### 5.5.1 New variables

The coder needs to provide a name for the variable and specify the levels. Often, the levels that are relevant to code a particular study are not exhaustive. The coder is urged to think of all possible values/levels a new variable might have. Values/levels that exist, but are not present in the study at hand should be provided as well, but within square brackets [...] (see example below).

| ☐ Motor vehicle - Number of wheels | 2 | 3 | [1] | [4] |

### 5.5.2 New values/levels

If a coding issue can be solved by adding levels to an existing variable in the sampling frame, the coder should first of all *copy the name of the existing variable* in the first column (recall that one should always use "paste values"). Next, the names of levels that are needed to code the study, including any existing level names should be provided.

| ☑ Injury severities | Moderate | AIS 3 | AIS 4 | |

> *When new values/levels are proposed in the "Custom info" sheet, <u>do not use the original variable in the "Core info" or "Flexible info" sheets anymore</u>. Always verify that the name of the variable and any existing levels that are provided in the "Custom info" sheet MATCH THE SPELLING OF THE ORIGNAL NAMES EXACTLY. To copy & paste values from the original variables, right mouse-click on tab-names, select "unhide". Select $codes, here you see a complete list of all variables and their possible values. You can copy from this sheet, but not edit it.*

## 5.6 RESULTS

As already noted, this sheet contains a table that will automatically shape itself according to:
   a. the number of effects that is indicated (see "Core info" sheet),
   b. the different design variables that are indicated (see "Core info", "Flexible info" and "Custom info" sheets),
   c. the choice with respect to the "Direction" in the design (see "Core info" sheet),
   d. the details of exposure variable(s) (see "$exposure" sheet) and/or,
   e. the details of the outcome(s) (see "$outcome" sheet).

Note that only those categories are included into the results sheet that are necessary to identify effects within the study but not information that is the same for all effects. Therefore for exposure -> outcome designs the outcome variable is usually not included (except if several different outcome variables have been specified) while for outcome->exposure designs the exposure categories are not presented, unless several exposure variables have been specified.

> *It is essential to verify that the structure of the results table indeed corresponds to the structure of the effects reported in the study BEFORE filling in the table.*

| Differences between effects | Effect 1 | Effect 2 | Effect 3 | Effect 4 | Effect 5 |
|---|---|---|---|---|---|
| Injury nature | Fracture; Internal; Open Wou | Fracture; Internal; Open Wou | Fracture; Internal; Open Wou | Fracture | Fracture |
| Injury severities | Moderate | AIS 3 | AIS 4 | AIS 3 | AIS 3; AIS 4 |
| | | | | | |
| Injury - Cases | Hospital; Head | Hospital; Head | Hospital; Head | Hospital; Head | Hospital; Head |
| Injury - Controls | Non-Head; Minor head | Non-Head; Minor head | Non-Head; Minor head | Non-Head; Minor head | Non-Head; Minor head |
| | | | | | |
| Measure of effect/association | Odds ratio | Odds ratio | Odds ratio | Odds ratio | Odds ratio |
| Specifications | Odds for wearing a helmet | Odds for wearing a helmet | Odds for wearing a helmet | Odds for wearing a helmet | Odds for wearing a helmet |
| | | | | | |
| Estimate | 0.5060 | 0.3790 | 0.2570 | 0.4370 | 0.2170 |
| Standard error of estimate | | | | | |
| Statistic [name(parameters)=x] | | | | | |
| p-value | <0.0001 | <0.0001 | <0.0001 | 0.1710 | <0.0001 |
| Sample size (x or n1=x1; n2=x2) | n (cyclist casualties)= 6745 | n (cyclist casualties)= 6745 | n (cyclist casualties)= 6745 | n (cyclist casualties)= 6745 | n (cyclist casualties)= 6745 |
| Confidence level | 0.9500 | 0.9500 | 0.9500 | 0.9500 | 0.9500 |
| Lower limit | 0.3880 | 0.2670 | 0.1480 | 0.1300 | 0.1320 |
| Upper limit | 0.6590 | 0.5360 | 0.4480 | 1.4660 | 0.3570 |
| | | | | | |
| Adjustment variables/Covariates | Speed limit; Collision vehicle | Speed limit; Collision vehicle | Speed limit; Collision vehicle | Speed limit; Collision vehicle | Speed limit; Collision vehicle |
| Conclusion | Significant positive effect on | Significant positive effect on | Significant positive effect on | Non-significant effect on roa | Significant positive effect on |

> *It will often occur that specific values need to be repeated many times across the results table. In Excel, a single cell value can be pasted into several other cells at once simply by selecting those cells before giving "paste" command. AGAIN, PLEASE MAKE SURE ONLY TO PASTE VALUES (SEE ABOVE).*

### 5.6.1 Design variables

If one or more variables in the sampling frame have been identified as design variables, these will occur at the top of the left-most column (see e.g. "Injury nature" and "Injury severity" in the example above).

For each effect (column), the coder can specify the corresponding levels of the design variables by selecting them from the drop-down list.

If effects correspond to a combination of different levels of a single design variable, the names of these levels need to be given within the same cell, separated by **semicolons**. This is also illustrated in the example above. Do not use "All" for effects that were calculated across all conditions treated in the study – this becomes meaningless outside the context of the study.

### 5.6.2 Exposure/Outcome variables

After the effects are defined with respect to the design variables, the specific **contrast** with respect to exposure or outcome levels should be determined for each effect.

If the direction of the design is **≠ Outcome -> ≠ Exposure**, as in the example above, each outcome variable X that occurs in the "$outcome" sheet will be represented for "X – <u>Cases</u>" and "X – <u>Controls</u>". In the example above, there is only one outcome variable ("Injury") and the effects differ with respect to the specific combinations of the levels that have been provided (using semicolons). Note that this design requires that all outcome variables are defined as dichotomous or categorical data types.

For **≠ Exposure -> ≠ Outcome** the contrast is instead defined with respect to one or more exposure variables. There are two possibilities here, depending on whether one is dealing with dichotomous/categorical exposure variables (as in most experiments or quasi-experiments) or with count/general numerical data types. For any dichotomous/categorical exposure variable X that occurs in the "$exposure" sheet, the coder needs to choose the level (or combination of levels using semicolons) with respect to the "X – Test group" and "X – Reference group", as illustrated below.

| ☐ Differences between effects | Effect 1 | Effect 2 | Effect 3 |
|---|---|---|---|
| Road network profile - Area | Urban | ▾ban | Rural |
| Road network profile - Segments | Intersection - Stop controlled | Intersection - Stop controlled | Intersection - Stop control |
| Accident severities | All | Injury | All |
| Road network profile - No. lanes per direction | 1 | 1 | 1 |
| | | | |
| Conversion to roundabout - Test group | Exposed | Exposed | Exposed |
| Conversion to roundabout - Reference group | Non-exposed | Non-exposed | Non-exposed |
| | | | |
| Measure of effect/association | Crash modification factor | Crash modification factor | Crash modification factor |
| Specifications | | | |
| | | | |
| Estimate | 0.28 | 0.12 | 0.42 |
| Standard error of estimate | 0.06 | 0.08 | 0.07 |
| p-value | | | |
| Confidence level | | | |
| Lower limit | | | |
| Upper limit | | | |
| | | | |
| Adjustment variables/Covariates | AADT | AADT | AADT |
| | | | |
| Conclusion | Significant improvement due | Non-significant improvement | Significant improvement c |
| | | | |
| Comments | | | |

As noted in the section on the "$exposure" sheet, ≠ Exposure -> ≠ Outcome designs might also in-volve effects that are slopes or parameters of non-linear functions of numerical exposure variables (including counts and general numerical data types). In those cases, there is no contrast with respect to a test and reference group. For instance, if the crash rates (outcome) are studied as a function of the age of a vehicle (exposure), the investigator might choose to report the results of a logistic re-gression with a linear age-effect (i.e., slope). Whenever exposure variables are identified as counts or general numerical data types in an ≠ Exposure -> ≠ Outcome design, the "Results" sheet will ask to specify the units of the variable in question (e.g. years).

| ☐ Differences between effects | Effect 1 | Effect 2 |
|---|---|---|
| Accident severities | Damage only | Injury |
| | | |
| Age of car - Unit | Years | Years |
| | | |
| Measure of effect/association | Slope | Slope |
| Specifications | slope in logistic regression | slope in logistic regression |
| Estimate | | |

### 5.6.3 Measure of effect/association

Depending on the specific design at hand and the method for statistical analysis, there exist several ways to quantify the change in outcome values across exposures (≠ Exposure -> ≠ Outcome) or the difference in exposure across different outcomes (≠ Outcome -> ≠ Exposure). Details about the dif-ferent possibilities are beyond the scope of these instructions and are provided in the guidelines document.

For each measure, further details about the measure itself or the statistical modelling that was in-volved can be provided in the "Specifications" field (see example above).

### 5.6.4 Numerical and statistical details

*The next set of variables concern the actual numerical values.*
***Estimate****: The value of the measure of effect/association.*
***Standard error of the estimate****: If available.*
***Statistic [name(parameters)=x]****: This field gives you the opportunity to report test-statistics like t-, Chi², or F tests. The parameters will usually be the degrees of freedom.*
    *Examples:*
    *t(37)=10*
    *Chi-square(2)=7.6*

*Note however, that statistical tests should only be reported when they concern the contrast between test and control group that is specified in the effect-column in question. An overall F-test that compares 3 or more conditions (even if one of them is the control group) is unfortunately not sufficient.*

*P-value: This can either be an actual value (e.g. .0234) or a cut-off ("<.001"; ">.2").*

*Sample size: It is often difficult to decide which entity the sample size refers to. As an example, if you have 26 crossings with crash counts, is your sample size then 26 or the total number of crashes? If the analysis is based on a 2X2 table (e.g. for odds-ratios) it might be handy to give all four cell sizes. We leave it to you to give the most informative indication of the size of the study. Make sure that each number is labelled so that others can understand to what they refer to – if that takes too much place, use n1, n2… and give explanations in the comment section.*

> *Examples:*
> *n(crashes)=…*
> *n(exposed)=…; n(non-exp)=…*
> *n1=…; n2=…; n3=…; n4= …*

*Confidence level: If confidence limits are reported, please specify the level (e.g. .80, .95).*

> *Lower limit: Smallest value of the confidence interval, if available.*
> *Upper limit: Largest value of the confidence interval, if available.*

### 5.6.5 Adjustment variables / Covariates

*These fields allow the coder to specify any variables that were put under numerical/statistical control. Please use semicolons to separate different variable names (e.g. "age; traffic volume").*

### 5.6.6 *Conclusions*

This variable is essential, as it codes the basic conclusion for each effect: *Significant negative effect on road safety*, *Significant positive effect on road safety*, *Non-significant effect on road safety*.

### 5.6.7 Differences between effects/Interactions

A final feature of the "Results" sheet is that one can specify the significance of differences between effects – if that information is provided. To do so, the coder should check the "Differences between effects" checkbox in the upper left corner of the sheet. Once checked, corresponding fields will appear at the bottom of the sheet, as illustrated below. Where applicable, the coder can indicate the presence/absence of significant difference by selecting from the drop-down list.

In the example below, the effect of an energy drink (Effect1) or an energy drink + rest (Effect2) was investigated. Both leading to a significant improvement in lane stability. Effect2 (energy drink + rest) was significantly larger than Effect1 (energy drink alone). This is coded below Effect 1 in the dark-blue fields, where a significant difference to Effect 2 is indicated.

| Differences between effects ☑ | Effect 1 | Effect 2 | Effect 3 | Effect 4 |
|---|---|---|---|---|
| OUTCOME DEFINITION | Lane stability | ⌄ ne stability | Steering stability | Steering stability |
| | | | | |
| Energy drink - Test group | Exposed | Exposed | Exposed | Exposed |
| Energy drink - Reference group | Non-exposed | Non-exposed | Non-exposed | Non-exposed |
| Rest - Test group | Non-exposed | Exposed | Non-exposed | Exposed |
| Rest - Reference group | Non-exposed | Non-exposed | Non-exposed | Non-exposed |
| | | | | |
| Measure of effect/association | Absolute difference | Absolute difference | Absolute difference | Absolute differe |
| Specifications | | | | |
| | | | | |
| Estimate | 0.08 | 0.13 | | |
| Standard error of estimate | | | | |
| p-value | | | | |
| Confidence level | | | | |
| Lower limit | | | | |
| Upper limit | | | | |
| | | | | |
| Adjustment variables/Covariates | | | | |
| | | | | |
| Conclusion | Significant improvement due | Significant improvement due | Non-significant improvemen | Non-significant i |
| | | | | |
| *Differences between effects:* | | | | |
| *Effect 2* | Significant difference | | | |
| *Effect 3* | | | | |
| *Effect 4* | | | Non-significant difference | |
| *Effect 5* | | | | |
| *Effect 6* | | | | |

## 5.7 WHICH EFFECTS TO CODE?

### 5.7.1 Not completely reported effects

The coding template is based on the philosophy that quantitative results should be **tested statistically** and reported in sufficient detail so that **either the standard error or the confidence intervals** is known for each effect. In most research domains, **a paper not supplying this information would therefore be considered of minor quality** and should not be coded.

For some research domains, reporting statistical tests is however not common practice and it would be difficult to find results to code if applying this criterion. Therefore below a number of **exceptions to that general** rule are explained.

### Results without statistical testing

If the majority of results concerning a particular topic does not include testing for significance, it is advised to code the results anyway. In that case select "*Positive effect of road safety without statistical test*" or "Negative effect of road safety without statistical test". (Note however, that if the results *were* tested and they turned out to be not significant, "non-significant effect on road safety" – irrespective of whether the effect was in tendency positive or negative.) If no significance testing is reported it is all the more important to include any information (e.g. sample size) that gives an indication about the reliability of the results.

### Non-significant results

It is **very important to code non-significant results just as well as significant results**. Otherwise a strong bias (otherwise known as publication bias) would be introduced. A thought experiment: a measure has only a very small effect. Most tests yield null-results, however, a small number of studies found a significant effect. If these studies are selected – because they show significant results – and all others are left out, the impression would be given that the measure has reliably positive results and the size of the effect would be strongly overestimated.

In many research domains, the unfortunate habit has risen to report precise numerical results only for significant effects (and summarize the rest as "all other effects were non-significant"). ***Given the importance to also report non-significant effects, it is sometimes necessary to include effects for which not even the size of the effect has been indicated***.

Regression models form a special case: usually the authors departed from a large number of possible predictors but retained only the significant ones in the final model. If a summary would contain only studies in which a particular variable had been retained in the final model, a strong bias is created and it makes sense to check how many studies had the variable in their original list of predictors but dropped it (note however, that a variable can also be dropped from a model because another one is included that measures almost the same).

If (for whichever reason) an effect has to be reported for which it is only known whether it was significant or not, try to derive as much information as possible from the other effects reported in the paper (e.g. sample size, confidence level) and use the "Conclusion" field to report the effect.

**Results from graphs**

Generally, it should be avoided to code results that are displayed as graphic information rather than being directly reported in the paper. There can be exceptions to that rule, if the information in the graph is considered so important that it seems unwise to leave it out. This could happen if for a particular topic important studies present their results graphically rather than in numerical values. In that case, select "coded from graph" in the comment field below the conclusion field.

### 5.7.2 What if studies report many effects?

Many studies have coded a big wealth of effects. There can be no general rules which effects to code and which not.

- Only code effects that are related to the risk-factor or countermeasure identified under "Taxonomy" (5.3.3).
- If you want to code effects concerning different taxonomy levels you can select multiple categories and mark it as design variable (5.6.1).
- If results are established using different methods/models – choose the method that is concluded to be best by the authors (or by you).
- Code as many effects as necessary to convey all essential information and as few as possible.

Two examples can serve as inspiration how to decide which effects to code and which not.
**Example 1**, the evaluation of the effects of bicycle helmets by Bambach et al. (2013). It is based on hospital data about injured cyclists. In addition to helmet use, the study collected data on many potentially confounding factors, allowing the effects of these to be controlled in statistical analysis. A total of 18 estimates for the effect of helmets are presented in Table 4 of the paper. The estimates form two groups, depending on the type of comparison group that was used in the analysis. In each of the two main groups, estimates are presented for four types of injuries:

1. Head injury
2. Skull fracture
3. Intracranial injury
4. Open wound

Where Categories 2, 3, and 4 are sub-groups of Category 1. To avoid double counting, a choice must therefore be made between coding the results for the subgroups (2-4) or coding the results for all head injuries together (only 1). Unfortunately, for the fourth subgroup of head injuries (multiple or

other), no results are listed. Hence, results for this group would not be included if the sub-group results would be coded, but are included in the summary results for all head injuries. Thus, coding each of the types of injury results in a loss of information.

Another aspect that should be considered when deciding which effects to code is completeness. Here, a distinction was made between four levels of injury severity:

1. Moderate injury
2. Serious injury
3. Severe injury

However, only for "head injuries" as a total and for one subgroup (intracranial injury) results are reported for all 3 levels of severity. As the effect of helmets is largest for the most severe injuries, it seems to make most sense to code only the results for all Head injuries together, but keep the distinction between different severity types.

There is, in addition, a choice to be made between comparison group 1 and comparison group 2. It does not increase the informative value to code results for both comparison groups. In the first place, the results are quite similar, although the confidence intervals tend to be larger for comparison group 2. In the second place, users of the decision support system may start wondering why "the same" results are reported twice and which set of results should be trusted. One avoids any need for explaining this by simply coding one set of results. The statistics for the logistic regressions listed in Table 5 of the paper make it clear that comparison group 1 produces the most precise estimates and should therefore be preferred.

It is concluded that, all in all, the most informative level of detail is represented by the first three Odds Ratios reported. They are estimate for all head injuries together the effect of helmets on moderate, serious, and severe injuries, relative to comparison group 1. The other 15 effects reported in Table 4 of the paper are, for various reasons, less informative.

**Example 2** to be discussed is the meta-analysis of studies estimating the safety-in-numbers effect presented by Elvik and Bjørnskau (2017). Table 4 of the paper presents a total of 44 summary estimates of regression coefficients indicating a safety-in-numbers effect. The coefficients refer to motor vehicles, cyclists or pedestrians. A distinction is made between three levels of study: the micro level, the meso level and the macro level. For each level four summary estimates of regression coefficients are presented, based on:

1. Simple arithmetic mean; no statistical weighting
2. Mean weighted by the number of crashes
3. Fixed-effects mean based on meta-analysis
4. Random-effects mean based on meta-analysis

It is clearly not informative to present summary estimates based on all these approaches. In fact, only the meta-analyses make sense. The other two approaches were included because there are doubts among some meta-analysts about whether using inverse-variance meta-analysis to combine regression coefficients makes sense or not. As a check on whether a meta-analysis lead astray, simpler summary estimates were included for comparison.

Therefore, it is only informative to present the summary estimates based on meta-analysis. The next question is whether one should present the fixed-effect summary estimates, the random-effects summary estimates or both. It is again instructive to imagine oneself in the position of a user of the decision support system. From the user perspective, one might wonder why duplicate estimates of the same effect are presented and why they differ. Since the reason for this is rather technical and can only be fully understood if you know a little about meta-analysis, it would be confusing rather than enlightening to present both fixed-effects and random-effects estimates. Preference should go to the random-effects estimates, as these are, from the point of view of sampling theory, more general than the fixed-effects estimates.

Are there any systematic differences in summary estimates between the three levels of study? Considering the wide confidence intervals of estimates for each level, the answer must be no. It does not add any useful information to code the summary coefficients for each of the levels micro, meso and macro.

The conclusion is therefore that only three results are coded. These are the random-effects summary estimate of the coefficient for motor vehicles, cyclists and pedestrians. This example shows that although a study may present a large number of estimates, it does not follow that coding all of these is informative. In this example, coding the smallest number possible – three estimates – can reasonably be argued to be the most informative.

## 5.8 "SUMMARY" SHEET

*The "Summary" sheet is a regular sheet where the coder should synthesize the study, USING PLAIN TEXT ONLY. It is most practical to prepare the synthesis in a separate document and paste the text (as plain text) in the "Summary" sheet*

The text should explain which risk factor or countermeasure was investigated, how the design looked like, how the results were analysed and which conclusions were formulated. Coders should provide this information for all aspects of the study they are coding and should mention if there are other aspects they are not coding (or that are coded in another file because it concerns another risk factor / countermeasure). Summaries serve two purposes: (1) abstracts that are provided by authors do not always give all the appropriate information and (2) as in the reviewing process for scientific articles, an adequate summary shows that the coder identified all relevant aspects of a study.

# 6 Examples of Coded Studies

## 6.1 BEFORE-AFTER STUDY: STEP-BY-STEP EXAMPLE

*Bhagwant N. Persaud, Richard A. Retting, Per E. Garder, Dominique Lord (2011). Observational Before-After Study of the Safety Effect of U.S. Roundabout Conversions Using the Empirical Bayes Method. Transportation Research Record 1751, 1-8*

### 6.1.1 Core/Flexible/Custom info

This study asses the gain in road safety by roundabouts. Crash counts were compared before and after the conversion of 23 stop sign and traffic signal intersections in the US. To account for the phenomenon of regression to the mean as well as natural changes in traffic volumes the comparison was adjusted using the Empirical Bayes method.

### Coder - "Core info"

The variables in this section (name, institution and date) serve to identify the person that was responsible for coding a certain study. This information is important to trace and solve potential problems during the automatic transfer of the coding sheets to the database or when questions are raised about the content.

| Coder | Name | Kevin Diependaele |
|---|---|---|
| | Institution | BRSI |
| | Date (dd/mm/yyyy) | 08/12/2015 |

### Reference - "Core info"

It is essential that end-users of the repository can return to the original study. Please make sure that all information is available in this section to allow that.

| Reference | Authors | Bhagwant N. Persaud, Richard A. Retting, Per E. Garder, Dominique Lord |
|---|---|---|
| | Title | Observational Before-After Study of the Safety Effect of U.S. Roundabout Conversions Using the Empirical Bayes Method |
| | Year | 2001 |
| | Source | Transportation Research Record 1751, 1-8 |
| | URL | http://dx.doi.org/10.3141/1751-01 |

### Topic - "Core info"

The purpose of these variables is that end-users of the repository can quickly navigate towards studies that are of interest to them. First, a main distinction is made between studies that deal with road safety *risk factors* and those that assess the effectiveness of *countermeasures*. Second, studies will be grouped according to the SafetyCube *Work Packages* that are concerned (WP4: Behavior; WP5: Infrastructure; WP6: Vehicle). The following three variables serve to classify studies according to the *taxonomy* that is defined within each Work Package.

| Topic | Risk factor or Countermeasure? | Countermeasure |
|---|---|---|
| | WP | WP5 |
| | Header 5 - Infrastructure element | Alignment-junctions |
| | Header 6 - Countermeasure | At-grade junctions treatments |
| | Header 7 - Specific countermeasure | convert junction to roundabout |
| | Abstract | Modern *roundabout*s are designed to control traffic flow at *intersection*s without the use of stop signs or traffic s |
| | Keywords | |

The last two variables concern the abstract and keywords. Providing an *abstract* will help users to grasp the specific context of a study. If no abstract is available, please consider writing a brief de-

scription of the study yourself. Providing a list of **keywords** is very helpful, as it will allow the clustering of studies and improve the database search options. As an alternative to typing in a list of keywords, asterisks can be used to highlight *keywords* directly in the abstract.

Roundabouts clearly qualify as countermeasures to reduce collisions.
In the example, keywords are indicated using asterisks in the abstract, as follows:
*Modern *roundabout*s are designed to control traffic flow at *intersection*s without the use of stop signs or traffic signals. *U.S.* experience with … This interest has created a need for data regarding the *safety effect* of roundabouts. Changes in *motor vehicle crashes* following conversion of 23 intersections from stop sign and traffic signal control to modern roundabouts are evaluated. The settings, located in seven states, are a mix of urban, suburban, and rural environments with the urban sample consisting of both *single-lane* and *multilane* designs and the rural sample consisting of only single-lane designs. A *before-after* study was conducted using the *empirical Bayes* procedure, which accounts for regression to the mean and traffic volume changes that usually accompany conversion of intersections to roundabouts....*

---

*WARNING: when COPY-PASTING information into or within the template, ONLY PASTE VALUES. Excel provides several ways to do this.*
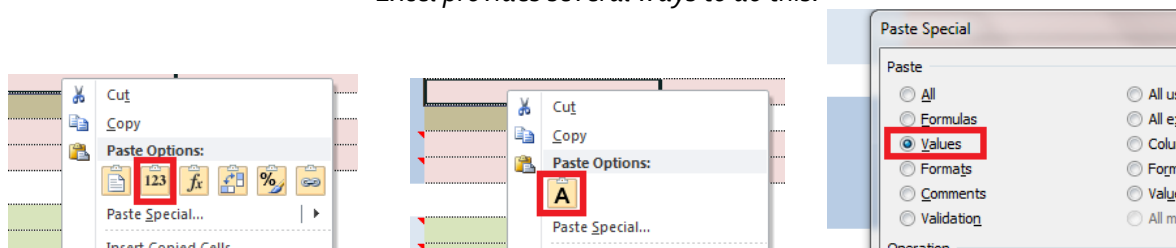


NEVER USE <u>CUT</u>-AND-PASTE. If you need to remove values, USE DELETE instead.

---

**Sampling Frame - "Core info", "Flexible info", "Custom info"**

These variables are used to specify the *scope* of the study and are also used to define the *design*. Each variable can take one or more values. Multiple values are simply specified using different columns. The columns are used freely, i.e., there is no need to align values across different rows.

| Sampling frame | | | | | |
|---|---|---|---|---|---|
| ☐ | Countries | United States | | | |
| ☐ | Administrative Level | National | | | |
| | | | | | |
| ☐ | Road user profile - Modes | Car | LGV | HGV | PTW |
| ☐ | Road user profile - Type | All | | | |
| ☐ | Road user profile - Subgroup | | | | |
| ☐ | Road user profile - Age | | | | |
| ☐ | Road user profile - Gender | | | | |
| | | | | | |
| ☑ | Road network profile - Area | Urban road | Rural road | | |
| ☑ | Road network profile - Segments | Intersection - Signalized | Intersection - Stop controlled | | |
| | | | | | |
| ☑ | Accident severities | All | Injury | | |
| ☐ | Injury severities | | | | |
| | | | | | |
| | Comments | "The analyses were confined to 7 states — Colorado, Florida, Kansas, Maine, Maryland, South Carolina, and Vermont | | | |

The study used data on the national level in the US. We code that the study was only concerned with motor-vehicle crashes by selecting all the corresponding modes "Car", "LGV", "HGV" and "PTW". Since the results are not separated for these modes, we do <u>not</u> identify this field as a design variable. The study does make a distinction in the results for rural and urban areas. It also distinguishes two types of segments: intersections that were converted from a signalized intersection and those that were previously stop-controlled. Separate results are also provided for all crash severities and injury crashes. Finally, in the "<u>Flexible info</u>" sheet, we can specify that the study distinguishes single (1) and multi-lane (2+) intersections.

| Sampling frame | ☐ Traffic volume (AADT) ☐ Relative share of road users (Mode : %) | | |
|---|---|---|---|
| | ☑ No. lanes per direction ☐ Speed limit (km/h) ☐ Speed limit (miles/h) | 1 | >1 |

Also on the "Flexible info" sheet, we code that the data were collected between 1992 and 1997.

| | ☐ Data collection - Start-Stop (dd/mm/yyyy-...) | 1992-1997 | |
|---|---|---|---|

### Design – "Core info"

The section starts with a field where the basic **features of the design** can be given as keywords. These keywords correspond to the terms described in the chapter on study designs of the guidelines.

We are dealing with a quasi-experimental design, since exposure levels (no roundabout, roundabout) are manipulated, however, not under the control of the investigators. The study follows a before-after design with repeated measures; the same intersections are tested before and after the conversion to a roundabout. The investigators control for selection bias (regression to the mean) and traffic-volume changes through the application of the "Empirical Bayes" technique.

| Design | Features | Quasi-experimental | Crossover/Repeated-measur | Before-after | Empirical Bayes |
|---|---|---|---|---|---|
| | Direction | ≠ Exposure -> ≠ Outcome | | | |
| | EXPOSURE DEFINITION | Conversion to roundabout | | | |
| | OUTCOME DEFINITION | Crash count | | | |
| | Total number of effects | 9 | | | |
| | Comments | | | | |

The "Direction" field is concerned with the direction in which effects are specified.

> *Choosing the right option for the "Direction" field is critical, as it determines whether in the "Results" sheet effects denote differences in outcomes (≠ **Exposure -> ≠ Outcome**) or exposures (≠ **Outcome -> ≠ Exposure**).*

In experimental or quasi-experimental studies investigators always start from a manipulation of exposure levels and monitor potential effects on outcome variables. Hence we choose the option "≠ Exposure -> ≠ Outcome".

The following two fields cannot be edited directly. The coder needs to follow the links to the "$exposure" and "$outcome" sheets. They provide a dedicated interface to specify, respectively, the exposure and outcome variables in the study. Please consider that correct input in these fields is perhaps the most critical part of the coding activity.

**$exposure**
There is a single dichotomous exposure variable, which we have named "Conversion to roundabout". The two levels are simply "Non-exposed" (i.e., during the "before" period) and "Exposed" (i.e., during the "after" period).

DEFINITION OF MEASURE(S) OF EXPOSURE          BACK TO CORE INFO
                                             BACK TO RESULTS

| Name of Countermeasure | Specifications | Data type | Level 1 | Level 2 |
|---|---|---|---|---|
| 1. Conversion to roundabout | | Exposed/Non-exposed | Exposed | Non-exposed |

**$outcome**
The outcome is also quite simple: here we are dealing with crash counts.

| Name | Type | Specifications | Data type |
|------|------|----------------|-----------|
| 1. Crash count | Crash | | Count |

The final field in the "Design – Core info" section simply requires the coder to specify the total number of effects for which details will be provided in the "Results" sheet.

As can be seen in Table 8, there are 4 groups of results:
1. Single Lane, Urban, Stop Controlled
2. Single Lane, Rural, Stop Controlled
3. Multilane, Urban, Stop Controlled
4. Urban, Signalized

For each of the groups, there is (a) an overall result (i.e., all crashes), expressed as "Index of Effectiveness" and "Percent Reduction in Crashes" and (b) a result for injury crashes only. Since only one of the two numerical expressions is needed, there are 2x4=8 effects at the group level. In the bottom, Table 8 also includes across-groups effects (all crashes and injury crashes only). At this point the total number of effects would be 10. However, since the effect for injury crashes is not available for the group "Multilane, Urban, Stop Controlled", we eventually indicate 9 effects.

### 6.1.2    Results

This sheet contains a table that will automatically shape itself according to
   (a) the number of effects that is indicated (see "Core info" sheet),
   (b) the different design variables that are indicated (see "Core info", "Flexible info" and "Custom info" sheets),
   (c) the choice with respect to the "Direction" in the design (see "Core info" sheet)
   (c) the details of exposure variable(s) (see "$exposure" sheet) and
   (d) the details of the outcome(s) (see "$outcome" sheet).

> *It is essential to verify that the structure of the results table indeed corresponds to the structure of the effects reported in the study BEFORE filling in the table.*

| ☐ Differences between effects | Effect 1 | Effect 2 | Effect 5 | Effect 6 |
|---|---|---|---|---|
| Road network profile - Area | Urban road | Urban road | Urban road | Urban road |
| Road network profile - Segments | Intersection - Stop controlled | Intersection - Stop controlled | Intersection - Stop controlled | Intersection - Signalized |
| Accident severities | All | Injury | All | All |
| No. lanes per direction | 1 | 1 | >1 | 1; >1 |
| | | | | |
| Conversion to roundabout - Test group | Exposed | Exposed | Exposed | Exposed |
| Conversion to roundabout - Reference group | Non-exposed | Non-exposed | Non-exposed | Non-exposed |
| | | | | |
| Measure of effect/association | Crash modification factor | Crash modification factor | Crash modification factor | Crash modification factor |
| Specifications | | | | |
| | | | | |
| Estimate | 0.2800 | 0.1200 | 0.9500 | 0.6500 |
| Standard error of estimate | 0.0600 | 0.0800 | 0.1000 | 0.0900 |
| p-value | | | | |
| Confidence level | | | | |
| Lower limit | | | | |
| Upper limit | | | | |
| | | | | |
| Adjustment variables/Covariates | AADT | AADT | AADT | AADT |
| | | | | |
| Conclusion | Significant positive effect on | Non-significant effect on roa | Significant positive effect on | Significant positive effect on |
| | | | | |
| Comments | | | | |

The screenshot only shows the coding of effects 1, 2, 5 and 6.

> *It will often occur that specific values need to be repeated many times across the results table. In Excel, a single cell value can be pasted into several other cells at once simply by selecting those cells before giving "paste" command. AGAIN, PLEASE MAKE SURE ONLY TO PASTE VALUES (SEE ABOVE).*

## Design variables

The 9 effects that are reported in Table 8 of Persaud et al. can all be identified using the 4 design variables we have indicated in the sampling frame. Note that the group "Urban, Signalized" collapses data across single and multilane intersections. This is coded by joining the two levels with a semicolon: "1; 2+"

## Exposure/Outcome variables

After the effects are defined with respect to the design variables, the specific **contrast** with respect to exposure or outcome levels should be determined for each effect.

For **≠ Exposure -> ≠ Outcome** the contrast is defined with respect to one or more exposure variables. For any dichotomous/categorical exposure variable X that occurs in the "$exposure" sheet, the coder needs to choose the level (or combination of levels using semicolons) with respect to the "X – Test group" and "X – Reference group".

The contrasts concern a comparison of crash counts before and after exposure (conversion to roundabout). The measurements in the before period thus function as the "reference" for the "test" measurements in the after period. Accordingly, we choose "Exposed" for the "Conversion to roundabout – Test group" field and "Non-exposed" for the "Conversion to roundabout – Reference group".

## Measure of effect/association

Depending on the specific design at hand and the method for statistical analysis, there exist several ways to quantify the change in outcome values across exposures (≠ Exposure -> ≠ Outcome) or the difference in exposure across different outcomes (≠ Outcome -> ≠ Exposure).

As already noted, the group-level and across-groups effects are expressed in two ways: "Index of effectiveness" and "Percentage Reduction of Crashes". On page 7 we read:
*"Table 8 summarizes the estimated crash reductions and provides two measures of safety effects. The first is 'index of safety effectiveness' (θ), which is approximately equal to the ratio of the number of crashes occurring after conversion to the number expected had conversion not taken place. The second is the more conventional percent reduction in crashes, which is equal to 100(1−θ)."*
It is clear that the two expressions are transformations of each other. When we are dealing with such a situation we try to avoid redundancy and choose the "best" expression. In this case, we choose "Index of effectiveness" as it is in fact another name for "Crash modification factor", which is more common. Also note that information about the precision of estimates (standard errors) is also only provided for the first expression.

## Numerical and statistical details

*The next set of variables concern the actual numerical values.*
1. ***Estimate**: The value of the measure of effect/association.*
2. ***Standard error of the estimate**: If available.*
3. ***P-value**: This can either be an actual value (e.g. .0234) or a cut-off ("<.001"; ">.2")*
4. ***Confidence level**: If confidence limits are reported, please specify the level (e.g. .80, .95)*
5. ***Lower limit**: Smallest value of the confidence interval, if available*
6. ***Upper limit**: Largest value of the confidence interval, if available*

*For each effect we can only provide the estimate for the crash modification factor and the standard error.*

**Adjustment variables/Covariates**

*These fields allow to specify any variables that were put under numerical/statistical control. Please use semicolons to separate different variable names (e.g. "age; traffic volume")*
*The empirical Bayes method implies that traffic volume effects were controlled in the analyses.*

**Conclusions**

This variable is essential, as it codes the basic conclusion for each effect: *Significant risk factor, Non-significant risk factor, Significant improvement due to measure, Non-significant improvement due to measure.*
*Although Table 8 does not provide p-values we can derive the conclusions from the text.*

### 6.1.3 Summary

---

*The "Summary" sheet is a regular sheet where the coder should synthesize the study, USING PLAIN TEXT ONLY. It is most practical to prepare the synthesis in a separate document and paste the text (as plain text) in the "Summary" sheet*

---

The text should explain which risk factor or countermeasure was investigated, how the design looked like, how the results were analyzed and which conclusions were formulated. Coders should provide this information for all aspects of the study they are coding and should mention if there are other aspects they are not coding (or that are coded in another file because it concerns another risk factor / countermeasure). Summaries serve two purposes: (1) abstracts that are provided by authors do not always give all the appropriate information and (2) as in the reviewing process for scientific articles, an adequate summary shows that the coder identified all relevant aspects of a study.

*The study addresses the reduction of crashes at intersections after their conversion to roundabouts. The data cover 23 intersections in 7 U.S. states. Intersections were a mix of:*
- *converted from stop sign control or traffic signal control*
- *single lane or multilane designs*
- *located in an urban or rural environment*

*A before-after design was used comparing changes in crash counts with no-conversion model predictions ("Empirical Bayes").*
*Overall, the results show that converting intersections from stop sign or traffic light control to roundabouts can lead to significant reductions of crashes.*

## 6.2 "CASE-CONTROL" STUDY WITHOUT MATCHING

*M.R. Bambach, R.J. Mitchell, R.H. Grzebieta, J. Olivier (2013). The effectiveness of helmets in bicycle collisions with motor vehicles: A case-control study. Accident Analysis and Prevention, 53, pp. 78-88.*

### 6.2.1 Summary and topic

From an Australian crash database (New South Wales, 2001-2009) cyclists with hospitalised with head-injuries (cases) are compared to those either hospitalised with another injury or with only minor injuries (controls). Different severities and natures of head injuries are considered separately. The countermeasure investigated is helmets and cyclists who wore a helmet (exposed) were compared to those who did not. The odds for riders to have worn a helmet are consistently smaller in the head-injury groups as compared to the control group with other or minor injury. The effect is consistently larger for more severe injury categories.

| Coder | Name | Kevin Diependaele | | | |
|---|---|---|---|---|---|
| | Institution | BRSI | | | |
| | Date (dd/mm/yyyy) | 08/12/2015 | | | |

| Reference | Authors | M.R. Bambach, R.J. Mitchell, R.H. Grzebieta, J. Olivier | | | |
|---|---|---|---|---|---|
| | Title | The effectiveness of helmets in bicycle collisions with motor vehicles: A case-control study | | | |
| | Year | 2013 | | | |
| | Source | Accident Analysis and Prevention, 53, pp. 78-88 | | | |
| | URL | doi:10.1016/j.aap.2013.01.005 | | | |

| Topic | Risk factor or Countermeasure? | Countermeasure | | | |
|---|---|---|---|---|---|
| ☐ | WP | WP6 | | | |
| ☐ | Service | Passive Safety (VRU) | | | |
| ☐ | System | | | | |
| ☐ | | | | | |
| ☐ | | | | | |
| ☐ | | | | | |
| | Abstract | There has been an ongoing debate in Australia and internationally regarding the effectiveness of bicycle helmets in p | | | |
| | Keywords | bicycle helmets; head injury; case-control; bicycle collisions with motor vehicles | | | |

| Sampling frame | ☐ Countries | Australia | | | |
|---|---|---|---|---|---|
| | ☐ Administrative Level | Regional | | | |
| | ☐ Road user profile - Modes | Cyclist | | | |
| | ☐ Road user profile - Type | Rider | | | |
| | ☐ Road user profile - Subgroup | | | | |
| | ☐ Road user profile - Age | | | | |
| | ☐ Road user profile - Gender | All | | | |
| | ☐ Road network profile - Area | Rural road | Urban road | Suburban road | |
| | ☐ Road network profile - Segments | All | | | |
| | ☐ Accident severities | Injury | Fatal | | |
| | ☑ Injury severities | Hospital | AIS 3 | AIS 4+ | |
| | Comments | Region = New South Wales; Injury severities defined on the basis of Survival Risk Ratio (SRR); "Moderate": 0.965 < SRF | | | |

| Design | Features | Observational | Cross-sectional | Case-control | Unmatched |
|---|---|---|---|---|---|
| | Direction | ≠ Outcome -> ≠ Exposure | | | |
| | EXPOSURE DEFINITION | Cycling helmet | | | |
| | OUTCOME DEFINITION | Injury | | | |
| | Total number of effects | 18 | | | |
| | Comments | | | | |

| Limitations / Potential sources of bias | | Extent | Motivation | | |
|---|---|---|---|---|---|
| | | | | | |

**Figure 6.1** Example coding of a case-control study without matching, "core-info" sheet

Note that the coding of the taxonomy levels are probably not the final ones. The key words have been specified within the abstract by inserting asterisks before and after the words in question.

There has been an ongoing debate in Australia and internationally regarding the effectiveness of *bicycle helmets* in preventing *head injury*. This study aims to examine the effectiveness of bicycle helmets in preventing head injury amongst cyclists in crashes involving motor vehicles, and to assess the impact of 'risky cycling behaviour' among helmeted and unhelmeted cyclists. This analysis involved a retrospective, *case–control* study using linked police-reported road crash, hospital admission and mortality data in New South Wales (NSW), Australia during 2001–2009. The study population was cyclist casualties who were involved in a collision with a motor vehicle. Cases were those that sustained a head injury and were admitted to hospital. Controls were those admitted to hospital who did not sustain a head injury, or those not admitted to hospital. Standard multiple variable logistic regression modelling was conducted, with multinomial outcomes of injury severity. There were 6745 cyclist collisions with motor vehicles where helmet use was known. Helmet use was associated with reduced risk of head injury in *bicycle collisions with motor vehicles* of up to 74%, and the more severe the injury considered, the greater the reduction. This was also found to be true for particular head injuries such as skull fractures, intracranial injury and open head wounds. Around one half of children and adolescents less than 19 years were not wearing a helmet, an issue that needs to be addressed in light of the demonstrated effectiveness of helmets. Non-helmeted cyclists were more likely to display risky riding behaviour, however, were less likely to cycle in risky areas; the net result of which was that they were more likely to be involved in more severe crashes.

### 6.2.2 Sampling frame

Note that the coding of crash severity (injury & fatal) indicates the inclusion criteria for crashes to the database (and therefore for the analysis presented). *Injury severity is checked as a design variable* because it differentiates between the effects reported.

*Injury coding:* Two tables give the core of the results and are the ones that should potentially be coded. Table 6 gives the results for all types of head injuries jointly. It also gives the complete model that has been applied to estimate the Odds ratio's for helmet wear (presumably to the results in Table 4 as well).

**Table 6**
Odds ratios for associations with all head injury (compared with no head injury casualty controls), from multiple variable logistic regression analysis of cyclist casualties resulting from motor vehicle collisions, NSW 2001–2009 (model diagnostics are shown in Table 5).

| | Moderate vs possible minor injury | | | | Serious vs possible minor injury | | | | Severe vs possible minor injury | | | | Variable |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Odds | $CI_L$ | $CI_U$ | $p$ | Odds | $CI_L$ | $CI_U$ | $p$ | Odds | $CI_L$ | $CI_U$ | $p$ | Chi-square |
| Speed limit (km/h) 0–50 | 1 | | | | 1 | | | | 1 | | | | |
| 70–90 | 1.368 | 0.946 | 1.979 | 0.089 | 1.572 | 0.961 | 2.571 | 0.066 | 3.104 | 1.659 | 5.805 | <0.0001 | 16.89 |
| 100–110 | 1.945 | 1.138 | 3.325 | 0.013 | 2.370 | 1.219 | 4.609 | 0.010 | 2.450 | 0.872 | 6.885 | 0.083 | 12.50 |
| Collision with a car/van/utility | 1 | | | | 1 | | | | 1 | | | | |
| Bus/light/heavy truck | 1.064 | 0.766 | 1.477 | 0.706 | 1.866 | 1.279 | 2.723 | 0.001 | 1.803 | 1.026 | 3.169 | 0.037 | 14.14 |
| 4WD | 1.522 | 1.039 | 2.230 | 0.028 | 0.987 | 0.525 | 1.855 | 0.967 | 0.177 | 0.023 | 1.356 | 0.089 | 8.07 |
| Age group (years) 50+ | 1 | | | | 1 | | | | 1 | | | | |
| 0–12 | 1.016 | 0.642 | 1.608 | 0.946 | 3.651 | 1.995 | 6.684 | <0.0001 | 1.230 | 0.531 | 2.847 | 0.622 | 18.53 |
| 13–19 | 0.760 | 0.506 | 1.142 | 0.178 | 1.806 | 1.009 | 3.231 | 0.042 | 0.631 | 0.282 | 1.410 | 0.252 | 7.95 |
| 20–29 | 0.921 | 0.643 | 1.319 | 0.647 | 0.690 | 0.364 | 1.308 | 0.246 | 0.490 | 0.212 | 1.133 | 0.089 | 4.09 |
| 30–39 | 0.582 | 0.398 | 0.850 | 0.004 | 1.139 | 0.651 | 1.991 | 0.642 | 0.376 | 0.157 | 0.901 | 0.025 | 13.15 |
| 40–49 | 0.677 | 0.451 | 1.018 | 0.056 | 1.048 | 0.569 | 1.931 | 0.879 | 0.535 | 0.218 | 1.314 | 0.164 | 8.43 |
| Serious injury other than head | 6.390 | 4.872 | 8.381 | <0.0001 | 9.917 | 7.070 | 13.911 | <0.0001 | 13.263 | 8.048 | 21.859 | <0.0001 | 375.7 |
| Helmet | 0.506 | 0.388 | 0.659 | <0.0001 | 0.378 | 0.267 | 0.536 | <0.0001 | 0.257 | 0.148 | 0.448 | <0.0001 | 70.48 |
| Disobeyed a traffic control | 1.738 | 1.159 | 2.605 | 0.006 | 0.967 | 0.522 | 1.789 | 0.912 | 3.166 | 1.632 | 6.140 | 0.001 | 18.14 |
| BAC over 0.05 | 3.468 | 2.257 | 5.329 | <0.0001 | 4.086 | 2.289 | 7.294 | <0.0001 | 0.897 | 0.202 | 3.974 | 0.884 | 48.89 |
| Riding on the footpath | 0.510 | 0.361 | 0.721 | <0.0001 | 0.343 | 0.212 | 0.556 | <0.0001 | 1.014 | 0.569 | 1.808 | 0.962 | 32.77 |

Moderate (0.965 < SRR ≤ 1.0), serious (0.854 < SRR ≤ 0.965), severe (SRR ≤ 0.854).
$CI_L$, lower 95% confidence limit; $CI_U$, upper 95% confidence limit.

Table 4 gives the results per type of injury (skull fracture, intracranial injury, open wound) and for all injuries jointly (-> head injury, which are the same as reported more extensively in Table 6).

**Table 4**

Odds ratios for associations with head injury resulting from helmet use compared with non-helmet use, from multiple variable logistic regression analysis of cyclist casualties resulting from motor vehicle collisions, NSW 2001–2009.

| | Injury severity | No head injury casualty controls | | | Control 1 | No head injury hospitalisation controls | | | Control 2 |
|---|---|---|---|---|---|---|---|---|---|
| | | Odds[a] | $CI_L$ | $CI_U$ | $p$ | Odds[a] | $CI_L$ | $CI_U$ | $p$ |
| Head injury | Possible minor injury | 1 | | | | 1 | | | |
| | Moderate | 0.506 | 0.388 | 0.659 | <0.0001 | 0.577 | 0.431 | 0.773 | <0.0001 |
| | Serious | 0.378 | 0.267 | 0.536 | <0.0001 | 0.435 | 0.300 | 0.629 | <0.0001 |
| | Severe | 0.257 | 0.148 | 0.448 | <0.0001 | 0.285 | 0.163 | 0.500 | <0.0001 |
| Skull fracture | Possible minor injury | 1 | | | | 1 | | | |
| | Moderate | 0.437 | 0.130 | 1.466 | 0.171 | 0.466 | 0.137 | 1.584 | 0.212 |
| | Serious/severe | 0.217 | 0.132 | 0.357 | <0.0001 | 0.246 | 0.149 | 0.408 | <0.0001 |
| Intracranial injury | Possible minor injury | 1 | | | | 1 | | | |
| | Moderate | 0.626 | 0.408 | 0.961 | 0.029 | 0.718 | 0.461 | 1.117 | 0.134 |
| | Serious | 0.356 | 0.200 | 0.633 | <0.0001 | 0.374 | 0.207 | 0.677 | 0.001 |
| | Severe | 0.284 | 0.159 | 0.506 | <0.0001 | 0.302 | 0.169 | 0.542 | <0.0001 |
| Open wound | Possible minor injury | 1 | | | | 1 | | | |
| | Moderate/serious/severe | 0.200 | 0.122 | 0.330 | <0.0001 | 0.224 | 0.135 | 0.372 | <0.0001 |

[a] Odds ratios for wearing a helmet vs not; $CI_L$, lower 95% confidence limit; $CI_U$, upper 95% confidence limit.

Table 4 gives the main results but the authors have calculated additional models for which the effects of the helmet variable are given in Table 6. The result pattern for the different types of injuries (fracture, intracranial, open wound) reported separately and two types of control groups have been considered (all casualties vs. only hospitalised casualties). Given that none of the subgroups gives a very different pattern of result compared to the overall pattern for all head injuries together and that there is no substantial difference in the pattern of result with the two different control groups, we would ***normally*** suggest that it ***suffices to code the main results*** presented in Table 4. *The results for different injury types and those using all "hospitalised casualties" as controls are coded **for demonstration purposes** only.*

In the **flexible coding sheet**, consequently ***Injury nature*** was also checked as a design variable.

| | | | | |
|---|---|---|---|---|
| ☐ | *Accident types* | | | |
| ☐ | *Accident - Opponent (a)* | Motor vehicle (incl PTW) | | |
| ☐ | *Accident - Opponent b* | | | |
| ☐ | *Accident - Collisions* | | | |
| ☐ | *Accident - CDC: Directions of force* | | | |
| ☐ | *Accident - CDC: Areas of deformation* | | | |
| ☐ | *Accident - CDC: Types of damage distribution* | | | |
| | | | | |
| ☑ | *Injury nature* | Fracture | Internal | Open wound |
| ☐ | *Injury - Body region* | Head | | |
| ☐ | *Injury scale* | AIS | | |
| | | | | |
| ☐ | *Data collection - Start-Stop (dd/mm/yyyy-…)* | 2001-2009 | | |

### 6.2.3    Design

Although the authors claim that this is a case-control study, there is no matching between cases and controls, and actually this is simply a cross-sectional study observational study.

***Direction:*** As in all cross-sectional studies, the direction is somewhat arbitrary. However, because in the multiple logistic regression analyses the ***injury outcomes*** (and other variables) were used ***as predictors to explain the odds of the victim "being exposed"*** to a helmet, this study is coded as ***$outcome->$exposure study***.

### $Exposure

| DEFINITION OF MEASURE(S) OF EXPOSURE | | | BACK TO CORE INFO | |
|---|---|---|---|---|
| | | | BACK TO RESULTS | |
| Name of Countermeasure | Specifications | Data type | Level 1 | Level 2 |
| 1. Cycling helmet | | Exposed/Non-exposed | *Exposed* | *Non-exposed* |

### $Outcome

| Name | Type | Specifications | Data type | Level 1 | Level 2 | ... |
|------|------|----------------|-----------|---------|---------|-----|
| 1. Injury | Injury | | Categories | Hospital; Head | Non-Head; Minor head | Hospital; Non-Head |

The outcome variable "Injury" has 3 categories to be able to report effect estimates with different control groups. In Table 6, the control groups to which cyclists with head injuries (Hospital; Head -> Cases) were compared were either to hospitalised cyclists without head injuries (Hospital; Non-Head -> hospital controls) or to all injured cyclists who either had no head-injury or whose head-injury was so minor that they were not admitted to hospital (Non-Head; Minor Head -> casualty controls).

### *Limitations*

No limitations were noted. Generally this seems to be a good study.

- High detail of reporting - for data-sources, outcome variables (injuries) and their categorization, for control variables, and for the statistical analysis applied.
- Fair study design: Cyclists with head-injuries are compared to cyclists with other injuries or only minor injuries. Due to reporting problems uninjured cyclists could not be taken into account.
- N= 6745 -> large sample. Small sample size within design cells was avoided by joining several cells.
- Potential bias – fair: Possible differences between helmeted and non-helmeted riders (type of road, safety behaviour, age, alcohol, etc...) have been corrected for by including them (or related variables) as covariates into the regression analysis.
- Generalizability – fair: One Australian region tested. Study spans 9 years of crash data. Consistent results across different level of injury nature and injury severity. Covers only police reported collisions with motor vehicles but the authors convincingly argue that protection effects for collisions without motor vehicles are unlikely to be smaller.

## 6.2.4   Results

| ☐ Differences between effects | Effect 1 | Effect 2 | Effect 3 |
|------------------------------|----------|----------|----------|
| Injury severities | Hospital | AIS 3 | AIS 4+ |
| Injury nature | Fracture; Internal; Open Wou | Fracture; Internal; Open Wou | Fracture; Internal; Open Wou |
| | | | |
| Injury - Cases | Hospital; Head | Hospital; Head | Hospital; Head |
| Injury - Controls | Non-Head; Minor head | Non-Head; Minor head | Non-Head; Minor head |
| | | | |
| Measure of effect/association | Odds ratio | Odds ratio | Odds ratio |
| Specifications | Odds for wearing a helmet | Odds for wearing a helmet | Odds for wearing a helmet |
| | | | |
| Estimate | 0.5060 | 0.3790 | 0.2570 |
| Standard error of estimate | | | |
| Statistic [name(parameters)=x] | | | |
| p-value | <0.0001 | <0.0001 | <0.0001 |
| Sample size (x or n1=x1; n2=x2) | n (cyclist casualties)= 6745 | n (cyclist casualties)= 6745 | n (cyclist casualties)= 6745 |
| Confidence level | 0.9500 | 0.9500 | 0.9500 |
| Lower limit | 0.3880 | 0.2670 | 0.1480 |
| Upper limit | 0.6590 | 0.5360 | 0.4480 |
| | | | |
| Adjustment variables/Covariates | Speed limit; Collision vehicle | Speed limit; Collision vehicle | Speed limit; Collision vehicle |
| | | | |
| Conclusion | Significant positive effect on | Significant positive effect on | Significant positive effect on |
| | | | |
| Comments | | | |

The first 3 effects are those analysing all injury-types jointly (which is why all 3 types are selected under "injury nature" – separated by semicolon). The analysis concerns all cyclist casualties (n=6745) and the control consists thus of cyclists with non-head injuries and with minor head injuries (who were not hospitalized). The covariates are all variables listed in the analysis presented in Table 4.

| Effect 4 | Effect 5 | Effect 6 | Effect 7 | Effect 8 | Effect 9 |
|---|---|---|---|---|---|
| AIS 3 | AIS 3; AIS 4+ | Hospital | AIS 3 | AIS 4+ | Hospital; AIS 3; AIS 4+ |
| Fracture | Fracture | Internal | Internal | Internal | Open wound |
| | | | | | |
| Hospital; Head | Hospital; Head | Hospital; Head | Hospital; Head | Hospital; Head | Hospital; Head |
| Non-Head; Minor head | Non-Head; Minor head | Non-Head; Minor head | Non-Head; Minor head | Non-Head; Minor head | Non-Head; Minor head |
| | | | | | |
| Odds ratio | Odds ratio | Odds ratio | Odds ratio | Odds ratio | Odds ratio |
| Odds for wearing a helmet | Odds for wearing a helmet | Odds for wearing a helmet | Odds for wearing a helmet | Odds for wearing a helmet | Odds for wearing a helmet |
| | | | | | |
| 0.4370 | 0.2170 | 0.6260 | 0.3560 | 0.2840 | 0.2000 |
| | | | | | |
| | | | | | |
| 0.1710 | <0.0001 | 0.0290 | <0.0001 | <0.0001 | <0.0001 |
| n (cyclist casualties)= 6745 | n (cyclist casualties)= 6745 | n (cyclist casualties)= 6745 | n (cyclist casualties)= 6745 | n (cyclist casualties)= 6745 | n (cyclist casualties)= 6745 |
| 0.9500 | 0.9500 | 0.9500 | 0.9500 | 0.9500 | 0.9500 |
| 0.1300 | 0.1320 | 0.4080 | 0.2000 | 0.1490 | 0.1220 |
| 1.4660 | 0.3570 | 0.9610 | 0.6330 | 0.5060 | 0.3300 |
| | | | | | |
| Speed limit; Collision vehicle | Speed limit; Collision vehicle | Speed limit; Collision vehicle | Speed limit; Collision vehicle | Speed limit; Collision vehicle | Speed limit; Collision vehicle |
| | | | | | |
| Non-significant effect on roa | Significant positive effect on | Significant positive effect on | Significant positive effect on | Significant positive effect on | Significant positive effect on |

Effects 4-9 concern different types of injuries. Note that the categories that are analysed jointly are all listed, separated by semicolons.

| Effect 10 | Effect 11 | Effect 12 |
|---|---|---|
| Hospital | AIS 3 | AIS 4+ |
| Fracture; Internal; Open Wou | Fracture; Internal; Open Wou | Fracture; Internal; Open Wou |
| | | |
| Hospital; Head | Hospital; Head | Hospital; Head |
| Hospital; Non-Head | Hospital; Non-Head | Hospital; Non-Head |
| | | |
| Odds ratio | Odds ratio | Odds ratio |
| Odds for wearing a helmet | Odds for wearing a helmet | Odds for wearing a helmet |
| | | |
| 0.5770 | 0.4350 | 0.2850 |
| | | |
| | | |
| <0.0001 | <0.0001 | <0.0001 |
| n (cyclists hospitalized)=1859 | n (cyclists hospitalized)=1859 | n (cyclists hospitalized)=1859 |
| 0.9500 | 0.9500 | 0.9500 |
| 0.4310 | 0.3000 | 0.1630 |
| 0.7730 | 0.6290 | 0.5000 |
| | | |
| Speed limit; Collision vehicle | Speed limit; Collision vehicle | Speed limit; Collision vehicle |
| | | |
| Significant positive effect on | Significant positive effect on | Significant positive effect on |

Effects 10-12 show the main effects with all injury types, but with a different control group (only hospitalized cyclists without head injuries). Not that the sample size is much smaller for these effects.

## 6.3 EXPERIMENTAL STUDY

*Kureckova, V., Gabrhel, V., Zamecnik, P., Rezac, P., Zaoral, A., & Hobl, J. (2017). First aid as an important traffic safety factor – evaluation of the experience–based training, European Transport Research Review, 9:5.*

### 6.3.1 Summary and topic

This study compares the efficacy of two first-aid trainings for learning drivers in Czech Republic: the official and current one, lasting 4 hours, and a new conception of a 16-h experience-based first-aid course, designed by the authors and focused on knowledge, skills and psychological set-up. Thirty participants were randomly divided into two groups of 15 participants each. The first group went through the standard training; the second group went through the new experience-based training. Three levels of competencies were tested: 1. Knowledge; 2. Skills; 3. Performance in a simulated situation. A pilot test showed a remarkable difference between the two groups: compared to the 4-h standard trainings, the 16-h experience-based-training participants demonstrate more knowledge as well as better skills. They also perform better during simulated emergency situations and declare better comfort and fewer inadequate emotional reactions.

| Coder | Name | Charles Goldenbeld | | | |
|---|---|---|---|---|---|
| | Institution | SWOV | | | |
| | Date (dd/mm/yyyy) | 22/02/2017 | | | |
| | | | | | |
| Reference | Authors | Kureckova, V., Gabrhel, V., Zamecnik, P., Rezac, P., Zaoral, A., & Hobl, J. | | | |
| | Title | First aid as an important traffic safety factor – evaluation of the experience–based training | | | |
| | Year | 2017 | | | |
| | Source | European Transport Research Review, 9:5 | | | |
| | URL | http://link.springer.com/article/10.1007/s12544-016-0218-4 | | | |
| | | | | | |
| Topic | Risk factor or Countermeasure? | Countermeasure | | | |
| ☐ | WP | WP7 | | | |
| ☐ | Topic | First aid training drivers | | | |
| ☐ | Subtopic | First aid training drivers | | | |
| ☐ | | 0 | | | |
| ☐ | | | | | |
| | Abstract | Introduction: First aid is a factor that reduces damage to health and loss of life in traffic accidents. It is therefore nece | | | |
| | Keywords | First aid, Driving school, Training | | | |

References of this study are mentioned in the "Reference" box of the "Core info" sheet. There are essential for the end-users who want to return to the original study. The "Topic" box clearly indicates that the paper is talking about "First aid training drivers", a countermeasure which is part of the Work Package 7 "Serious road injuries, analysis and strategy", as it can reduce damage to health and loss of life in road crashes.

Here, the keywords are specified in the dedicated variable. Another option would have been to indicate it using asterisks in the abstract, as follows: *"Introduction: *First aid* is a factor that reduces damage to health and loss of life in road crashes. It is therefore necessary to make even the lay population ready to give at least basic first aid. Czech *driving schools* offer only 4-h first-aid *trainings* that do not provide the appropriate level of competencies. Our team has designed a new conception of a 16-h experience-based first-aid course and compared its efficacy with the standard 4-h training. […]"*

### 6.3.2    Sampling frame

| Sampling frame | ☐ Countries | Czech Republic |
|---|---|---|
| | ☐ Administrative Level | |
| | ☐ Road user profile - Modes | |
| | ☐ Road user profile - Type | Driver |
| | ☐ Road user profile - Subgroup | |
| | ☐ Road user profile - Age | |
| | ☐ Road user profile - Gender | All |
| | ☐ Road network profile - Area | |
| | ☐ Road network profile - Segments | |
| | ☐ Accident severities | |
| | ☐ Injury severities | |
| | Comments | |

Information about the *sampling* is quite limited in the paper, which only mentioned the number, the gender and the age of the participants: "*The participants in group 1 (standard training), comprising 13 male and 2 female, were aged 19 to 52 years, the average age being 34 years; the participants in group 2 (experience-based training), comprising 14 male and 1 female, were aged 20 to 52 years, the average age being 35 years.*" The coder could also have mentioned the age in the "Sampling frame" box, but this was not essential as this experiment does not focus on a specific age-group.

### 6.3.3    Design

| Design | Features | Experimental | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Direction | ≠ Exposure -> ≠ Outcome | | | | | | |
| | EXPOSURE DEFINITION | First aid  training | | | | | | |
| | OUTCOME DEFINITION | Knowledge: emergency num | Knowledge: proceedings sce | Knowledge: internal injury is | Knowledge: spinal injury issi | Knowledge: priority issues | Knowledge: when to take ou | Kno |
| | Total number of effects | 14 | | | | | | |
| | Comments | | | | | | | |

As the exposure is actively manipulated (participants are exposed to different trainings), we are dealing with an experimental study. In experimental studies investigators always start from a manipulation of exposure levels and monitor potential effects on outcome variables. Hence the coder chose the option "≠ Exposure -> ≠ Outcome". Here, at the end of the experiment, the researchers compared the knowledge, skills and performances of the two groups. They wanted to evaluate the impact of the type of first-aid training (exposure) on the knowledge, skills, performance in a simulated situation of the participants (outcome).

Information about "exposure definition" and "outcome definition" has to be filled in the "$exposure" and "$outcome" sheets. The two levels of the exposure are the two types of first-aid training. The impact on knowledge, skills and performances was measured through 14 indicators, as described in the "$outcome" sheet. These indicators correspond to the measured "effects". For each outcome variable a comparison between both groups is presented, amounting to a total number of 14 effects.

DEFINITION OF MEASURE(S) OF EXPOSURE                                    BACK TO CORE INFO
                                                                        BACK TO RESULTS

| Name of Countermeasure | Specifications | Data type | Level 1 | Level 2 |
|---|---|---|---|---|
| 1. First aid  training | | Categories | *experience based (16h)* | *standard* |

| | Name | Type | Specifications | Data type |
|---|---|---|---|---|
| 1. | Knowledge: emergency number | Questionnaire | Mean score (1 best, 5 worst) | General numerical |
| 2. | Knowledge: proceedings scene accident | Questionnaire | Mean score (1 best, 5 worst) | General numerical |
| 3. | Knowledge: internal injury issues | Questionnaire | Mean score (1 best, 5 worst) | General numerical |
| 4. | Knowledge: spinal injury issues | Questionnaire | Mean score (1 best, 5 worst) | General numerical |
| 5. | Knowledge: priority issues | Questionnaire | Mean score (1 best, 5 worst) | General numerical |
| 6. | Knowledge: when to take outside car? | Questionnaire | Mean score (1 best, 5 worst) | General numerical |
| 7. | Knowledge: average score | Questionnaire | Mean score (1 best, 5 worst) | General numerical |
| 8. | Skill: life function check | Other (Please provide name) | Exercise checking of the life functions (u | General numerical |
| 9. | Skill: CPR artificial breathing | Other (Please provide name) | Exercise CPR – artificial breathing ; the p | General numerical |
| 10. | Skill: CPR chest compression | Other (Please provide name) | Exercise CPR – chest compressions (appr | General numerical |
| 11. | Performance at simulation: active reacti | Other (Please provide name) | A simulated situation was arranged – a ti | General numerical |
| 12. | Performance at simulation: fast reaction | Other (Please provide name) | A simulated situation was arranged – a ti | General numerical |
| 13. | Performance at simulation: accuracy | Other (Please provide name) | A simulated situation was arranged – a ti | General numerical |
| 14. | Performance at simulation: risk awarene | Other (Please provide name) | A simulated situation was arranged – a ti | General numerical |

## 6.3.4 Results

The "Results" sheet is partially automatically filled based on information given in other sheets. For each of the 14 effects, the coders had to specify the test/reference groups, the measure of effects and all other statistical information. He also had to conclude on the direction of (possible) effect and its significance.

Information to fill in this sheet can be found in the results tables of the papers. The association between exposure and outcome is expressed in terms of a simple absolute difference in the outcome value between the test and reference conditions: for each outcome/effect, the coder calculated by itself (as mentioned in comments) the difference between the mean score of the group exposed to the "improved" first-aid training and the mean score of the group exposed to the standard training. He also reported the Mann-Whitney U and the p-value for each estimate.

| ☐ Differences between effects | Effect 1 | Effect 2 | Effect 3 | Effect 4 | Effect 5 | Effect 6 |
|---|---|---|---|---|---|---|
| OUTCOME DEFINITION | Knowledge: emergency num | Knowledge: proceedings sce | Knowledge: internal injury is | Knowledge: spinal injury issu | Knowledge: priority issues | Knowledge: when to take ou |
| | | | | | | |
| First aid training - Test group | experience based (16h) | experience based (16h) | experience based (16h) | experience based (16h) | experience based (16h) | experience based (16h) |
| First aid training - Reference group | standard | standard | standard | standard | standard | standard |
| | | | | | | |
| Measure of effect/association | Absolute difference | Absolute difference | Absolute difference | Absolute difference | Absolute difference | Absolute difference |
| Specifications | Difference in mean score | Difference in mean score | Difference in mean score | Difference in mean score | Difference in mean score | Difference in mean score |
| | | | | | | |
| Estimate | 1.3300 | 2.4000 | 3.9000 | 1.3300 | 1.3300 | 2.4000 |
| Standard error of estimate | | | | | | |
| Statistic [name(parameters)=x] | Mann-Whitney U=52.5 | Mann-Whitney U=10.5 | Mann-Whitney U=0.000 | Mann-Whitney U=0.000 | Mann-Whitney U=75 | Mann-Whitney U=0 |
| p-value | p < 0.01 | p < 0.001 | p < 0.001 | p < 0.001 | p < 0.05 | p < 0.001 |
| Sample size (x or n1=x1; n2=x2) | n1=n2=15 | n1=n2=15 | n1=n2=15 | n1=n2=15 | n1=n2=15 | n1=n2=15 |
| Confidence level | | | | | | |
| Lower limit | | | | | | |
| Upper limit | | | | | | |
| | | | | | | |
| Adjustment variables/Covariates | | | | | | |
| | | | | | | |
| Conclusion | Significant positive effect on | Significant positive effect on | Significant positive effect on | Significant positive effect on | Significant positive effect on | Significant positive effect on |
| | | | | | | |
| Comments | difference score calculated b | difference score calculated b | difference score calculated b | difference score calculated b | difference score calculated b | difference score calculated b |

| Effect 7 | Effect 8 | Effect 9 | Effect 10 | Effect 11 | Effect 12 | Effect 13 | Effect 14 |
|---|---|---|---|---|---|---|---|
| Knowledge: average score | Skill: life function check | Skill: CPR artificial breathing | Skill: CPR chest compression | Performance at simulation: a | Performance at simulation: f | Performance at simulation: a | Performance at simulation: r |
| | | | | | | | |
| experience based (16h) | experience based (16h) | experience based (16h) | experience based (16h) | experience based (16h) | experience based (16h) | experience based (16h) | experience based (16h) |
| standard | standard | standard | standard | standard | standard | standard | standard |
| | | | | | | | |
| Absolute difference | Absolute difference | Absolute difference | Absolute difference | Absolute difference | Absolute difference | Absolute difference | Absolute difference |
| Difference in mean score | Difference in mean score | Difference in mean score | Difference in mean score | Difference in mean score | Difference in mean score | Difference in mean score | Difference in mean score |
| | | | | | | | |
| 2.5900 | 2.4700 | 1.8700 | 1.1700 | 1.2000 | 0.8000 | 2.2700 | 3.2400 |
| | | | | | | | |
| not reported | Mann-Whitney U=121.5 | Mann-Whitney U=20.5 | Mann-Whitney U=20.5 | Mann-Whitney U=65.0 | Mann-Whitney U=82.5 | Mann-Whitney U=4.5 | Mann-Whitney U=0 |
| | p < 0.001 | p < 0.001 | p < 0.001 | p < 0.05 | p > 0.05 | p < 0.001 | p < 0.001 |
| | n1=n2=15 | n1=n2=15 | n1=n2=15 | n1=n2=15 | n1=n2=15 | n1=n2=15 | n1=n2=15 |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| Positive effect on road safety | Significant positive effect on | Significant positive effect on | Significant positive effect on | Significant positive effect on | Non-significant effect on roa | Significant positive effect on | Significant positive effect on |
| | | | | | | | |
| difference score calculated b | difference score calculated b | difference score calculated b | difference score calculated b | difference score calculated b | difference score calculated b | difference score calculated b | difference score calculated b |

**Table 3** Comparison of the level of knowledge (average ratings: 1 = the best; 5 = the worst)

| Type of Training | Emergency line number | Desired proceedings on the scene of traffic accident | Internal injury issues | Spinal injury issues | Priority issues – massive bleeding or resuscitation | When to take the casualty outside of the car? | Average knowledge rating |
|---|---|---|---|---|---|---|---|
| **Improved** | | | | | | | |
| Mean | 1 | 1.6 | 1.1 | 1 | 1 | 1.53 | 1.21 |
| Std. Deviation | 0 | 0.83 | 0.26 | 0 | 0 | 0.52 | 0.23 |
| Median | 1 | 1 | 1 | 1 | 1 | 2 | |
| **Standard** | | | | | | | |
| Mean | 2.33 | 4 | 5 | 2.33 | 2.33 | 3.93 | 3.8 |
| Std. Deviation | 1.45 | 1 | 0 | 1.95 | 1.95 | 0.96 | 0.62 |
| Median | 3 | 4 | 5 | 5 | 1 | 4 | |
| **Difference (n1 = n2 = 15)** | | | | | | | |
| Mann-Whitney U | 52.5** | 10.5*** | 0*** | 0*** | 75* | 0*** | |
| r | 0.59 | 0.79 | 0.97 | 0.97 | 0.44 | 0.87 | |

*p < .05; ** p < .01; ***p < .001

**Table 4** Comparison of the level of skills (average ratings: 1 = the best; 5 = the worst)

| Type of training | Life functions check – unconscious casualty treatment | CPR – artificial breathing | CPR – chest compression |
|---|---|---|---|
| **Improved** | | | |
| Mean | 1.2 | 1.6 | 1.1 |
| Std. Deviation | 0.41 | 0.74 | 0.26 |
| Median | 1 | 1 | 1 |
| **Standard** | | | |
| Mean | 3.67 | 3.47 | 2.27 |
| Std. Deviation | 0.72 | 1.2 | 0.96 |
| Median | 4 | 3 | 2 |
| **Difference (n1 = n2 = 15)** | | | |
| Mann-Whitney U | 121.5*** | 20.5*** | 20.5*** |
| r | 0.88 | 0.72 | 0.77 |

***p < .001

**Table 5** Comparison of the performance (average ratings: 1 = the best; 5 = the worst)

| Type of training | Reaction (active – passive) | Reaction (fast – slow) | Accuracy of performance | Safety risk awareness |
|---|---|---|---|---|
| **Improved** | | | | |
| Mean | 1.53 | 1.73 | 1.2 | 1.36 |
| Std. Deviation | 0.74 | 0.8 | 0.4 | 0.63 |
| Median | 1 | 2 | 1 | 1 |
| **Standard** | | | | |
| Mean | 2.73 | 2.53 | 3.47 | 4.6 |
| Std. Deviation | 1.62 | 1.55 | 1.1 | 0.51 |
| Median | 2 | 2 | 3 | 5 |
| **Difference (n1 = n2 = 15)** | | | | |
| Mann-Whitney U | 65.5* | 82,5 | 4.5*** | 0*** |
| r | 0.38 | 0.24 | 0.85 | 0.87 |

*p < .05; ***p < .001

12 out of the 14 reported effects have a "significant positive effect on road safety", meaning that knowledge/skills/performances of participants exposed to the improved training (test-group) are "better" than those of the participants exposed to the standard training (reference group). There was only one case in which the difference was not statistically significant: Reaction Fast-Slow (effect 12). As the significance level of the "Average knowledge rating" is not mentioned in the paper, the coder noted "Positive effect on road safety without statistical test" as a conclusion for this outcome (effect 7). This pilot testing proved that the experience-based first-aid training focused on knowledge and skills as well as the psychological set-up is an effective part of a driver's education.

## 6.4    QUASI-EXPERIMENTAL STUDY

*Nasvadi, G. E., Vavrik, J. (2007). Crash risk of older drivers after attending a mature driver education program. Accident Analysis and Prevention 39, 1073–1079*

### 6.4.1    Summary and topic

The primary aim of this study was to determine if the crash rate of aging drivers can be mitigated by post-license driver education. It focused on the 55-Alive/Mature driver course (British Columbia, Canada) which provides information on rules of the road hazard recognition and age-related changes that effect driving. It also covers information on reducing exposure to complex situations and planning for driving cessation. This study was conducted in three phases. Phase 1, which examined self-selection bias of seniors attending the driver education program, and Phase 2, which examined changes in crash rate after attending the program, were carried out through analysis of driving records before and after attending the course. In Phase 3, the use of selection, optimization, and compensation strategies by older male drivers who attended 55 Alive/Mature Driving was addressed through focus group interviews.

884 older drivers (male and female, age range 55-94) who had attended the 55 Alive/Mature Driving classroom based refresher course were compared to control drivers matched for age, gender, postal code region and for number of crashes over a 2 year period prior to training course date. Drivers in both the intervention and control groups were labelled as 'crash' or 'non-crashed' drivers. 'Crash drivers' were those who had been involved in crash following training /date of training (controls) where the older driver was considered at least 25% liable. Binary logistic regression was used to determine if there was a difference between the two groups and odds ratios were calculated.

Results presented and included in the factsheet concern the whole sample for all ages, but are also split into the age ranges 55-74 and 75-94 and also divided by gender for the same age categories. No statistically significant benefit was identified but training appeared to increase the at fault crash risk for males in the older age category (75-94) (Odds ratio 3.8; $\beta$ = 1.344, p=0.005). The study also examined the pre-training crash risk of participants attending the 55 Alive/Mature Driving course compared with controls matched for age, gender and postal code region and those attending the course had significantly more police attended crashes ($\chi^2$=23.634, p<0.001) and total number of crashes ($\chi^2$=9,310, p=0.010) than controls.

| Coder | Name | Rachel Talbot | | | |
|-------|------|---------------|--|--|--|
| | *Institution* | Loughborough University | | | |
| | *Date (dd/mm/yyyy)* | 21/03/2017 | | | |
| | | | | | |
| **Reference** | *Authors* | Nasvadi, G. E., Vavrik, J. | | | |
| | *Title* | Crash risk of older drivers after attending a mature driver education program | | | |
| | *Year* | 2007 | | | |
| | *Source* | Accident Analysis and Prevention 39, 1073–1079 | | | |
| | *URL* | http://dx.doi.org/10.1016/j.aap.2007.02.005 | | | |
| | | | | | |
| **Topic** | *Risk factor or Countermeasure?* | Countermeasure | | | |
| ☐ | *WP* | WP4 | | | |
| ☐ | *Topic* | Education and voluntary trainings/programs | | | |
| ☐ | *Subtopic* | Elderly | | | |
| ☐ | *Specific countermeasure* | Driving | | | |
| ☐ | | | | | |
| | *Abstract* | Purpose: The primary aim of this study was to determine if the crash rate of aging drivers can be mitigated by post-li | | | |
| | *Keywords* | Self-selection bias; Driving self-regulation; Self-awareness; Selection; Optimization; Compensation, elderly, drivers | | | |

A taxonomy has been defined for each Work Package of the Safety Cube project. As this study tries to evaluate the impact of a post-license driver education dedicated to elderly, it belongs to the following categories in the taxonomy: Countermeasure - WP4 "Road user behaviour analysis" – Educa-

tion and voluntary trainings/programs – Elderly – Driving. The keywords proposed by the coder provide additional information to the taxonomy.

## 6.4.2 Sampling frame

| Sampling frame | ☐ Countries | Canada | | |
|---|---|---|---|---|
| | ☐ Administrative Level | Local | | |
| | | | | |
| | ☐ Road user profile - Modes | Car | | |
| | ☐ Road user profile - Type | Driver | | |
| | ☐ Road user profile - Subgroup | Elderly | | |
| | ☑ Road user profile - Age | 55-74 | 75-94 | 55-94 |
| | ☑ Road user profile - Gender | Female | Male | All |
| | | | | |
| | ☐ Road network profile - Area | | | |
| | ☐ Road network profile - Segments | | | |
| | | | | |
| | ☐ Accident severities | | | |
| | ☐ Injury severities | | | |
| | | | | |
| | Comments | A sample of both male and female older drivers (n=884) who had took part in the 55 Aliv | | |

As the name suggests, the box "Sampling frame" provides all information related to sampling. This study took place in Canada, at a local level, in British Colombia exactly. It focused on elderly car drivers, aged 55-94. As results will be disaggregated per age group and gender, these variables are defined as "design variables" and the little boxes in front of "Road user profile – Age" and "Road user profile – Gender" are checked.

In the comments, the coders specifies that "A sample of both male and female older drivers (n=884) who had took part in the 55 Alive/Mature classroom based driving programme. A control was identified for each of these participants matched for age, gender, postal code region and for number of crashes over a 2 year period prior to training course date."

The time period examined varied between participants as data was available until 31st December 2003 and training was conducted between January 2000 and July 2003. This information in mentioned by the coder in the "Flexible info" sheet: 01/01/2000-21/12/2003.

| ☐ Data collection - Start-Stop (dd/mm/yyyy-...) | 01/01/2000-21/12/2003 |
|---|---|

## 6.4.3 Design

For the Phase 1, a retrospective cohort design was used to compare the crash rates of older drivers who attended 55 Alive/Mature Driving to a matched control group of those who did not attend the educational program. Results of this phase show that elderly drivers who attended the 55 Alive/Mature Driving refresher program were more likely to have been involved in at-fault collisions prior to the course.

To accurately measure the impact of the program, it was therefore necessary to compare the 55 Alive/Mature Driving group with drivers who had similar crash rates but who were not exposed to the intervention. A matched pre-post-comparison design was used. For each subject and matched control, driving records were extracted for the same period of time before and after the date of attendance at the course. The data were subsequently dichotomized to "crashed" versus "non-crashed" drivers (following the course). Binary logistic regression was used to determine if there was a difference between the two groups, and odds ratios were calculated.

| Design | Features | | Quasi-experimental | Cohort | Matched |
|---|---|---|---|---|---|
| | Direction | | ≠ Exposure -> ≠ Outcome | | |
| | EXPOSURE DEFINITION | | Classroom based training pro | | |
| | OUTCOME DEFINITION | | Crash caused by older driver | | |
| | Total number of effects | | 9 | | |
| | Comments | | | | |

So, as mentioned by the authors, the study used a matched-pairs cohort design correcting for self-selection bias. This is therefore a ***quasi-experimental design***. Although the researchers had to rely on a sample of elderly drivers who had chosen themselves to take part in the program (and had been shown to be more crash prone than the usual population of elderly drivers in the first part of the paper), the researchers corrected this by matching them one by one to randomly selected other drivers with the same age, gender, and pre-training crash record.

The results listed in the coding sheet by the coder are those of Phase 2. Hence, the "direction of the design" is "≠ Exposure -> ≠ Outcome", with having followed (or not) the training program as the exposure and being in involved (or not) in a crash following the course as the outcome (where the older driver was considered at least 25% liable). Information about "exposure definition" and "outcome definition" have been filled in the "$exposure" and "$outcome" sheets, as follow.

*DEFINITION OF MEASURE(S) OF EXPOSURE*                    BACK TO CORE INFO
                                                         BACK TO RESULTS

| Name of Countermeasure | Specifications | Data type | Level 1 | Level 2 |
|---|---|---|---|---|
| 1. Classroom based training program | Taking part in the 55-Alive/Mature drive | Exposed/Non-exposed | *Exposed* | *Non-exposed* |

*DEFINITION OF ROAD SAFETY OUTCOME VARIABLE(S)*           BACK TO CORE INFO
                                                         BACK TO RESULTS

| Name | Type | Specifications | Data type | Level 1 | Level 2 |
|---|---|---|---|---|---|
| 1. Crash caused by older driver | Crash | Involved in crash following training (Inte | Yes/No | *Yes* | *No* |

### 6.4.4   Results

As the impact of the training program has been measured for three different age-groups and for each gender within these age-groups, 9 effects in total have to be coded. Each time, elderly "exposed" to the training program are the test group. The measure of effect/association is clearly mentioned in the paper as odds ratio coming from a binary logistic regression. All information required to fill the "Results" sheet can be found in the table 4 of the paper. Although this is not strictly speaking a statistical correction, it is chosen to make it clear that drivers were matched on age, gender, and pre-training crash record by entering this into the row for "adjustment variables/Covariates".

| ☐ Differences between effects | Effect 1 | Effect 2 | Effect 3 | Effect 4 |
|---|---|---|---|---|
| Road user profile - Age | 55-94 | 55-74 | 75-94 | 55-94 |
| Road user profile - Gender | All | All | All | Female |
| | | | | |
| Classroom based training program - Test group | Exposed | Exposed | Exposed | Exposed |
| Classroom based training program - Reference group | Non-exposed | Non-exposed | Non-exposed | Non-exposed |
| | | | | |
| Measure of effect/association | Odds ratio | Odds ratio | Odds ratio | Odds ratio |
| Specifications | Binary logistic regression, cra | Binary logistic regression, cra | Binary logistic regression, cra | Binary logistic regression, cra |
| | | | | |
| Estimate | 1.1500 | 0.8100 | 1.5300 | 1.0300 |
| Standard error of estimate | | | | |
| Statistic [name(parameters)=x] | $\beta = 0.141$ | $\beta = -0.213$ | $\beta = 0.425$ | $\beta = 0.025$ |
| p-value | 0.4270 | 0.4250 | 0.0780 | 0.9100 |
| Sample size (x or n1=x1; n2=x2) | Training and crash = 74/884; c | Training and crash = 28; contr | Training and crash = 46; contr | Training and crash = 43; contr |
| Confidence level | | | | |
| Lower limit | | | | |
| Upper limit | | | | |
| | | | | |
| Adjustment variables/Covariates | age; gender; pre-training cra | age; gender; pre-training cra | age; gender; pre-training cra | age; gender; pre-training cra |
| | | | | |
| Conclusion | Non-significant effect on roa | Non-significant effect on roa | Non-significant effect on roa | Non-significant effect on roa |
| | | | | |
| Comments | | | | Author's report that this is marginally significant |

| Effect 5 | Effect 6 | Effect 7 | Effect 8 | Effect 9 |
|---|---|---|---|---|
| 55-74 | 75-94 | 55-94 | 55-74 | 75-94 |
| Female | Female | Male | Male | Male |
| | | | | |
| Exposed | Exposed | Exposed | Exposed | Exposed |
| Non-exposed | Non-exposed | Non-exposed | Non-exposed | Non-exposed |
| | | | | |
| Odds ratio | Odds ratio | Odds ratio | Odds ratio | Odds ratio |
| Binary logistic regression, cra | Binary logistic regression, cra | Binary logistic regression, cra | Binary logistic regression, cra | Binary logistic regression, cra |
| 1.0600 | 1.0000 | 1.3900 | 0.5500 | 3.8000 |
| $\beta = 0.062$ | $\beta = 0.000$ | $\beta = 0.326$ | $\beta = -0.596$ | $\beta = 1.344$ |
| 0.8600 | 1.0000 | 0.2570 | 0.1570 | 0.0050 |
| Training and crash = 18; contr | Training and crash = 25; contr | Training and crash = 31; contr | Training and crash = 10; contr | Training and crash = 21; contr |
| age; gender; pre-training cra | age; gender; pre-training cra | age; gender; pre-training cra | age; gender; pre-training cra | age; gender; pre-training cras |
| Non-significant effect on roa | Non-significant effect on roa | Non-significant effect on roa | Non-significant effect on roa | Significant negative effect or |

Table 4
Post-course crash involvement of subjects and controls

| Age group | Involved in a crash | | $\beta$ | $p$ | OR |
|---|---|---|---|---|---|
| | Subjects | Controls | | | |
| Whole sample | | | | | |
| All ages | 74 | 65 | 0.141 | .427 | 1.15 |
| 55–74 years | 28 | 34 | −0.213 | .425 | 0.81 |
| 75–94 years | 46 | 31 | 0.425 | .078 | 1.53 |
| Women | | | | | |
| All ages | 43 | 42 | 0.025 | .910 | 1.03 |
| 55–74 years | 18 | 17 | 0.062 | .860 | 1.06 |
| 75–94 years | 25 | 25 | 0.000 | 1.000 | 1.00 |
| Men | | | | | |
| All ages | 31 | 23 | 0.326 | .257 | 1.39 |
| 55–74 years | 10 | 17 | −0.596 | .157 | 0.55 |
| 75–94 years | 21 | 6 | 1.344 | .005 | 3.80 |

Findings showed no significant positive effects of the training. For men aged 75 and older, the program is even associated with an *increased* number of crashes , which is all the more notable, as the men taking part in the training already had a higher than normal crash record. But even when compared to a group with a similarly heightened crash record before the training, they still had an even higher crash record after following the training. The focus group sessions (Phase 3) suggested older men who attended the program used fewer strategies to cope with their declining skills.

## 6.5    ACCIDENT PREDICTION MODELS

*Caliendo C., De Guglielmo M.L., Guida M. (2013). A crash prediction model for road tunnels. Accident Analysis and Prevention 55, 107– 115.*

### 6.5.1    Summary and topic

This paper seeks to quantify, for the first time for Italian motorway tunnels, the effects on the expected number of crashes of the following variables: tunnel length, traffic flow, percentage of trucks, number of lanes and sidewalks, with a view to suggesting countermeasures for improving tunnel safety. Both non-severe and severe crashes were jointly investigated in order to propose a prediction model. For this purpose, a 4-year monitoring period extending from 2006 to 2009 was considered. The database consisted of 260 tunnels with unidirectional traffic and having two or three lanes. During the monitored period, crash data and traffic flow were collated. A Bivariate Negative Binomial distribution was used to model the random variation of the number of crashes. It has shown that the number of both non-severe and severe crashes occurring in tunnels increases with the tunnel length, the annual average daily traffic per lane, the percentage of trucks and the number of lanes. In contrast, the sidewalk variable was not found to be statistically significant.

| Coder | Name | Akis Theofilatos | | |
| --- | --- | --- | --- | --- |
| | Institution | National Technical University of Athens | | |
| | Date (dd/mm/yyyy) | 13/06/2016 | | |
| | | | | |
| **Reference** | Authors | Caliendo C., De Guglielmo M.L., Guida M. | | |
| | Title | A crash-prediction model for road tunnels | | |
| | Year | 2013 | | |
| | Source | Accident Analysis and Prevention 55, 107– 115 | | |
| | URL | http://dx.doi.org/10.1016/j.aap.2013.02.024 | | |
| | | | | |
| **Topic** | *Risk factor or Countermeasure?* | Risk factor | | |
| ☐ | *WP* | WP5 | | |
| ☑ | *Header 5 - Infrastructure element* | Alignment - Road segments | Cross-section - Road segments | |
| ☑ | *Header 6 - Risk factor* | Horizontal/vertical alignmen | Lanes / ramps deficiencies | |
| ☑ | *Header 7 - Specific risk factor* | tunnel | number of lanes | |
| ☐ | | | | |
| ☐ | | | | |
| | Abstract | Considerable research has been carried out into open roads to establish relationships b | | |
| | Keywords | Road tunnels; Crash-prediction model; Non-severe and severe crashes; Bivariate Negat | | |

In the Work Package 5 "Infrastructure safety analysis", only two of the items analysed in this paper are listed in the taxonomy: tunnel infrastructure as such and the number of lanes. Even if those both items concern a characteristic of a road segment, they belong to different headers in the taxonomy, as mentioned by the coder. As the results of the analyses distinguish between the effect of the length of the tunnel and the effect of the number of lanes on crashes, these variables/headers have been checked as design variables. Hence, they will automatically appear on the first line of the "Results" sheet.

Here, the coder decided not to use the asterisks in the abstract to indicate the keywords, but provided a very complete list of keywords in the cell specially provided for this purpose: *Road tunnels; Crash prediction model; Non-severe and severe crashes; Bivariate Negative Binomial regression; Random Effects Binomial regression; Negative Multinomial regression; Traffic flow; Trucks; Length; Number of lanes.*

### 6.5.2    Sampling frame

The scope of this study is pretty clear: it goes over non-severe and severe crashes in Italian unidirectional motorway tunnels with two or three lanes during the period 2006-2009. All this information was encoded in the "sampling frame" (in the "Core info" sheet, but also in the "Flexible info" sheet)

which is designed to specify the scope of the study. In the paper, the authors explained that the term "severe crashes" includes injury and fatal crashes. Therefore it can be deducted that "non-severe crashes" corresponds to damage-only crashes. That's what is mentioned for the "Accident severities" variable.

| Sampling frame | ☐ Countries | Italy | | |
|---|---|---|---|---|
| | ☐ Administrative Level | National | | |
| | ☐ Road user profile - Modes | All | | |
| | ☐ Road user profile - Type | All | | |
| | ☐ Road user profile - Subgroup | | | |
| | ☐ Road user profile - Age | | | |
| | ☐ Road user profile - Gender | All | | |
| | ☐ Road network profile - Area | Motorway | | |
| | ☐ Road network profile - Segments | | | |
| | ☐ Accident severities | Damage only | Injury | Fatal |
| | ☐ Injury severities | | | |
| | Comments | Only 2- and 3- lane tunnels in motorways were considered. | | |

| Sampling frame | ☐ Traffic volume (AADT) | | |
|---|---|---|---|
| | ☐ Relative share of road users (Mode : %) | | |
| | ☐ No. lanes per direction | two-lanes | three-lanes |
| | ☐ Speed limit (km/h) | | |
| | ☐ Data collection - Start-Stop (dd/mm/yyyy-...) | 2006-2009 | |

Also note that the sampling frame does not provide any variable to indicate that the analysis relates to tunnels. The "Custom info" sheet is intended to add this type of missing information, as done here by the coder.

| | New/Existing variable name | Level names - Use square brackets [...] for levels that are new, but are NOT present in the study | | | |
|---|---|---|---|---|---|
| | ☐ Location | Tunnel | Level 2 | [Level 3] | [Level 4] |

### 6.5.3   Design

In the study, researchers analysed the number of crashes in tunnels according to different characteristics of these tunnels. They considered simultaneously the accidentally (the outcome) and the infrastructural characteristics of the tunnels, at a specific moment, in order to identify a possible association between these two elements. This is therefore a cross-sectional design. As different models have been tested by the researchers and are presented in the paper, the coder could (already) have specified in the "Comments" that effects reported in this coding sheet are those of the Bivariate Negative Binomial regression model (the best-fitting model). He decided to mention it in the "Results" sheet.

| Design | Features | Observational | Cross-sectional | |
|---|---|---|---|---|
| | Direction | ≠ Exposure -> ≠ Outcome | | |
| | EXPOSURE DEFINITION | Length of tunnel | Number of lanes | |
| | OUTCOME DEFINITION | Number of severe accidents | Number of non-severe accid | |
| | Total number of effects | 4 | | |
| | Comments | | | |

Detailed information about exposure and outcome variables are provided in the "$exposure" and "$outcome" sheets, as follow. As it can be seen on the Table 3 (from the paper) below, 7 different exposure items were included in the model, but the coder decided to only report in the coding sheet

items listed in the taxonomy, that is to say the (log of) length of the tunnel in kilometres and the number of lanes (two or three). There are two outcome variables, the number of severe crashes and the number of non-severe crashes. In total, 4 effects will thus be reported in the "Results" sheet.

*DEFINITION OF MEASURE(S) OF EXPOSURE*

| | Name of Risk factor | Specifications | Data type | Level 1 | Level 2 |
|---|---|---|---|---|---|
| 1. | Length of tunnel | in km | General numerical | | |
| 2. | Number of lanes | | Categories | *two-lanes* | *three-lanes* |

*DEFINITION OF ROAD SAFETY OUTCOME VARIABLE(S)*

| | Name | Type | Specifications | Data type |
|---|---|---|---|---|
| 1. | Number of severe accidents | Crash | | Count |
| 2. | Number of non-severe accidents | Crash | | Count |

### 6.5.4 Results

The coder reported here the information mentioned in the Table 3 of the paper.

In multivariable accident prediction models, as it is the case here, the typical estimator of effect is a regression coefficient. For continuous variables a coefficient indicates the slope (i.e. the change of the dependent variable for one unit of the independent variable). That's what's written in the coding sheet for the exposure variable "length of the tunnel". For discrete variables, such as here the number of lanes (here only two options, 2 or 3 lanes), this coefficient indicates the difference between test and reference condition.

As explained above, the coder specified here that the researchers applied a Bivariate Negative Binomial regression model. For the estimate and the standard error, he simply copied the values from Table 3. He also reported the value of the Likelihood Ratio test (LRT statistic, third column in Table 3) for each effect. The sample size corresponds to the number of crashes included in the analysis, i.e. 765 severe crashes and 1539 non-severe crashes.

Knowing that the model includes other variables than those included in the coding sheet, the coder mentions them as covariates: annual average daily traffic (AADT) (3 different indicators), percentage of trucks, presence of a sidewalk number and, according to the described effect, the number of lanes or the length of the tunnel.

| ☐ Differences between effects | Effect 1 | Effect 2 | Effect 3 | Effect 4 |
|---|---|---|---|---|
| Header 5 - Infrastructure element | Alignment - Road segments | Cross-section - Road segmen | Alignment - Road segments | Cross-section - Road segmen |
| Header 6 - Risk factor | Horizontal/vertical alignmen | Lanes / ramps deficiencies | Horizontal/vertical alignmen | Lanes / ramps deficiencies |
| Header 7 - Specific risk factor | tunnel | number of lanes | tunnel | number of lanes |
| OUTCOME DEFINITION | Number of non-severe accid | Number of non-severe accid | Number of severe accidents | Number of severe accidents |
| | | | | |
| Length of tunnel - Unit | in km | | in km | |
| | | | | |
| Number of lanes - Test group | | three-lanes | | three-lanes |
| Number of lanes - Reference group | | two-lanes | | two-lanes |
| | | | | |
| Measure of effect/association | Slope | Absolute difference | Slope | Absolute difference |
| Specifications | Bivariate negative binomial r | Bivariate negative binomial r | Bivariate negative binomial r | Bivariate negative binomial r |
| | | | | |
| Estimate | 0.6575 | 0.6975 | 0.4756 | 0.2996 |
| Standard error of estimate | 0.0715 | 0.0841 | 0.0947 | 0.1127 |
| Statistic [name(parameters)=x] | Likelihood Ratio test (tunnel | Likelihood Ratio test (numbe | Likelihood Ratio test (tunnel | Likelihood Ratio test (numbe |
| p-value | | | | |
| Sample size (x or n1=x1; n2=x2) | 1539 | 1539 | 765 | 765 |
| Confidence level | 0.9000 | 0.9000 | 0.9000 | 0.9000 |
| Lower limit | | | | |
| Upper limit | | | | |
| | | | | |
| Adjustment variables/Covariates | AADT; percentage of trucks; s | AADT; percentage of trucks; s | AADT; percentage of trucks; s | AADT; percentage of trucks; s |
| | | | | |
| Conclusion | Significant negative effect or | Significant negative effect or | Significant negative effect or | Non-significant effect on roa |
| | | | | |
| Comments | AADT could not be coded because it was entered as a function. The estimates of the best-fitting models were entere | | | |

**Table 3**
Parameters of bivariate and univariate negative binomial regression models for non-severe and severe accidents.

| | Variables | Bivariate NB | | |
|---|---|---|---|---|
| | | Point estimate | Standard error | LRT statistic |
| Non-severe accidents | Constant | 3.64063 | 0.03226 | 77.21 |
| | Log of length in km | 0.65747 | 0.07146 | 53.66 |
| | Log of AADT per lane/10,000 | 1.45384 | 0.05707 | 87.99 |
| | D1 (1 if AADT per lane < 5000, 0 otherwise) | 0.30106 | 0.07840 | 3.27 |
| | D2 (1 if AADT per lane > 13,000, 0 otherwise) | −0.17063 | 0.09833 | 1.31 |
| | Log of percentage of trucks/100 | 1.04928 | 0.02057 | 20.24 |
| | Number of lanes (0 two lanes, 1 three lanes) | 0.69753 | 0.08413 | 29.20 |
| | Sidewalk (1 if present, 0 if absent) | 0.06850 | 0.04766 | 0.39 |
| | Overdispersion parameter | – | | |
| | Log-likelihood | – | | |
| Severe accidents | Constant | 2.59513 | 0.04060 | 25.08 |
| | Log of length in km | 0.47560 | 0.09466 | 18.06 |
| | Log of AADT per lane/10,000 | 2.08129 | 0.08492 | 112.38 |
| | D1 (1 if AADT per lane < 5000, 0 otherwise) | 0.61642 | 0.11962 | 7.20 |
| | D2 (1 if AADT per lane > 13,000, 0 otherwise) | −0.63012 | 0.11231 | 12.85 |
| | Log of percentage of trucks/100 | 0.69250 | 0.02551 | 5.42 |
| | Number of lanes (0 two lanes, 1 three lanes) | 0.29956 | 0.11267 | 3.28 |
| | Sidewalk (1 if present, 0 if absent) | −0.07900 | 0.06551 | 0.32 |
| | Overdispersion parameter | 8.08584 | 1.48524 | |
| | Log-likelihood | −1068.789 | | |

The results showed that the number of both non-severe and severe crashes occurring in tunnels over the 4-year monitoring period increases with (among others) tunnel length and number of lanes. It can therefore be said that these both factors have a "Significant negative effect on road safety". To explain these results, researchers mention the following hypotheses: more crashes may be expected in longer tunnels due to the drivers' diminishing concentration with increasing length and more crashes may be expected in three-lanes tunnels (compared to two-lanes ones) given that an increase in the number of lanes increases the opportunities for lane change.

## 6.6 CROSS SECTIONAL STUDY

*Kattan, L., Tay, R., & Acharjee, S. (2011). Managing speed at school and playground zones. Accident Analysis and Prevention, 43, 887-1891.*

### 6.6.1 Summary and topic

The purpose of this paper is to examine the effectiveness of school and playground zones (and their characteristics) in reducing traffic speed. The analysis covers a sample of 11 schools and 16 playgrounds randomly located in the City of Calgary in Alberta, Canada, where the speed is limited to 30 km/h. The free flow speed of 4580 was recorded and four indicators were calculated: the mean speed, the $85^{th}$ percentile speed, the proportion of vehicles that were driven over the speed limit of 30 km/h and the proportion of vehicles that were driven at 10 km/h or more over the speed limit. In addition to the traffic speed, several site characteristics were recorded to examine their impact on traffic speed: school or playground, children present or absent, 2 lane or 4 lane road, with or without fencing, with or without speed display, local or collector road, length of zone, distance from the road and controlled or uncontrolled intersection. A series of simple ANOVA tests were performed to determine if these speed measures were affected by the characteristics of the sites. In addition, a multivariable analysis using a linear regression model was conducted to determine the effects of different site characteristics on vehicle speed.

| Coder | Name | Charles Goldenbeld |
|---|---|---|
| | Institution | SWOV |
| | Date (dd/mm/yyyy) | 03/04/2017 |

| Reference | Authors | Kattan, L., Tay, R., & Acharjee, S. |
|---|---|---|
| | Title | Managing speed at school and playground zones |
| | Year | 2011 |
| | Source | Accident Analysis and Prevention, 43, 887-1891 |
| | URL | http://www.sciencedirect.com/science/article/pii/S000145751100090X |

| Topic | Risk factor or Countermeasure? | Countermeasure | |
|---|---|---|---|
| ☐ | WP | WP5 | |
| ☐ | Header 5 - Infrastructure element | Infrastructure safety management | |
| ☐ | Header 6 - Countermeasure | Speed management & enforcement | |
| ☐ | Header 7 - Specific countermeasure | school zones speed reduction measures | |
| ☐ | | | |
| | Abstract | Since speeding is one of the major causes of frequent and severe traffic accidents arour | |
| | Keywords | Road safety, School zone, Playground zone, Speed | |

The position of this study in the SafetyCube's taxonomy is rather clear: it falls under the Work Package 5 devoted to infrastructure, particularly the field of speed management in school area.

### 6.6.2 Sampling frame

In this case, information that can be provided in the Sampling frame (whether in the "Core info" sheet or in the "Flexible info" sheet) is very limited. It can only be mentioned that the study was conducted in Canada, at a local level. It does no concern any particular road user type, road network profile or crash severity. On the "flexible info" sheet, one may mention the speed limit of 30 km/h and the weather conditions, as the paper specifies that measurements were taken under dry conditions.

| Sampling frame | ☐ Countries | Canada | |
| | ☐ Administrative Level | Local | |
| | | | |
| | ☐ Road user profile - Modes | | |
| | ☐ Road user profile - Type | | |
| | ☐ Road user profile - Subgroup | | |
| | ☐ Road user profile - Age | | |
| | ☐ Road user profile - Gender | | |
| | | | |
| | ☐ Road network profile - Area | | |
| | ☐ Road network profile - Segments | | |
| | | | |
| | ☐ Accident severities | | |
| | ☐ Injury severities | | |
| | | | |
| | Comments | | |

| Sampling frame | ☐ Traffic volume (AADT) | |
| | ☐ Relative share of road users (Mode : %) | |
| | | |
| | ☐ No. lanes per direction | |
| | ☐ Speed limit (km/h) | 30 |
| | ☐ Speed limit (miles/h) | |

| | ☐ Time of the day (hh:mm) | |
| | ☐ Day of the week | |
| | ☐ Weather conditions | Dry/clear |
| | ☐ Light conditions | |

### 6.6.3    Design

In this study, researchers looked at the impact of a certain infrastructural countermeasures (speed limit in school zones/playgrounds).They observe the speed of vehicles in areas dependent on the absence/presence of school zones and playgrounds, taking also into account other characteristics.

They wanted to know if the characteristics of the area have an impact on the speed of the drivers. They look simultaneously at the exposure and the outcome, in order to find a potential association between these elements. This is thus a cross-sectional analysis.

| Design | Features | Cross-sectional | | | | |
| | | | | | | |
| | Direction | ≠ Exposure -> ≠ Outcome | | | | |
| | EXPOSURE DEFINITION | Type zone | Presence children | Number lanes | Presence fencing | Presence speed mo |
| | OUTCOME DEFINITION | Mean speed (km/h) | | | | |
| | | | | | | |
| | Total number of effects | 18 | | | | |
| | | | | | | |
| | Comments | | | | | |

As explained above, researches collected the speed of the vehicles and calculated four indicators based on these measures. Results for all these four indicators are not detailed in the paper, reason why the mean speed is the only indicator mentioned as outcome variable in the coding sheet.

| DEFINITION OF ROAD SAFETY OUTCOME VARIABLE(S) | | | BACK TO CORE INFO |
| | | | BACK TO RESULTS |
| Name | Type | Specifications | Data type |
| 1. Mean speed (km/h) | Other (Please provide name) | | General numerical |

In total, the impact of nine characteristics of the studied zones is analysed in the study. These are listed in the "$exposure" sheet. All these are categorical variables, with two or three possible values.

*DEFINITION OF MEASURE(S) OF EXPOSURE*

| | Name of Countermeasure | Specifications | Data type | Level 1 | Level 2 | ... |
|---|---|---|---|---|---|---|
| 1. | Type zone | | Categories | Playground | Schoolzone | ... |
| 2. | Presence children | | Categories | No | Yes | ... |
| 3. | Number lanes | | Categories | 2 lane road | 4 lane road | ... |
| 4. | Presence fencing | | Categories | No | Yes | ... |
| 5. | Presence speed monitoring device | | Categories | No | Yes | ... |
| 6. | Road classification | | Categories | Local | Collector | ... |
| 7. | Length of zone | | Categories | < 200 meter | 200-300 meter | > 300 meter |
| 8. | Distance from road | | Categories | Abut road | <=50 m | >50 m |
| 9. | Traffic control | | Categories | No | Yes | ... |

## 6.6.4 Results

In order to evaluate the association between the outcome and the exposure variables, researchers conducted two types of analysis. First, they simply ran an ANOVA test for each of the exposure. The results of the tests are reported in the Table 1 of the paper and in the first nine columns of the "Results" sheet of the coding Excel document. The measure of association is an absolute difference between the mean speed of the test condition (for example when children are present in the area) and the mean speed of the reference condition (when there is no child). Differences across each site characteristics are statistically significant at α=0.05. In total, eleven effects are described in the paper, but only nine were reported in the coding sheet. All these have a significant positive impact on road safety, in the sense that mean speed is lower in the test condition than in the reference condition. The main limit of this approach is that potential other confounding factors are absolutely not taken into account.

| ☐ Differences between effects | Effect 1 | Effect 2 | Effect 8 | Effect 9 |
|---|---|---|---|---|
| Type zone - Test group | Schoolzone | | | |
| Type zone - Reference group | Playground | | | |
| Presence children - Test group | | Yes | | |
| Presence children - Reference group | | No | | |
| Number lanes - Test group | | | | |
| Number lanes - Reference group | | | | |
| Presence fencing - Test group | | | | |
| Presence fencing - Reference group | | | | |
| Presence speed monitoring device - Test group | | | | |
| Presence speed monitoring device - Reference group | | | | |
| Road classification - Test group | | | | |
| Road classification - Reference group | | | | |
| Length of zone - Test group | | | | |
| Length of zone - Reference group | | | | |
| Distance from road - Test group | | | >50 m | |
| Distance from road - Reference group | | | Abut road | |
| Traffic control - Test group | | | | Yes |
| Traffic control - Reference group | | | | No |
| | | | | |
| Measure of effect/association | Absolute difference | Absolute difference | Absolute difference | Absolute difference |
| Specifications | | | | |
| | | | | |
| Estimate | -0.6000 | -0.7000 | -1.7000 | -3.1000 |
| Standard error of estimate | | | | |
| Statistic [name(parameters)=x] | | | | |
| p-value | $p < 0.05$ | $p < 0.05$ | $p < 0.05$ | $p < 0.05$ |
| Sample size (x or n1=x1; n2=x2) | N=4580 (n test=1230; n ref=3... | N=4580 (n test=3380; n ref=1... | N=4580 (n test=2400; n ref=1... | N=4580 (n test=840; n ref=374... |
| Confidence level | | | | |
| Lower limit | | | | |
| Upper limit | | | | |
| | | | | |
| Adjustment variables/Covariates | | | | |
| | | | | |
| Conclusion | Significant positive effect on | Significant positive effect on | Significant positive effect on | Significant positive effect on |
| | | | | |
| Comments | | | | |

**Table 1**
Speed and non-compliance by site characteristics.

| Site characteristics | Number of sub-samples | Number of observations | Simple non-compliance (%) | High-end non-Compliance (%) | Mean speed (km/h) | 85th percentile speed (km/h) |
|---|---|---|---|---|---|---|
| Total sample | 41 | 4580 | 54.4 | 10.0 | 32.0 | 38.8 |
| **Type of zones** | | | | | | |
| Playground | 30 | 3350 | 55.9 | 10.7 | 32.0 | 39.1 |
| School zone | 11 | 1230 | 51.4 | 8.5 | 31.4 | 38.0 |
| **Presence of children** | | | | | | |
| Yes ($N=25$; $n=3380$) | 25 | 3380 | 52.4 | 8.8 | 31.5 | 38.0 |
| No ($N=16$; $n=1200$) | 16 | 1200 | 55.6 | 10.7 | 32.2 | 39.3 |
| **Number of lanes** | | | | | | |
| 2 lane road | 35 | 3960 | 53.3 | 9.3 | 31.7 | 38.4 |
| 4 lane road | 6 | 620 | 58.7 | 12.7 | 33.1 | 40.2 |
| **Presence of fencing** | | | | | | |
| No | 8 | 800 | 64.8 | 13.0 | 33.3 | 40.5 |
| Yes | 38 | 3780 | 51.3 | 9.1 | 31.6 | 38.3 |
| **Presence of speed monitoring device** | | | | | | |
| No | 22 | 2400 | 64.3 | 10.2 | 33.1 | 38.9 |
| Yes | 19 | 2180 | 38.7 | 6.3 | 30.0 | 36.4 |
| **Road classification** | | | | | | |
| Local | 14 | 1550 | 46.3 | 8.3 | 30.8 | 37.4 |
| Collector | 27 | 3030 | 59.3 | 11.0 | 32.7 | 39.5 |
| **Length of zone** | | | | | | |
| Length < 200 m | 21 | 2340 | 56.3 | 11.1 | 32.2 | 39.4 |
| Length 200–300 m | 13 | 1440 | 51.9 | 9.4 | 31.6 | 38.4 |
| Length > 300 m | 7 | 800 | 52.7 | 7.6 | 31.7 | 37.6 |
| **Distance from road** | | | | | | |
| Abut road | 10 | 1060 | 51.9 | 8.6 | 31.4 | 37.9 |
| ≤50 m | 10 | 1120 | 70.3 | 14.7 | 34.4 | 41.9 |
| >50 m | 21 | 2400 | 55.7 | 13.4 | 33.1 | 40.5 |
| **Traffic control (lights, stop sign, etc.)** | | | | | | |
| No | 33 | 3740 | 67.2 | 14.9 | 34.2 | 41.6 |
| Yes | 8 | 840 | 47.8 | 8.0 | 31.1 | 37.6 |

*Note*: Differences in each of the speed measures across each site characteristics are statistically significant at $\alpha=0.05$.

In addition to this series of single-variable analyses, a multivariable analysis using a linear regression model was conducted to take into account simultaneously the effects of different site characteristics on vehicle speed. The estimates and p-value of the model are detailed in the Table 2 of the paper and in the columns 10 to 18 of the coding sheet. Here again, the coder only reported nine of the eleven effects, as he decided to only report one (of the two) effect for the length of the zone and the distance from road. The typical estimator of effects in multivariable models is the regression coefficient, what is correctly mentioned in the coding sheet. In this case, this is the slope (which is equal to the absolute difference between the test and the reference condition). Finally, all other (exposure) variables included in the regression are rightly cited as "Adjustment variables / Covariates". All estimated coefficients are statistically significant (p-value < 0.05) and all tested conditions have a significant positive impact on road safety, in the sense that mean speed is lower in such conditions than in the reference conditions.

| Effect 10 | Effect 11 | Effect 12 | Effect 13 | Effect 14 |
|---|---|---|---|---|
| Schoolzone | | | | |
| Playground | | | | |
| | Yes | | | |
| | No | | | |
| | | 2 lane road | | |
| | | 4 lane road | | |
| | | | Yes | |
| | | | No | |
| | | | | Yes |
| | | | | No |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| Slope | Slope | Slope | Slope | Slope |
| regression coefficient in line | regression coefficient in line | regression coefficient in line | regression coefficient in line | regression coefficient in line |
| -1.9600 | -1.4700 | -1.0100 | -1.0300 | -1.3100 |
| p < 0.0001 | p < 0.0001 | p = 0.0030 | p < 0.0001 | p < 0.0001 |
| N=4580 (n test=1230; n ref=3: | N=4580 (n test=3380; n ref120 | N=4580 (n test=3960; n ref=6; | N=4580 (n test=800; n ref=37; | N=4580 (n test=2400; n ref=21 |
| Variables in the regression w | Variables in the regression w | Variables in the regression w | Variables in the regression w | Variables in the regression w |
| Significant positive effect on | Significant positive effect on | Significant positive effect on | Significant positive effect on | Significant positive effect on |

| Effect 15 | Effect 16 | Effect 17 | Effect 18 |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| Local | | | |
| Collector | | | |
| | 200-300 meter | | |
| | > 300 meter | | |
| | | Abut road | |
| | | >50 m | |
| | | | Yes |
| | | | No |
| Slope | Slope | Slope | Slope |
| regression coefficient in line | regression coefficient in line | regression coefficient in line | regression coefficient in line |
| -1.6800 | -2.4400 | -0.8800 | -1.2100 |
| p < 0.0001 | p < 0.0001 | p = 0.0436 | p < 0.0001 |
| N=4580 (n test=1550; n ref=3( | N=4580 (n test=1440; n ref=8( | N=4580(n test=1060; n ref=24 | N=4580 (n test=840; n ref=374 |
| Variables in the regression w | Variables in the regression w | Variables in the regression w | Variables in the regression w |
| Significant positive effect on | Significant positive effect on | Significant positive effect on | Significant positive effect on |

**Table 2**
Estimation results using linear regression.

| Variables | Coefficient | P-value |
|---|---|---|
| Intercept | 36.70 | <0.0001 |
| Type of zone (ref: playground) | | |
| School zone | −1.96 | <0.0001 |
| Presence of children (ref: none) | | |
| Children present | −1.47 | <0.0001 |
| Number of lanes (ref: 4 lanes) | | |
| 2 lane road | −1.01 | 0.0030 |
| Presence of fencing (ref: none) | | |
| Fence | −1.03 | <0.0001 |
| Presence of speed monitoring device (ref: none) | | |
| Speed monitoring device | −1.31 | <0.0001 |
| Road classification (ref: collector) | | |
| Local road | −1.68 | <0.0001 |
| Length of zone (ref: over 300 m) | | |
| Length < 200 m | −0.45 | 0.0349 |
| Length 200–300 m | −2.44 | <0.0001 |
| Distance from road (ref: over 50 m) | | |
| Abut road | −0.88 | 0.0436 |
| Within 50 m | 1.41 | 0.0153 |
| Traffic control (ref: none) | | |
| Controlled intersection | −1.21 | <0.0001 |

*Note*: No of observations: 4580; *R*-square: 0.0749; *F*-statistics: 33.62; *P*-value: <0.0001.

## 6.7 META-ANALYSIS

As an example for coding a meta-analysis, the study "Safety-in-numbers: A systematic review and meta-analysis of evidence" by Elvik and Bjørnskau (2015) is chosen.

### 6.7.1 Background

Safety-in-numbers denotes a tendency for the number of crashes involving a certain group of road users to increase less than proportionally to the increase in traffic volume. Safety-in-numbers has been discussed in particular for crashes involving motor vehicles and pedestrians or cyclists. A recent meta-analysis (Elvik and Bjørnskau 2015) synthesised the findings of studies that have evaluated the existence and magnitude of a safety-in-numbers effect.

Most of these studies were multivariable accident prediction models of the form:

1

$$\text{Number of crashes} = e^{\beta_0} MV^{\beta_1} CYCL^{\beta_2} e^{\left(\sum_{n}^{i=1} \beta_n X_n\right)}$$

Where e denotes the exponential function, i.e. the base of the natural logarithms (2.71828) raised to the power of a regression coefficient $\beta$. The first term is the constant term. The next two terms refer to traffic volume. MV denotes motor vehicles, CYCL denotes cyclists (PED for pedestrians in models including pedestrian volume). Traffic volume typically enters models in the form of average daily traffic (AADT). The final term ($e^{(\Sigma \beta_n X_n)}$) is a set of predictor variables (X) other than traffic volume, which may influence the number of crashes. The final term was not included in all models. Some models employed traffic volumes as the only independent variables.

The outcome variables in the models are regression coefficients. A majority of the studies that were retrieved stated the standard errors of the regression coefficients and could therefore be included in an ordinary inverse variance meta-analysis. A random effects model was adopted and summary estimates of regression coefficients developed for motor vehicles, cyclists and pedestrians.

### 6.7.2 Coding

There are many choices to be made when coding this study. The first is about which countries to code. It was decided to code the countries from which studies were included in the meta-analysis.

The countries listed were (alphabetically):
Canada, Denmark, Netherlands, New Zealand, Norway, Sweden, United Kingdom, United States. Studies have also been reported in Australia and Belgium. These studies were not included in the meta-analysis and were therefore not coded.

The road network profile elements were difficult to use. A comment was therefore written that the studies were made at three different levels: micro, meso and macro.

Study design features were coded as: *Observational*, *Meta-analysis (random effects)*, *Cross-sectional*.

All studies included in the meta-analysis were cross-sectional. Although the paper provides results for several meta-models (see Table 4 of the paper), it was decided to code only the results of the random effects meta-analysis, since it represents the most appropriate choice in the current case.

The study was coded as exposure -> effect. The definition of *exposure* is traffic volume for three road user groups: cars, cyclists and pedestrians. The three exposure variables are expressed in annual average daily traffic (AADT) and, hence, are coded as "General numerical".

| Name of Risk factor | Specifications | Data type |
|---|---|---|
| 1. Motor vehicle volume | unit: annual average daily traffic (AADT) | General numerical |
| 2. Cyclist volume | unit: annual average daily traffic (AADT) | General numerical |
| 3. Pedestrian volume | unit: annual average daily traffic (AADT) | General numerical |

The road safety *outcome* consists of crash counts. As mentioned in the last paragraph of the introduction: "*Studies that use the number of injury crashes involving both a motor vehicle and a cyclist or pedestrian as dependent variable were treated as relevant.*". This information is coded as follows:

**3.** On the Core info sheet we indicate that the sampling frame involves injury crashes.

| ☐ *Accident severities* | Injury | | |
|---|---|---|---|

**4.** On the Flexible info sheet we use the two opponent fields to specify that crashes always involved a motor vehicle, on the one hand, and a pedestrian or cyclist, on the other hand.

| ☐ *Accident - Opponent (a)* | Motor vehicle (incl PTW) | | |
|---|---|---|---|
| ☐ *Accident - Opponent b* | Pedestrian | Cyclist | |

The next decision concerns how many results to code from the meta-analysis. Table 4 of the paper contains a total of 44 summary estimates. It was decided to code only three of these. These were the random-effects summary estimates of the regression coefficients for the three exposure variables: motor vehicle, cyclist and pedestrian volumes. These summary estimates were chosen because they are based on the most appropriate model of meta-analyses (random effects) and on studies for which the statistical weights were known.

For each exposure variable the unit is specified as "annual average daily traffic (AADT)".

The estimates can be interpreted as slopes if the general model underlying the individual studies (see formula 1 in the paper) is reformulated on natural logarithmic scale:

2

$$E(\ln(\boldsymbol{y})) = \ln(\boldsymbol{V})\,\boldsymbol{\zeta} + \boldsymbol{X\beta}$$

Where **y** is the vector of observed crash counts, the vector $\boldsymbol{\zeta}$ contains the estimated coefficients for the log-transformed traffic volumes $\boldsymbol{V}$ (matrix with different volumes cars/pedestrians/cyclists in the columns) and $\boldsymbol{X\beta}$ the linear model for predictor variables other than log-traffic volumes (including an intercept).

This information is coded in simplified form in the "Specifications" field (i.e., E(ln(y)) = Intercept + Estimate * ln(AADT) + Xb).

| ☐ Differences between effects | Effect 1 | Effect 2 | Effect 3 |
|---|---|---|---|
| Motor vehicle volume - Unit | annual average daily traffic (AADT) | | |
| Cyclist volume - Unit | | annual average daily traffic (AADT) | |
| Pedestrian volume - Unit | | | annual average daily traffic (A |
| | | | |
| Measure of effect/association | Slope | Slope | Slope |
| Specifications | E(ln(y)) = Intercept + Estimate | E(ln(y)) = Intercept + Estimate | E(ln(y)) = Intercept + Estimate |
| Estimate | 0.4990 | 0.4320 | 0.5110 |
| Standard error of estimate | | | |
| p-value | | | |
| Confidence level | 0.9500 | 0.9500 | 0.9500 |
| Lower limit | 0.3830 | 0.3330 | 0.3950 |
| Upper limit | 0.6150 | 0.5300 | 0.6270 |
| Adjustment variables/Covariates | | | |
| Conclusion | | | |
| Comments | | | |

### 6.7.3 Checking for sources of bias

The next stage of coding is to code potential sources of bias. As far as meta-analyses are concerned, a distinction must be made between sources of bias at two levels:

5. The primary studies that form the basis of the meta-analysis
6. The meta-analysis itself

In most meta-analyses, it is not possible to make changes in the primary studies or re-analyse them. In the paper, it is noted that none of the primary studies control adequately for all potentially confounding variables. In particular, no study has controlled fully for the quality of the infrastructure for cyclists and pedestrians. The code for confounding was therefore checked on the coding template.

With respect to the meta-analysis itself, each of the relevant points is discussed below.

### Systematic literature search

Relevant studies were identified by electronic searches in two databases using "safety-in-numbers" as search term (paper, section 2, page 1). A reviewer of the paper pointed out that the term safety-in-numbers only started to be used recently and may not identify older studies. Four older studies were identified by the ancestry method, i.e. by examining the list of references in studies that were found in the electronically searchable databases.
**Assessment:** A systematic literature search was performed (no bias).

### Listing of included and excluded studies

A total of 26 studies were identified. The studies are listed in Table 1 of the paper. 8 studies were not included in the meta-analysis. For each of these studies, the reason for their exclusion is stated.
**Assessment:** Study inclusion criteria have been stated and excluded studies listed with the reason for their exclusion (no bias).

### Variables coded for each study

For each study included in the meta-analysis, the following independent variables were coded (in addition to bibliographic information needed to identify the study): (1) Publication year; (2) Country; (3) Level of study units (micro, meso, macro); (4) Number of locations included; (5) Total number of crashes; (6) Type of crash model (four types); (7) Number of covariates (in addition to traffic volume) included in crash model. The following outcome variables were coded: (A) Coefficient for motor

vehicle volume and its standard error; (B) Coefficient for cyclist volume and its standard error; (C) Coefficient for pedestrian volume and its standard error.

All crashes are injury crashes. However, no distinction was made between different levels of crash severity. No data were available on the characteristics of cyclists or pedestrians. Very little information was available on the quality of the infrastructure for cyclists and pedestrians.
**Assessment:** There was incomplete data on some potentially relevant moderator variables (possible bias).

### Estimators of effect

All estimators of effect are identical, i.e. regression coefficients of negative binomial regression models. All models had the same mathematical form, but did not include exactly the same covariates.
**Assessment:** The estimators of effect are comparable (no bias).

### Exploratory analysis

Exploratory analysis was performed with respect to three problems: The comparability of the regression coefficients, the distribution of these coefficients, and the presence of outlying data points. Regression coefficients were found to be stable with respect to different model specifications; hence, although not all models were perfectly identical in all respects, the regression coefficients were judged to be sufficiently comparable for a meta-analysis to make sense. Outlying data points were identified both for motor vehicle volume, cyclist volume and pedestrian volume. It was decided to include the data points in the main analysis. The funnel plot for pedestrian volume was bimodal. As no reason could be found for this, all data points were nevertheless included in the analysis.
**Assessment:** The distributions of the regression coefficients had some minor anomalies, but these were not regarded as serious enough to invalidate an analysis (no bias).

### Main analysis

The main analysis was based on a random-effects model. This was appropriate given the heterogeneity of estimates. A subgroup analysis was made with respect to the level of the study units (micro, meso, macro). This analysis indicated that the summary coefficients were virtually identical for all levels of study units. Meta-regression was not attempted because of the limited number of estimates.
**Assessment:** No sources of bias can be identified in the main analysis.

### Sensitivity analysis

A sensitivity analysis was performed with respect to the possible presence of publication bias and number of confounding factors controlled for in primary studies. Evidence of publication bias was found for the coefficients referring to cyclist volume, but this evidence hinged on a single outlying data point. When that single data point was omitted, there was no evidence of publication bias.

Study quality was considered by comparing the value of coefficients estimated in models controlling for a different number of potentially confounding variables. There was no clear relationship between the number of confounding factors controlled for by a study and the estimated values of the coefficients. It was therefore concluded that safety-in-numbers is not likely to the result of confounding factors not controlled for.
**Assessment:** The sensitivity analysis confirmed that the results of the main analysis are robust (no bias).

# 7 Summarising Studies – the Synopses

## 7.1 SCOPE OF THE SYNOPSES

The SafetyCube Synopses aim to **summarise the existing effects of risk factors or measures**, either through a meta-analysis or, if a meta-analysis is not possible, through another type of comprehensive synthesis of existing results (e.g. vote-count table), in a way so that results can be easily identifiable and understood by the DSS users, and that the scientific background of the analysis is provided to those interested.

On the one hand, therefore, the main scope of the synopsis is to make a synthesis of studies results, and to complement and not replicate other DSS outputs (e.g. lists of studies, Tables comparing studies etc.). On the other hand, the transparency in the methods used within SafetyCube will certainly enhance the credibility of the DSS and will also will also serve as a guide to future DSS users for performing similar analyses with the DSS output. Therefore, the synopses needs to **balance the needs of different end users** (e.g. decision-makers looking for one global estimate vs. scientific users interested in more detailed aspects) **with the main scope and functionalities of the DSS,** as a dynamic clearinghouse of risks and measures effects.

## 7.2 SITUATING THE SYNOPSES IN THE DSS

On the basis of the DSS design principles, the main output of the DSS will be lists of available studies and effects, per risk factor (problem) or measure (solution), on the basis of the user search criteria (all available studies/effects per topic, or a subset concerning a specific road user group, road type, countries etc.). **The Synopses are therefore one additional type of "output" accessible through the DSS,** including the SafetyCube meta-analysis or other type of summary.

When a synopsis relevant to the search criteria used is available, it will be presented as **the top-level output in the list**, prioritised over existing meta-analyses, and over other existing studies.

An indicative example of a DSS output list for a topic is presented in the following Table 7.1. This corresponds to a user search through the taxonomy lists with search terms road user -> distraction / in-vehicle -> distraction due to mobile phone.

Table 7.1 Indicative example of DSS output list for distraction due to mobile phone

| Risk Factor | Source | Outcome variable(s) | Study type | Effect Type | Effect Size | Country |
|---|---|---|---|---|---|---|
| **Distraction / in-vehicle** | SafetyCube synopsis, 2016 (link to pdf.) | Various | Meta-analysis | | Significant effect on outcomes | |
| **Distraction due to mobile phone** | Caird et al., 2012 (link to URL) | Reaction time | Meta-analysis | | Significant effect on outcomes | |
| **Distraction due to mobile phone** | Yannis et al., 2016 (link to URL) | Speed, Crash risk | Original analysis | Odds ratio | Significant effect on outcomes | Greece |

| ... | ... | ... | ... | ... | ... | ... |
|-----|-----|-----|-----|-----|-----|-----|

## 7.3    SITUATING THE SYNOPSES IN THE TAXONOMY

The taxonomy of risk factors and measures has a hierarchical structure. As a consequence, one can imagine them as a tree in which the ends of each branch form the "leaves". As an example, Figure 7.1 shows a part of the taxonomy for the infrastructure work-package. The leaves for the branch "environment" are for instance "darkness", "fog" (under subtopic "visibility & lighting") and "rain", "snow/ice", and "wind" (under subtopic "Weather"). For the branch "Junctions" only some sub-categories and leaves are shown, and of course there are many more branches...
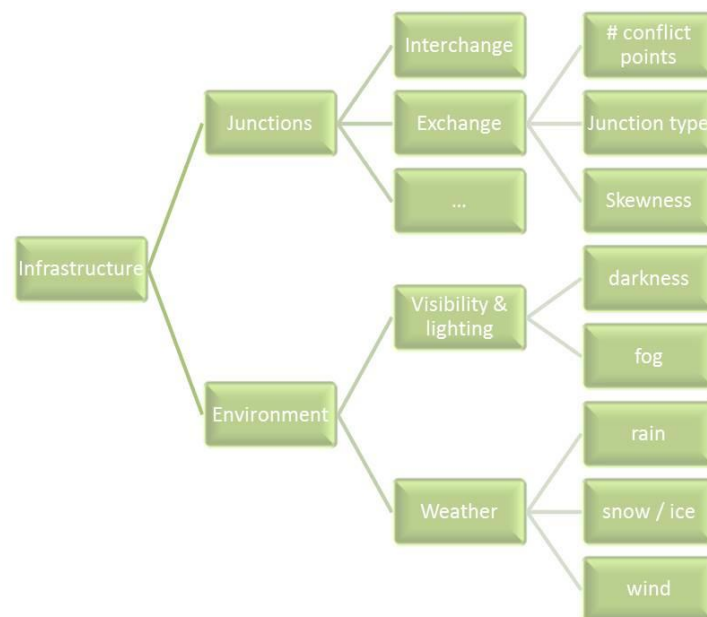


**Figure 7.1** Example of braches, subcategories, and leaves within taxonomy for infrastructure risk factors

**Ideally, each leaf of the DSS should include a synopsis** of the coded studies. It is noted, however, that **a synopsis will not always be available at the level of each specific topic (leaf)**, for either one or more of the following reasons: not meaningful to make a synopsis due to very detailed sub-topic (e.g. drugs - cannabis), synopsis not possible for a topic due to lack of studies (e.g. vehicle systems), or large differences in which the topic is defined and studied (e.g. road readability - self-explaining roads).

Consequently, **each Work Package (4-5-6) will define the taxonomy level at which synopses are possible / meaningful**, on a topic (leaf) specific basis.

## 7.4    OUTLINE OF THE SYNOPSES

The structure of the Synopses is proposed as follows, with **three distinct parts**:
a.  **Summary**: A two-page document including the abstract of the topic and an overview of effects and analysis methods by condition X indicator X risk type.
b.  **Scientific overview**: A 4-5 page document including a short synthesis of the literature, an overview of the available studies, a description of the analysis methods and an analysis of the effects by condition X indicator X risk type.
c.  **Supporting document** (no a priori page limit) describing the literature search, comparing the available studies in detail (optional) etc.

*When writing the synopsis the author would in most cases start with the Supporting Document (part c). For practical purposes, we will start these guidelines from the Supporting document. This Supporting document will serve as a basis for the scientific overview (b) and the eventual summary (a).*

The Supporting document (no a priori page limit) **aims to provide the most detailed information,** which may not be of interest for all users and which may not be presentable in a concise and user-friendly form for the summary or the scientific overview. It will be the "Appendix" of the summary and the scientific overview; it may start from the literature search, proceed to the analyses comparing the available studies and effects in detail etc., and finally include any details of the analysis carried out to summarise the results.

The Supporting document may include, in general, the following sections (which also correspond to the intuitive steps of analysing and summarising the topic):
- Methodology
  - Literature search strategy
  - Detailed analysis of study designs and methods
  - Exploratory analysis of results
- Details of analysis results
  - Meta-analysis
  - Vote-count analysis
  - Review type analysis
- Full list of studies

In the following sections, examples of all the different types of analysis and summarising are given. However, **some of the sections or sub-sections of the methodology and analysis will be eventually shifted** to the summary (a) and / or scientific overview (b) (analysis results, Tables, plots etc.), and only detailed results and long Tables will remain in the Supporting document. This is to be decided by the author of each Synopsis, together with the WP/Task Leader once the analysis is completed.

### 7.5.1    Methodology

#### Literature search strategy

First you should **document how you selected the studies** for which you have created a coding template (i.e. studies that you have coded and "non-codeable" studies for which you filled an "empty: coding template"). The documentation should contain the following.
- Literature database searched
- Search terms
- # of initial records
- # of records after abstract screening
- Possible additional sources (grey literature, etc.)
- Inclusion/exclusion criteria
- Prioritisation criteria
- -> Final # of records

#### Analysis of study designs and methods

Ideally we only have one definition of a risk factor and only one indicator (change to crash risk), and just a small number of modifying conditions (e.g. different crash modification factors for rural and urban roads). However sometimes the situation is more complicated and **the resulting risk estimate depends on the indicator that is studied**. As an example: the results of mobile phone use while driving (speaking, texting...) seem to differ depending on the type of study. While driving simulator results indicate a strongly impairing effect (and therefore increase of risk), naturalistic driving

studies do not seem to confirm such a large risk. Moreover, within simulator experiments, there seems to be a strong impairing effect on driver speed and reaction time, but less certainty as regards the effect on speed variability and lateral control.

In Figure 7.2 it is shown how different types of **risks (R)** (crash risk, injury risk, prevalence, behaviour), **studied with different indicators (I) under different conditions (C)** can potentially lead to a very complex pattern of results. For each of the relevant CIR combinations (i.e. for each sub-cube in Figure 7.2) a separate overview has to be given.
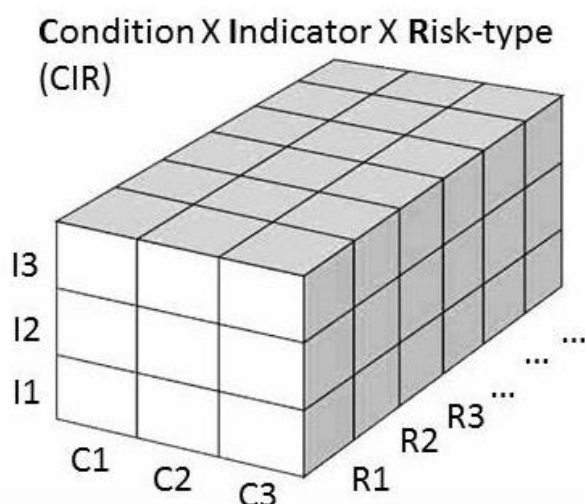


**Figure 7.2** Possible combinations of modifying conditions (C), type of indicator (I) and risk type (R)

It is therefore important to identify **the different conditions, indicators and types of risks studied and classify the studies and results accordingly**. You may present the results in a plot or tabulate them according to different variables. The variables should probably include some from the "core info" sheet (e.g. number of studies per road-type, vehicle type, or age-group involved), but can also come from the "flexible info" or "custom info" sheets depending on the topic.

Examples of such analyses using different variables are as follows, for ramp length (WP5) in Table 7.2 and for diabetes (WP4) in Table 7.3:

**Table 7.2** Description of coded studies designs / sample frames - ramp length (WP5)

| Author(s), Year | Sample and study design | Method of analysis | Outcome indicator | Main result |
|---|---|---|---|---|
| Chen et al., 2011 | One-lane exit ramps of Interchanges in the state of Florida, US.352 crashes in 60 sites were considered. | Poisson model | Crash frequency (number of crashes) | Longer exit ramps decrease the number of crashes for all passenger vehicles. |
| Chen et al., 2014 | 4 exit ramp types in the state of Florida, US. (573 crashes at 419 total exits). Only motorcycles were considered. | Negative binomial model | Crash frequency (number of crashes) | Longer exit ramps increase the number of motorcycle crashes. |
| Garnowski and Manner, 2011 | 3048 crashes at 197 ramps in Germany interchanges. | Random parameter Negative binomial model | Crash frequency (number of crashes) | Non-significant effect |

| Li et al., 2012 | 5538 crashes at 326 segments in the state of Florida. | Ordered probit model | Crash severity (5-scale) | Longer ramps increase severity of crashes |
| Wang et al., 2009 | 10,946 crashes at 231 exit segments in the state of Florida, US. | Ordered probit model | Crash severity (5-scale) | Longer ramps increase severity of crashes |
| Zhang et al., 2011 | 5539 crashes 326 motorway segments in Florida, US. | Ordered probit model | Crash severity (5-scale) | Longer ramps increase severity of crashes (at a 90% level) |

**Table 7.3** Description of coded studies designs and sampling frames - diabetes (WP4)

| Author, Year, Country | Sample, method/design and analysis | Risk group/ Cases | Control group/ Controls | Research conditions/ control variables |
|---|---|---|---|---|
| Bieber-Tregear, 2011 International | Meta-analysis. Random effects. 15 studies comparing crash involvement between diabetic and non-diabetic drivers | Diabetic drivers (crash/no crash) | Non-diabetic drivers (crash/no crash) | Comparison US and non-US studies |
| Bieber-Tregear, 2011 International | Meta-analysis. Random effects. 6 studies comparing crash involvement between insulin treated diabetic drivers and other 4 studies comparing prevalence of insulin-treated diabetic drivers among crash-involved and non-crash-involved drivers | Insulin treated diabetic drivers | Oral medication or diet treated diabetic drivers | Comparison US and non-US studies |
| Bieber-Tregear, 2011 International | Meta-analysis. Random effects. 4 studies comparing prevalence of diabetes between crash-involved and non-crash involved drivers | Crash-involved drivers (diabetes/no diabetes) | Non-crash in-volved drivers (diabetes/no diabetes) | 4 studies also reported on conditions of diabetes treatment (insulin, pharma-cotherapy, controlled diet alone). |
| Bieber-Tregear, 2011 International | Meta-analysis. Fixed effects. | Crash-involved drivers (insulin/ no insulin) | Non-crash in-volved drivers (insulin/no insulin) | Comparison US and non-US studies |
| Sagberg, 2006 Norway | Self-report questionnaires from 4448 crash-involved drivers. Odds ratio calcu-lated. | Cases: At fault (n = 2226) | Controls: Not at fault (n = 1840) | Analysis adjusted for age and annual driving distance. |
| Redel-Meier, 2009 Canada | In 2-y study interval 795 diabetic patients who had HbA1c values documented were reported to licensing authorities. Logistic regression. | Cases: 57 patients were involved in a crash | Controls: 738 were not involved in a crash | Analyses controlled for age, gender, medical complica-tion, history severe hypogly-caemia, age diabetes diag-nosed |
| Signorovitch, 2012 USA | Diabetes-2 people (not insulin treated) identified from a claims database (1998–2010). Crash occurrence leading to hospi-tal visits was com-pared between people with, and without claims for hypo-glycaemia. Analysis by multivariate Cox proportional hazard models. | n=5.582 people with claims for hypogly-caemia | n=27.910 with no such claims were | Analysis adjusted for demo-graphics, comorbidities, prior treatments and prior medical service use |
| Vingilis, 2012 Canada | Population-based large-scale panel research (N = 12.387). 524 (4.2%) report-ing an motor vehicle injury MVI 1996-2007. Path analyses examined the odds of subsequent MVI. | Diabetes reporting MVI, n =14 | Diabetic drivers not reporting MVI, n = 346 | Analysis controlled for age, gender and independent effects of medication use. |

| Orriols, 2014 France | 69.630 drivers involved in an injurious crash in France 2005-2008. Logistic regression analysis ; outcome = odds of being responsible for crash | Cases: drivers who were deemed responsible for the crash (n =33.200) | Controls: drivers who were not responsible for crash (n = 36.450). | Analysis adjusted for age, gender, socio-economic category, month, time of day, vehicle type, alcohol level, injury severity, expo-sure to medicines and other long-term diseases. |
| --- | --- | --- | --- | --- |

In some cases, **results may be too many and too heterogeneous to thoroughly consider each CIR combination**. You may choose the variables you find most appropriate to identify the differences in study designs and methods. Below in Table 7.4 is an extract of the example of such a case for rainfall (but such a level of detail may seldom be necessary) - the full Table is available in the respective example Synopsis.

*The main purpose of these Tables is to assist partners in properly summarizing the studies and in most cases they need not be included in the summary or the scientific overview. If you think that such Tables will not be helpful for your topic, you may skip this step.*

*Once the prototype DSS is available (expected date: September 2016), SafetyCube partners will be able to produce these Tables automatically by using the DSS (or at least a Table with all their coded studies per topic, to be easily customised in excel).*

**Table 7.4** Extract of description of coded studies designs and sampling frames - rainfall (WP5)

| Author(s) | Period covered | Country | Methodology /Design | Sample | Outcome variable | Road user type | Type of road | Weather variables | Other variables |
|---|---|---|---|---|---|---|---|---|---|
| Bergel, 2004 | 1975-1999 | France | Time series – DRAG model. Monthly data | 300 months | • Injury crashes<br>• Fatalities | • All | • All<br>• Secondary roads<br>• Urban roads<br>• Motorways<br>• Main roads | • Rainfall (mm) | • Temperature<br>• Frost days<br>• Traffic (on motorways and main roads only) |
| Bergel, Rattaire, Aron, Doucet, Violette, 2010 | 1995-2005 | France | Relative risk<br>Risk ratio | / | • Injury crashes | • All | • All<br>• Minor roads<br>• Secondary roads<br>• Motorways<br>• Main roads | • Rain (exposed or not) | / |
| Bijleveld, Churchill, 2009 | 1987-2006 | Netherlands | Approximate likelihood model. Daily data | 7304 days | • Fatalities plus hospitalized casualties | • All<br>• Pedestrians<br>• Cyclists<br>• Light mopeds<br>• Mopeds<br>• Motorcyclists<br>• Cars<br>• LGV<br>• HGV | • All | • Precipitation duration | |
| Brijs, Karlis, Wets, 2008 | 2001 | Netherlands | Time-series Poisson Interger Autoregressive model (INAR) for count data. Daily data | 365 days | • [Injury] crashes | • Cars | • All | • Precipitation duration<br>• Intensity of rain | • Temperature<br>• Sunshine<br>• Exposure (vehicles counts) |
| Edwards, 1998 | 1980-1990 | United Kingdom | Severity ratio Rain versus Fine weather | / | • Severity score | • All | • All | • Presence of rain (vs fine weather) | / |

### 7.5.2 Exploratory analysis of results

Based on the analysis of designs and methodological characteristics of the studies that you want to summarise, **you now have to decide what the appropriate CIR combinations are**. Which conditions can be merged into one? Which conditions lead to different results? The latter ones form the C's in your CIR cubes. The I's depend on the measure of effect. Generally you should only compare studies with the same measure of effect. The R's in the CIR cubes refer to different aspects of the risk variable. **Within each CIR group, there should be studies that are reasonably comparable**. They departed from (approximately) the same definition of the risk factor, they used the same measure of effect and they agree in the most important conditions. **For these studies it makes sense to summarize them numerically, for example in a meta-analysis**.

In principle, **the measure of effect selected in the coding template should be the same in order to be able to compare studies**. However, some measures of effects can be transformed into each other (e.g. Relative Risk; Percentage change in crashes; and Percent crash reduction). If the effects are slopes, they should only be compared if they refer to the same basic model (linear, logistic, etc.) and contain at least roughly the same other variables. If the effects are mean differences they should refer to identical or at least comparable variables. If the underlying variables are different but roughly comparable (e.g. resulting from different scales but in response to similar questions) one can express the differences in terms of the standardised indicators to make them comparable.

Assuming that for each study you have a measure of effect (e.g. a crash modification factor, an odds ratio, a mean decrease in a particular indicator, a slope in a regression analysis, etc.) and the Standard error (SE) of this effect, you can create a forest plot that gives an overview of the range of results in different studies. An example is given in Figure 7.3.
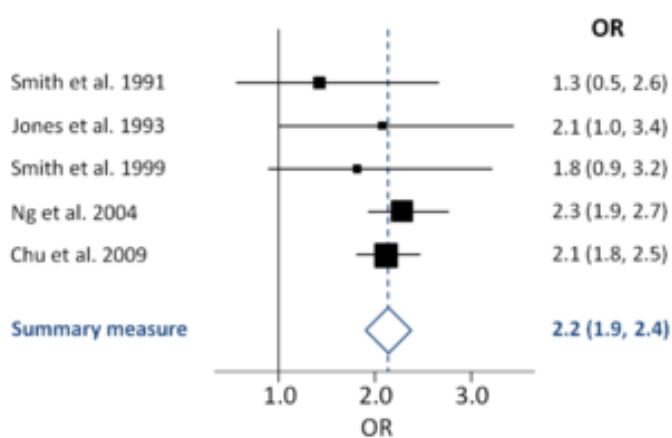


**Figure 7.3** Example for a forest plot to summarize the results of several (comparable) studies

If you have more than one CIR group, you have to present one plot for each cell. Make sure that each plot is clearly labelled. Below in Figure 7.4 is an example of a forest plot for the effect of ramp length on crash severity, by means of regression coefficients:
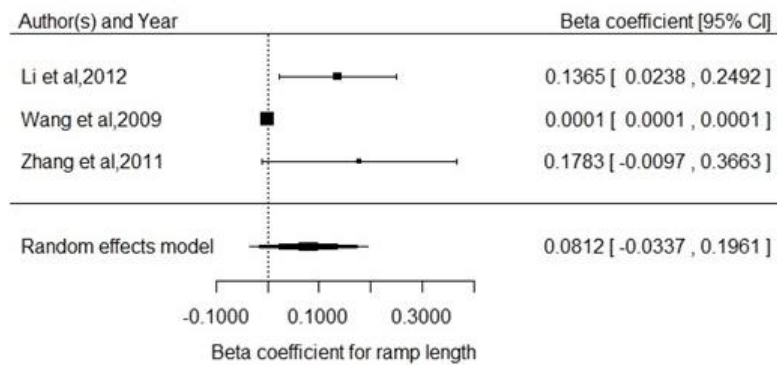
**Figure 7.4** Forest plot for ramp length effect on crash severity

**When the measures of effect for a CIR group differ substantially** and more detail is needed to get an overall picture of the results, you may create a custom Table providing an overview of the available estimates. Table 7.5 below presents such a Table for diabetes, where there are few and rather heterogeneous studies, therefore a simple listing Table with the main features of studies and effects is informative:

**Table 7.5** Overview of results of coded studies - Diabetes (WP4)

| Author, Year, Country | Risk factor | Study type | Outcome variable | Effects for Road Safety | Main outcome -description |
|---|---|---|---|---|---|
| Bieber-Tregear 2011 International | Diabetes 1 and 2 | Meta-analysis Random effects 15 studies | Crash in-volvement | RR=1.126; 95% CI: 0.847–1.497; p=0.415 | Increased crash risk was small and not statistically significant |
| Bieber-Tregear 2011 International | Insulin-treated diabetes | Meta-analysis. Random effects. 6 studies | Crash in-volvement | OR = 1.537; 95% CI: 0.603–3.915, p=0.368). | Non-significant increase in crash risk for insulin-treated drivers when compared with drivers treated with oral medication and/or diet alone. |
| Bieber-Tregear 2011 International | Diabetes 1 and 2 | Meta-analysis Random effect 4 studies | Crash in-volvement | OR = 1.052; 95%, CI: 0.970–1.141; p=0.220 | Drivers with diabetes are not over-represented among samples of drivers who have experienced a crash |
| Bieber-Tregear 2011 International | Insulin-treated diabetes | Meta-analysis Fixed effects 4 studies | Crash in-volvement | OR=1.212; 95% CI: 0.939–1.563, p=0.139 | Drivers with insulin controlled diabetes tend to be over-represented among samples of drivers who have experienced a crash; not statistically significant |
| Sagberg 2006 Norway | Diabetes 1 and 2 | Questionnaire study. Induced exposure: at fault crash-involved drivers compared not at fault. | Self-reported crash culpa-bility | Non-medicated diabetic drivers: (Diabetes Type II) OR=3.08, p = 0.05 | The adjusted odds ratio was significant for non-medicated diabetic drivers. For diabetic drivers on medication (Diabetes 1) the OR was non-significant. |
| Redelmeier 2009 Canada | Glycemic control | A population-based case control analy-sis | Crash in-volvement | OR= 1.26, 95% CI:1.03–1.54) | Crash risk increases 26% for each 1% reduction in HbA1c (finding robust after control for con-founders) |

| Signorovitch 2012 USA | Hypo-glycaemia (Diabetes 2) | Case-control comparing diabetes 2 patients with and without evidence hypoglycaemia | Crash involvement (resulting in hospital visit) | Hazard ratio (HR) = 1.82 (95% CL: 1.18, 2.80) People < 65 years; HR = 2.31 (95% CL: 1.44, 3.70) | After adjusting for baseline characteristics, hypoglycaemia significantly increased hazard |
|---|---|---|---|---|---|
| Vingilis 2012 Canada | Diabetes 1 and 2 | Population-based large-scale panel research | Motor vehicle injury | OR = 1.479, 95% CI: 0.743 - 2.944; p = 0.266 (NS). | No significantly increased odds of subsequent MVI was found for diabetes |
| Orriols 2014 France | Diabetes 1 and 2 | Case-control analysis comparing responsible vs. non-responsible crash-involved drivers. | Crash culpability (estimated by standard method) | Diabetes type 1: OR = 1.47; 95% Cl 1.12–1.92; p = 0.0047 | Significantly increased risk of being responsible for a crash found for drivers with type 1 diabetes. Type 2 diabetes not selected in final risk model. |

However, **if this Table was shifted in the Scientific Overview, a simplified version would be recommended** e.g. not reporting both CI and p-value, not reporting results on the same data twice (here meta-analysis with and without random effects model). You may choose the most adequate / reliable estimate and report that one.

In other cases, there may be no comparable results to be found in the studies for the risk factor hence, there will be no reported statistical numbers (see Table 7.6 below for headway distance).

**Table 7.6** Overview of results of coded studies - Headway distance (WP4)

| Author, year, country | Risk factor | Study type | Outcome variable | Effects for Road Safety* | Main outcome - description |
|---|---|---|---|---|---|
| Dingus, 2016, USA | Following too closely | Naturalistic driving, Case Control, randomized | Crash events (all crashes) | ↓ | Risk to be involved in a crash when "following too closely" is 13.5 times higher (Odds Ratio); |
| Summala, 2014, Sweden | Risky driving (speeding, crossing of no-passing lanes, close following and/or driving in the left lane or middle of the road) | Case control + matched control group, statistical control for various person characteristics and mileage | Police recorded traffic offences (2009-2011), number of offences | ↓ | Risky drivers have sig. more reported offenses in their driver records (hazard ratio) |
| Summala, 2014, Sweden | Risky driving | | Police recorded traffic offences Type of offence: traffic violations | ↓ | Risky drivers have sig. more traffic violation |
| Summala, 2014, Sweden | Risky driving | | Police recorded traffic offences Type of offence: endangering traffic safety | – | No difference between risky drivers and non-risky drivers in reported serious traffic offences |
| Duan, 2013, | Oncoming traffic | Simulator-experiment, | Measured headway in s | ↓ | Oncoming traffic decreases sig. headway |

| China | | randomised | (risk rate) | | distance |
|---|---|---|---|---|---|

There will be some cases where **the available studies per CIR are both too many and too heterogeneous** in their methods and types of outcomes used. It may still be informative to define the CIR groups and qualitatively summarise the effects per CIR and other necessary features. Below in Table 7.7 is a related extract of the example regarding rainfall (the full Table is available in the respective example Synopsis) - this Table would marginally "qualify" to be included in the Summary or the Scientific Overview, but it would be a helpful step towards a vote-count analysis:

**Table 7.7** Extract of overview of results - Rainfall (WP5)

| Author(s) | Country | Period covered | Dependant / outcome type | Effect on road safety | Traffic volume taken into account? | Main outcome - Description |
|---|---|---|---|---|---|---|
| Bergel, 2004 | France | 1975-1999 | Injury crashes – All | ↗ | N | - Rainfall height was linked, positively, to the total number of injury crashes and fatalities |
| | | | Injury crashes – Secondary roads | ↗ | N | |
| | | | Injury crashes – Urban roads | ↗ | N | |
| | | | Injury crashes – Motorways | ↗ | Y | |
| | | | Injury crashes – Toll Motorways | ↗ | Y | |
| | | | Injury crashes – Main roads | ↗ | Y | |
| | | | Fatalities – All | ↗ | N | |
| | | | Fatalities – Secondary roads | ↗ | N | |
| | | | Fatalities – Urban roads | ↗ | N | |
| | | | Fatalities – Motorways | ↗ | Y | |
| | | | Fatalities – Toll Motorways | ↗ | Y | |
| | | | Fatalities – Main roads | ↗ | Y | |
| Bergel, Rattaire, Aron, Doucet, Violette, 2010 | France | 1995-2005 | Injury crash risk – All | ↗ | Y | - The added risk appears high (2.4 in average in 2004). |
| | | | Injury crash risk – Motorways | ↗ | Y | - In 2005, the added risk is the highest on main roads (2.64), second on secondary roads (2.41) and on motorways (2.38), third on minor roads (1.88). |
| | | | Injury crash risk – Main roads | ↗ | Y | |
| | | | Injury crash risk – Secondary roads | ↗ | Y | |
| | | | Injury crash risk – Minor roads | ↗ | Y | - The added risk is higher outside built-in areas (2.6) than the average value for the whole of France, and thus than inside built-in area (2) |
| | | | Injury crash risk – Inside built-in area | ↗ | N | |
| | | | Injury crash risk – Outside built-in area | ↗ | N | |
| | | | Fatalities | - | N | |
| Bijleveld, Churchill, 2009 | Netherlands | 1987-2006 | Fatalities plus hospitalized casualties - All | ↘ | N | - The effect is different for different levels of crash severity and this effect is different for vulnerable transport modes and less vulnerable transport modes. |
| | | | Fatalities plus hospitalized casualties - Pedestrians | ↗ | N | |
| | | | Fatalities plus hospitalized casualties - Cyclists | ↗ | N | - The number of fatalities appears to be less sensitive to the duration of precipitation than the number of in-patients, which in turn is less sensitive to the duration of precipitation than the number of slightly injured. |
| | | | Fatalities plus hospitalized casualties – Light mopeds | ↗ | N | |
| | | | Fatalities plus hospitalized casualties - Mopeds | ↗ | N | |
| | | | Fatalities plus hospitalized casualties - Motorcyclists | ↗ | N | |
| | | | Fatalities plus hospitalized casualties - Cars | ↘ | N | - Coefficients among vulnerable modes of transport are relatively low, and some are even negative |
| | | | Fatalities plus hospitalized casualties - LGV | ↘ | N | |

| | | | Fatalities plus hospitalized casualties - HGV | ↘ | N | |
|---|---|---|---|---|---|---|
| Brijs, Karlis, Wets, 2008 | Nether-lands | 2001 | Car injury crashes | ↗ | Y | - Intensity of rain and precipitation duration are highly significant: the intensity/duration of the rain increases, then this leads to a higher number of crashes |
| Edwards, 1998 | United Kingdom | 1980-1990 | Severity | ↘ | N | - Crash severity decreases signifi-cantly in rain compared with fine weather |

### 7.5.3 Summarising the results

Once there is a clear picture of the CIR groups which are applicable for each topic, and the character-istics of the available studies and the effects there-in have been identified, the appropriate way to make a synthesis of the results should be decided. Three ways are proposed to summarise the re-sults for the SafetyCube Synopsis, in the following order of priority:

a. **Meta-analysis,** if the conditions apply (see details below).
b. **Vote-count analysis,** if a meta-analysis is not possible due to large differences in studies, and there is a sufficient number of studies.
c. **Review-type analysis, e.g. summary Table of effects,** if the number of studies is small and vote-count analysis is not meaningful.

**Criteria for conducting a meta-analysis**

Each CIR group should in principle contain **studies that are more or less comparable** and therefore suitable to conduct a meta-analysis.

- Did the studies employ (roughly) the same designs? (e.g. experimental study with treatment group and control group, safety performance function, time series analysis etc.)
- Are there at least 3 of the studies of sufficient quality?
- Is the same measure of effect and its standard error reported or deducible for each study (e.g. classic ANOVA F-test, negative binomial model parameters, ARIMA time series analyses etc.)?

If this is the case, you can proceed to **test the homogeneity of the studies in a funnel plot**. If this is "well-behaved", you can conduct a meta-analysis – if it is not, you might have to rethink your CIR groups. [In the Chapter on Meta-analyses, it will be described what is meant by "well-behaved"].

For example, in the analysis of ramp-length (WP5), the respective funnel plot (Figure 7.5) has been created from 3 studies. Publication bias is controlled for, but significant heterogeneity is present in the results. This means a random effects meta-analysis should be performed (see Table 7.8). Please note that **all 3 studies concerned the effect of ramp length on crash severity, which was studied in all 3 cases by of ordered probit models, with the same 5-scale variable of severity.**
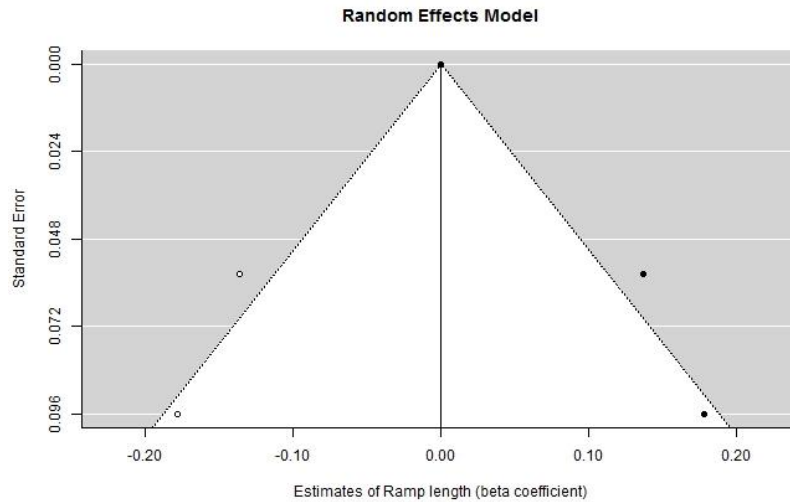
**Figure 7.5** Adjusted Funnel Plot for Publication Bias - Effect of ramp length on crash severity (WP5 meta-analysis)

**Table 7.8** Random effects meta-analysis for ramp length effects on crash severity

| Variable | Estimate | Std. Error | p-value | 95% CI |
|---|---|---|---|---|
| Ramp length | 0.0812 | 0.0586 | 0.1659 | (-0.0337, 0.1961) |

### Vote-count analysis

**If the studies are too different to compare them in a formal analysis,** you should summarise them as a vote-count. Each study is considered to give a vote for or against the risk-factor/ measure. You should calculate the following percentages:

- Studies with significant positive effect
- Studies without significant effect
- Studies with significant negative effect

Unfortunately, a vote count does not increase the power of the joint analysis as compared to the single analyses. If many non-significant results are analysed in a meta-analysis, it will make a difference whether the non-significant results all show the same tendency (in which case it can become significant in the meta-analysis) or whether they are equally spread around a null-effect. In a vote count this is not the case.

You may **choose the appropriate variables to differentiate the vote-count analysis** on the basis of your CIR groups: apart from the different measures of effects, there may be interest in performing the vote-count analysis for different road user types, road types etc. See below the example for rainfall effects per road type (Table 7.9).

**Table 7.9** Vote-count analysis results for risk and crash occurrence by road type - rainfall (WP5)

| | Total number of effects tested | Result (number of effects) | | | | Result (% of effects) | |
|---|---|---|---|---|---|---|---|
| | | ↗ | - | ↘ | | ↗ | ↘ |
| **Risk** | 32 | 30 | | 2 | | 94% | 6% |
| **all** | 20 | 18 | | 2 | | 90% | 10% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **motorway** | 6 | 6 | | | | 100% | 0% |
| **rural** | 6 | 6 | | | | 100% | 0% |
| **Crash Occurrence** | **52** | **24** | **10** | **18** | | **46%** | **35%** |
| **all** | 42 | 15 | 10 | 17 | | 36% | 40% |
| **motorway** | 2 | 2 | | | | 100% | 0% |
| **rural** | 4 | 4 | | | | **100%** | **0%** |
| **urban** | 4 | 3 | | 1 | | 75% | 25% |

**Review-type analysis**

In some cases, a vote-count analysis will not be meaningful, e.g. **in cases where the results are both rather few** (i.e. it is not meaningful to calculate percentages for 5 studies) **and too heterogeneous** (to perform a meta-analysis). In this case, the most informative way of summarising the results is through a qualitative summary Table, with the related interpretation, as the ones presented in Table 7.5 and Table 7.6 for the effects of diabetes or headway distance (WP4) - but possibly with even less detailed information.

*Such a Table should preferably provide more information and insights than that of the standard DSS output, to qualify for inclusion in the summary.*

### 7.5.4 Full list of studies

At the end of the Supporting document, provide the list of all studies. For each study consider giving (if not already included in one of the above Tables)
- Study summary
- N
- Biases (if any and not mentioned in summary)

*On the basis of the analysis and synthesis of results, the scientific overview of the topic and eventually the summary should be drafted. A selection should be made of Tables, plots and analysis results to be presented in the scientific overview and / or the summary. All other background or detailed contents (e.g. large Tables of exploratory analysis or detailed study / effects comparisons) should remain in the Supporting document.*

## 7.6 SCIENTIFIC OVERVIEW

The scientific overview should be a 4-5 page document describing the analysis carried out to produce the results and information as per the effect of the risk factor / measure and its transferability conditions. This document aims to describe the way the reported effects have been estimated, with a full analysis of the methods and results, in order to give the user all the necessary information to understand the results and assess their validity. But it does not include "appendix-type" information, like the literature search strategy or the full lists of studies, detailed Tables etc. that may have been used as a background for performing the analysis (these will be included only in the Supporting document).

A typical structure could be as follows:
- *Literature review*

- *Description of the available studies*: description of study designs, measures of effects examined, limitations etc.
- *Presentation of the analysis methods and results*: either a meta-analysis, a vote-count analysis of a review-type analysis, present and interpret the results, including potential biases and limitations.

### 7.6.1 Literature review

A literature review may further the knowledge presented in the summary, on the **definitions, mechanisms and modifying conditions** of the risk factor or measure. It may include **qualitative aspects** that are mentioned in studies which are not "codeable", and guide the DSS user to further reading (e.g. fact-sheets, review papers with no quantitative results, also non "codeable"). You may also include a key graph from the literature that highlights the mechanism or the determinants. You may include all the relevant citations.

### 7.6.2 Crash data or scenarios

You may include a graph or Table of macroscopic and / or in-depth data on the risk factor. For macroscopic data, you may consider a summary Table or graph from the latest version of the ERSO basic fact sheets (BFS) based on the CARE/CADAS data, if available. For in-depth data, you may include (if possible) a radar graph on the basis of SafetyCube crash scenarios (on the basis of GIDAS, LAB, IGlad, or DaCoTA data) ). Figure 7.6 shows a radar plot example for rainfall.

**Figure 7.6** Distribution of crash types under dry and wet road conditions



Source GIDAS, 2016
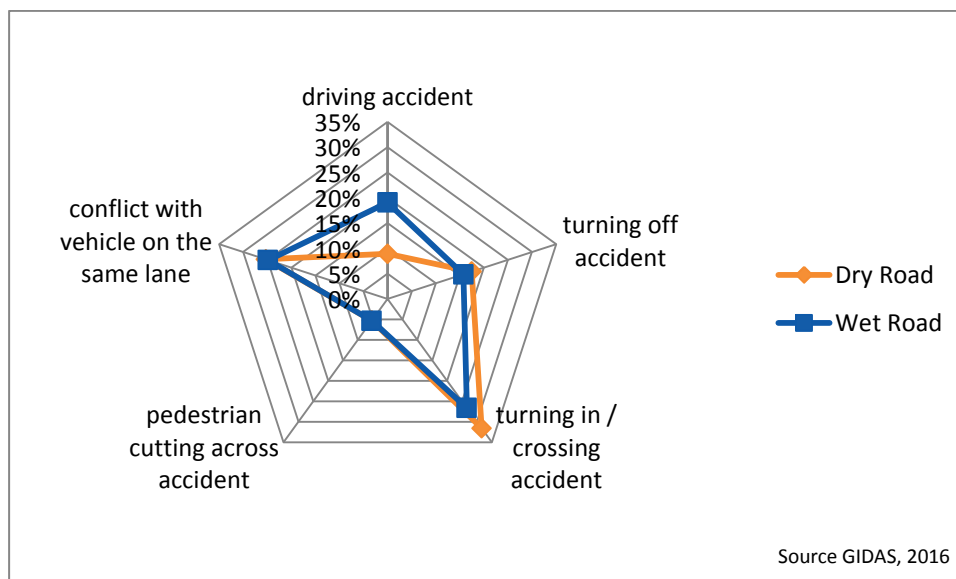
### 7.6.3 Description of studies

The studies should be described in detail at a content level as well as at a methodological level. This section should give an impression how well the topic has been studied, under which different conditions and in which methodological designs. It should give the DSS user an idea how reliable and unbiased the results are and how much is known about their transferability across different conditions. In a sense, it will be a more elaborated version of what is included in the summary (see Section 7.7). You may provide a figure or graph with the main study designs (as for example in the WP4 diabetes example in Figure 7.7 below), or another type of summary Table (but not one with the full list of studies of these are too many, as this will be available through the DSS output on the topic).
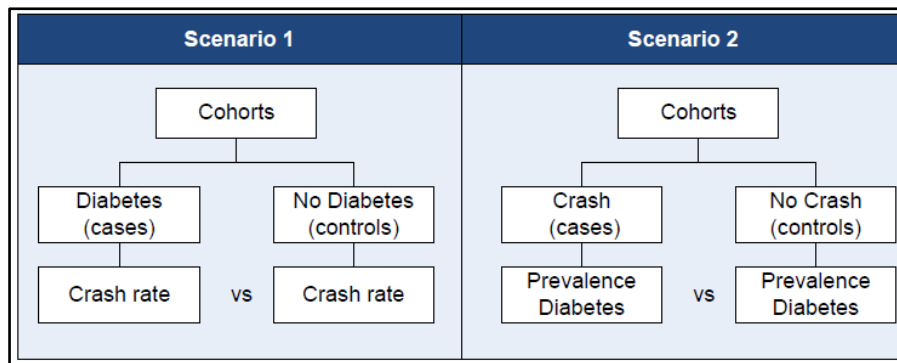
**Figure 7.7** Scenarios for Investigating Risk of Crash in Diabetes (Bieber-Tregear et al., 2011) - from the WP4 diabetes example

At the methodological level, you should look at the different type of analyses that have been conducted to study the topic. This can concern the measure of effect that is used to measure the effect of the risk factor or measure, the study design (pre-post, with / without control group, Empirical Bayes, crash prediction model, case-control study, matching / no matching), the type of regression model used (if any) and the variables that have been statistically controlled for. Features that are common to all studies on the topic can be described in the text.

Aspects on which the studies differ can be listed in a table indicating the number of each study-type. Sampling frame and methodology can be described in one joint table if the study-population is relatively homogeneous, in the sense that there are more common characteristics than differences that need to be listed in the table. When studies vary strongly in sample size, consider giving an indication of the total N (sum of the studies sample sizes) in addition to the number of studies.

*This section should allow the reader to understand how the CIR groups for the analysis of the effects are defined on the basis of the existing studies.*

### 7.6.4 Description of analysis carried out

If you have done a meta-analysis, **present the meta-analysis** (not necessarily with all the details included in the Supporting document) - but with enough detail to be transparent. Report the study selection criteria for the studies included in the meta-analysis, heterogeneity and publication bias tests, the method used (e.g. fixed vs. random effects) and the reasons for selecting it. Discuss the results in detail, explaining *how these lead to providing (or not) a "best estimate" of the effect.*

If you have done **a vote-count analysis, present the Table(s)** (not necessarily with all the details included in the Supporting document). Discuss the results in detail, explaining *how these lead to providing (or not) a "best estimate" of the effect.*

If you have done a **review-type analysis**, consider including a not too overwhelming summary Table of effects. This should include more insights and information than the standard DSS output Table, but not a full listing of effect-per-effect and study-per-study comparisons. Comment and interpret the results, explaining *how these lead to providing (or not) a "best estimate" of the effect.*

## 7.7 SUMMARY

A two-page document including the abstract of the topic and an overview of effects and analysis methods, as follows:

- *Colour code*
- *Key-words*
- *Abstract:* Main result, general gist of how sure we are and main modifying conditions.
- *Background:* Knowledge text, i.e. general information on the mechanisms by which this measure/risk causes/avoids crashes or affects road safety, relevant target groups, and relation to other risk factors / countermeasures.
- *Overview results:* summary of risk factor results by condition X indicator X risk type
- *Notes on analysis methods*: Brief description of analysis methods, potential biases and limitations, disclaimers etc.

The summary should report the key aspects of the topic, the main results and transferability conditions; **it is the actual synopsis.** The analysis carried out to report these main results and conditions should be described in the scientific overview.

### 7.7.1 Colour code

Indicate how important the risk factor is/ how effective the countermeasure, on the basis of the following scale:

| Risk factor | | Countermeasure | |
|---|---|---|---|
| Red | Results consistently show an increased risk when exposed to the risk factor concerned. | Green | Results consistently show that the countermeasure reduces road safety risk. |
| Yellow | There is some indication that exposure to the risk factor increases risk, but results are not consistent. | Light green | There is some indication that the countermeasure reduces road safety risk, but results are not consistent. |
| Grey | No conclusion possible because of few studies with inconsistent results, or few studies with weak indicators, or an equal amount of studies with no (or opposite) effect. | Grey | No conclusion possible because of few studies with inconsistent results, or few studies with weak indicators, or an equal amount of studies with no (or opposite) effect. |
| Green | Results consistently show that exposure to the presumed risk factor does not increase risk. | Red | Results consistently show that this measure does NOT reduce road safety risk and may even increase it. |

**Figure 7.8** Colour codes for risk-factors and countermeasures

Risk factors

Risky (red)

Appoint this category when the results are relatively consistently showing an increased risk upon exposure to the risk factor in question.

- Good number of studies (at least 3 per relevant condition, at least 5 in total)
- Fair quality (at least for a number of studies showing the result)
- Consistency across studies & conditions
- Good indicator (proven relation with road crashes)

### Probably risky (yellow)

Appoint this category if there is some indication that exposure to the risk factor increases the crash/injury risk, but the results are not consistent. This could either be due to conditions under which the risk factor has been shown to be unproblematic, or because the study results are inconsistent (but with the majority of studies pointing to an increased risk).

You should also chose this category (rather than "risky") if you have only very few studies that properly investigated the effect (say less than 5) or if the effect on road safety is tested via an indicator of which the relation with crashes is not very strong. A strong indicator would for instance be a proven effect on driven speed, a weak indicator would be a proven effect on self-reported speed.
- Few studies (of sufficient quality)
- Inconsistent results (but majority showing risk effect)
- Weak indicator (unsure relation to crashes)
-> if several of these apply, however, chose "unclear".

### Unclear (grey)

If several of the above issues apply (few studies with inconsistent results, or few studies with weak indicators or an equal amount of studies with no (or opposite) effect as those that show a risk effect.
- No studies that investigate effect
- Mixed results
- Insufficient quality & quantity of studies that might show a risk effect

### Probably not risky (green)

The absence of a risk effect is a null-effect which is notoriously difficult to "prove". Moreover, a risk factor of which it has been shown that it does *not* pose a risk, is not very likely to be included in the DSS. So there will probably not be too many factors in this category.

An example could be snow or pot-holes. Both are assumed by many laymen to pose a big threat to road safety. However, research seems to indicate that due to compensatory behaviour on the road user side they seem to be protective rather than risk factors.

This category can only be chosen if there is a reasonably large number of studies with fair quality – such that one could expect a risk effect (if it existed) to become apparent in them.
- Sufficient quantity and quality of studies
- Large majority of studies show null-effect or opposite effect
- Meta-analysis on a large number of studies shows no significant effect or opposite effect

### Countermeasures

### Effective (dark green)

Appoint this category when the results are relatively consistently showing a countermeasure to be effective.
- Good number of studies (at least 3 per relevant condition, at least 5 in total)
- Fair quality (at least for a number of studies showing the result)
- Consistency across studies & conditions
- Good indicator (proven relation with crashes)

### Probably effective (light green)

Appoint this category if there is some indication that the countermeasure is effective, but the results are not consistent. This could either be due to conditions under which the countermeasure has been shown to be ineffective, or because the study results are inconsistent (but with the majority of studies pointing to an effective measure).

You should also chose this category (rather than "effective") if you have only very few studies that properly investigated the effect (say less than 5) or if the effect on road safety is tested via an indicator of which the relation with crashes is not very strong. A strong indicator would for instance be a proven effect on driven speed, a weak indicator would be a proven effect on self-reported speed.

- Few studies (of sufficient quality)
- Inconsistent results (but majority showing risk effect)
- Weak indicator (unsure relation to crashes)

-> if several of these apply, however, chose "unclear".

### Unclear (grey)

If several of the above issues apply (few studies with inconsistent results, or few studies with weak indicators or an equal amount of studies with no (or opposite) effect as those that show a positive effect of the countermeasure.

- No studies that investigate effect
- Mixed results
- Insufficient quality & quantity of studies that might show a risk effect

### Ineffective or even counterproductive (red)

Showing that a countermeasure is not effective is a null-effect which is notoriously difficult to "prove". And the fact that no positive effect of a countermeasure could be measured so far, could also mean that the methods used are not sensitive enough to pick up an effect. However, if there are a number of good quality studies and none of them have found an effect, or worse if the majority of the studies showed a negative effect, the DSS user should be made aware of this.

An example could be skidding schools, which seem to be counterproductive because they boost the confidence of the participants making them less careful.

This category can only be chosen if there is a reasonably large number of studies with fair quality – such that one could expect a risk effect (if it existed) to become apparent in them.

- Sufficient quantity and quality of studies
- Large majority of studies show null-effect or opposite effect
- Meta-analysis on a large number of studies shows no significant effect or opposite effect

### 7.7.2 Key-words

Include a number of key-words associated with the risk factor, including the related taxonomy levels, and any additional ones that are relevant to the topic (e.g. road user types, road types etc.). This will also help to properly "tag" the synopses in the DSS so that they appear in all related searches.

### 7.7.3 Abstract

The following topics should be mentioned in the abstract. These are to be further elaborated in the main text of the summary and scientific overview.

- Definitions of risk-factor or measure
- Outcome measures
- Main result
- Modifying conditions
- Transferability
- Disclaimers (e.g. Publication bias, few studies or poor study quality, etc.)

**Examples of abstracts**

**Crash cushions**

It is found that the likelihood of fatality is lower when colliding with a crash cushion than when colliding with any of the investigated fixed objects. However, the probability of injury increases for most objects. For instance, for colliding with a pole, crash cushions are estimated to reduce fatality crashes by 62%, but increase injury crashes by 74 %. The study results are rather heterogeneous, and may be affected by regression to the mean. No studies compared different types of crash cushions.

**Guard rails**

Comparing risk of crashes (fatal and injury) from hitting wire, steel or concrete guardrails along the roadside. Estimates are based on studies that compare the risk of injury/fatality for colliding with different types of guardrail. It appears to be less dangerous to collide with a wire guardrail than with steel or concrete guardrails. The results are, however, not statistically significant, and the confidence intervals are large. Colliding with a steel type compared to concrete type guardrail appears to increase the risk for fatality, but reduce the risk of injury. However, for motorcyclists the risk of both fatality and injury is higher for colliding with steel than concrete guardrails. The effect on crashes of rigid versus yielding poles is unclear. Generally, guardrails that are more rigid reduce the risk of fatal and injury crashes, but not necessarily for motorcyclists.

### 7.7.4 Background

*This section aims to summarise the main points of the literature review and study description carried out for the scientific overview. You may use questions as headings e.g. see the diabetes example. The order of headings below is indicative and each author may decide the most useful way of presenting the background.*

**Introduction to risk measure and its effect on crashes or injuries**

This should be the basic knowledge text introducing the ***mechanisms*** (if known) how a risk factor or a measure affect road safety (i.e. either traffic or driving behaviour or vehicle functioning). Why does it increase / decrease the crash or injury risk? Introduce the ***main modifying conditions***. Explain why the risk factor / measure has different effects under different circumstances.

**Prevalence**

In case of risk factors give also information about the prevalence (if available).

**Definitions of risk factor or measure**

Define the risk factor or measure in question. If there are different definitions, which lead to different results, explain. This is probably in particular relevant for summaries of measures, where different variants of a measure (e.g. different types of guard rails) might lead to different results.

### Measures of effect

It should be clear how the effect of the risk factor or road safety measure has been measured. Optimally, the effect will be measured as a change in crash risk or injury risk, but it could also have been measured as a change in a related variable (e.g. speeding, road safety attitude, simulator variables like lane deviation). Keep it simple – no details, just the general type of outcome measure.

### Study methods

Define how the effect of the risk factor / measure is usually investigated. Are there different ways (e.g. simulator studies vs. naturalistic driving studies, before-and-after studies with or without control group)? What are the pro's and con's? Do they normally lead to the same results?

### 7.7.5 Overview of results

*This section aims to summarise the main estimates and points of caution mentioned in the scientific overview.*

### Main results / Summary of effects

Depending on the differences between the results in each CIR group, a summary estimate for all groups can be given. This makes sense when the differences between groups are of a quantitative nature (i.e. all effects go in the same direction but differ in size). It does not make sense to present a summary estimate that covers up opposite tendencies.

You are encouraged to **provide a "best-estimate" of the effect**, this could be a crash modification factor (e.g. resulting from meta-analysis) but it could also just be a general tendency (… is more than ..; varies depending on…). Summarise the effects either by reporting:

- *the meta-analysis results,*
- or the range of effects per CIR group (minimum and maximum) e.g. **in a forest plot,**
- or the **general tendencies** of the vote-count or review-type analysis.

If you have done a meta-analysis mention whether the resulting effect is significant or not. If you have done a vote-count report percentage of studies with significant positive effect.

### Modifying conditions

Often risk-factors or measures do not have a homogeneous effect, but it varies depending on the environment or the road-user or vehicle type. As examples: Roundabouts reduce the crash risk for passenger cars but not for heavy goods vehicles or for cyclists. Drink-driving bears a particularly high risk on young inexperienced drivers. Make sure you name the most important differences in estimated effects. ***Describe the result pattern as simply as possible as complex as necessary***.

### Transferability

Indicate whether the risk / measure has been investigated under a broad range of conditions (e.g. in different countries, urban/rural/suburban settings, with different age groups, looking at different transport modes, etc.). If this is not the case indicate the main restriction. E.g. "has only been studied in a suburban context with middle class children" or "has only been studied in north-west European countries".

### 7.7.6 Notes on analysis methods

**Disclaimers**

If there are reasons not to trust the results (completely) give the main reasons. These could be publication bias; too few studies; studies likely to be biased; heterogeneous results, etc.

**Conclusion: how well is the topic studied?**

From the description of the study methods and measures of effect you should draw a conclusion about the quality and transferability of the results. Describe whether the effect has been tested under all conditions that seem relevant to you and name those under which it yet remains to be done or under which the results are yet unclear.

By unclear we mean that there are studies but for some reason you don't trust their results, or different studies have results in different directions. List the most obvious problems. Typical problems would be for example that pre-post studies did not have a control group, or that they had one but did not correct for the regression to the mean, or that in Case control studies there were obvious differences between the two groups compared or that a task used in an experimental study had little to do with the actual risk situation in traffic. If you see no particular problems and the number of studies is not too small you can carefully conclude that the issue seems well studied.

# 8  Meta-analysis

## 8.1    ANALYTICAL METHODS FOR META-ANALYSIS

By far the most common technique for doing meta-analysis of studies of risk factors or road safety measures is the inverse-variance technique. According to this technique, each estimate of risk or effect is assigned a statistical weight which is inversely proportional to its sampling variance. To be included in a meta-analysis using the inverse-variance technique, a study should provide two pieces of information:

1.   One or more estimates of the result of interest
2.   The standard error of each estimate

The summary estimate of risk or effect based on g individual estimates is:

$$\text{Summary mean} = \bar{Y} = \frac{\sum_{i=1}^{g} Y_i \cdot W_i}{\sum_{I=1}^{g} W_i}$$

Here $\bar{Y}$ is the estimate of the weighted summary mean, based on g individual estimates, each of which is assigned a statistical weight as follows:

$$\text{Statistical weight} = W = \frac{1}{SE_i^2}$$

### 8.1.1    Investigating the distribution of effects

The distribution of effects can best be investigated in a funnel plot. A funnel plot is a scatterplot of treatment effect against a measure of sample size. The horizontal axis of a funnel plot gives the size of effect, while the vertical axis gives a measure of precision, usually the standard error of the estimates (SE). Smaller studies will scatter widely at the bottom of the graph. Funnel plots are used primarily as a visual aid for detecting bias or systematic heterogeneity.
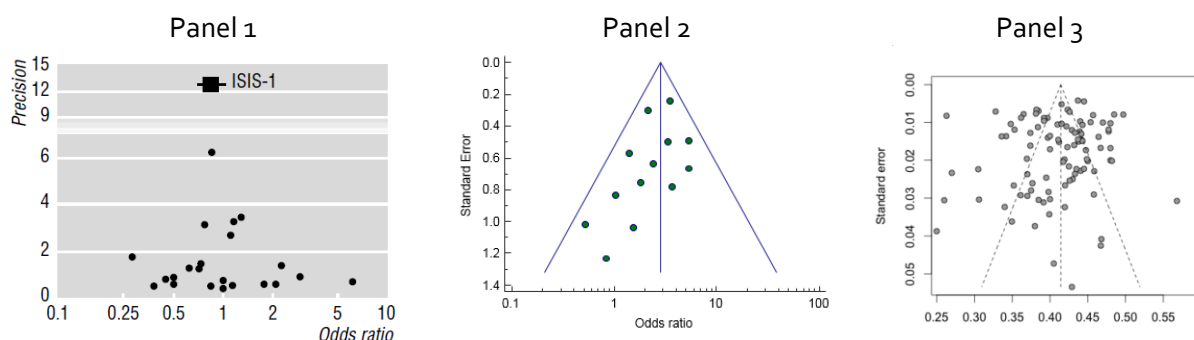


**Figure 8.1** Funnel plots with different degrees of symmetry (-> to be up-dated with SafetyCube examples)

A symmetric inverted funnel shape, as in panel 1 of Figure 8.1, arises from a 'well-behaved' data set, in which publication bias is unlikely. An asymmetric funnel, as in panel 2, indicates a relationship between treatment effect estimate and study size. This suggests the possibility of either publication bias or a systematic difference between smaller and larger studies ('small study effects'). If the dis-

tribution is not funnel shaped, like in panel 3, the data show heterogeneity, which might suggest that the studies are not measuring the same. In the worst case there might be more than one peak, in which case the studies should not be meta-analysed together – at least not without further investigation what caused the differences in results.

### 8.1.2  Assessing heterogeneity

To determine whether there is systematic between-study variation in results, a statistical test is performed by means of the following test statistic:

$$Q = \sum_{i=1}^{g} W_i \cdot Y_i^2 - \frac{\left(\sum_{i=1}^{g} W_i \cdot Y_i\right)^2}{\sum_{i=1}^{g} W_i}$$

Q is an estimate of variance. It is Chi-square distributed with g – 1 degrees of freedom. If Q is significant, the variance between studies is larger than would be expected on the basis of the within study variation.

Whether Q is significant or not depends – next to the heterogeneity – also on the sample size. With a very large sample, Q would practically always be significant and with a very small sample almost never. Therefore it has been suggested to calculate the percentage of variance that is due to heterogeneity between studies $I^2$.

$$I^2 = \left(\frac{Q - (g - 1)}{Q}\right) * 100\%$$

This describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance).

Thresholds for the interpretation of $I^2$ can be misleading, since the importance of inconsistency depends on several factors (e.g. the magnitude and direction of the effects). A rough guide given by the Cochrane Handbook for systematic reviews is as follows:
- 0% to 40%: might not be important;
- 30% to 60%: may represent moderate heterogeneity;
- 50% to 90%: may represent substantial heterogeneity;
- 75% to 100%: considerable heterogeneity (it is probably not appropriate to combine these studies in a meta-analysis).

### 8.1.3  What to do with heterogeneous data?

From Cochrane Handbook for systematic reviews:
http://handbook.cochrane.org/

#### Check your data

Severe heterogeneity can indicate that data have been incorrectly extracted or entered into your analysis. For example, if standard errors have mistakenly been entered as standard deviations for continuous outcomes, this could manifest itself in overly narrow confidence intervals with poor overlap and hence substantial heterogeneity. Including studies with different outcome measures, effect measures, or conditions that were compared will also increase the heterogeneity.

### Do not do a meta-analysis

A systematic review need not contain any meta-analyses. If there is considerable variation in results, and particularly if there is inconsistency in the direction of effect, it may be misleading to quote an average value for the intervention effect. It may be considered better to do a good critical review than an inappropriate meta-analysis.

### Explore heterogeneity

It is clearly of interest to determine the causes of heterogeneity among results of studies. Heterogeneity may be explored by conducting separate analyses for different conditions or sub groups e.g. considering studies using different age groups in separate analyses or meta-regression. Subgroup analysis should only be completed if there are sufficient studies with the specified characteristic to make it worthwhile, each study should only be assigned to one subgroup. Be aware that too many comparisons may increase false positives and negatives.

The interpretation of these results is however problematic. Explorations of heterogeneity that are devised after heterogeneity is identified (as opposed to modifying conditions that have been identified prior to conducting the meta-analysis) can at best lead to the generation of hypotheses. They should be interpreted with caution. Also, investigations of heterogeneity when there are very few studies are of questionable value.

### Perform a random-effects meta-analysis

A random-effects meta-analysis may be used to incorporate heterogeneity among studies. This is not a substitute for a thorough investigation of heterogeneity. It is intended primarily for heterogeneity that cannot be explained. See next Section (8.1.4).

### Change the effect measure

Heterogeneity may be an artificial consequence of an inappropriate choice of effect measure. For example, when studies collect continuous outcome data using different scales or different units, extreme heterogeneity may be apparent when using the mean difference but not when the more appropriate standardized mean difference is used. Furthermore, choice of effect measure for dichotomous outcomes (odds ratio, relative risk, or risk difference) may affect the degree of heterogeneity among results. In particular, when control group risks vary, homogeneous odds ratios or risk ratios will necessarily lead to heterogeneous risk differences, and vice versa. However, it remains unclear whether homogeneity of intervention effect in a particular meta-analysis is a suitable criterion for choosing between these measures.

### Exclude studies

Heterogeneity may be due to the presence of one or two outlying studies with results that conflict with the rest of the studies. In general it is unwise to exclude studies from a meta-analysis on the basis of their results as this may introduce bias. However, if an obvious reason for the outlying result is apparent, the study might be removed with more confidence. Since usually at least one characteristic can be found for any study in any meta-analysis which makes it different from the others, this criterion is unreliable because it is all too easy to fulfil. It is advisable to perform analyses both with and without outlying studies as part of a sensitivity analysis. Whenever possible, study characteristics that might lead to such situations should be specified in the protocol.

### 8.1.4   Fixed vs. random effects models

There are two models for inverse-variance meta-analysis: the fixed-effects model and the random-effects model. The fixed-effects model is based on the assumption that the variation in individual results consists of sampling variance only (random variation only occurs within studies), i.e. there is one true effect and all variance is fully explained in terms of the sampling random variation within studies. This is rarely appropriate as there are usually differences between studies e.g. due to the environment they are conducted in. However, if studies are conducted in the same environment and with the same sort of participants, this would suggest there should be a single true effect and a fixed-effects model should be used. The random-effects model is based on the assumption that there is systematic between-study variation in results (random error occurs both within and between studies), i.e. the true effect could vary from study to study, variation greater than sampling variance accounts for the difference in effect. For example, variation in effect may be due to variation in the age of participants or difference between geographical regions. If there is a lot of between-study variation (e.g. significant Q statistic, high $I^2$), a random-effects model of meta-analysis should be adopted. The statistical weights are then modified by adding a variance component, tau-squared ($\tau^2$), which is estimated as follows:

$$\tau^2 = \frac{Q - df}{C}$$

C is estimated as follows:

$$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i}$$

The modified statistical weight for each study becomes:

$$W = \frac{1}{SE_i^2 + \tau^2}$$

Please note that while the sampling variance ($SE^2$) varies from study to study, the between-study variance parameter ($\tau^2$) is a constant. It therefore flattens the statistical weights, which can vary considerably less from study to study in the random-effects model than in the fixed-effects model.

## 8.2   META-REGRESSION

### 8.2.1   Introduction

Meta-regression (or moderator analysis or meta-regression model or mixed effects model) is a tool used in meta-analysis to examine the impact of moderator variables on study effect size using regression-based techniques. In other words, meta-regression models are linear models that investigate the impact of (one or more) moderator variables on the outcomes. In most meta-regression approaches, the unit of analysis, (e.g. each observation in the regression model) is a study. It is noted that continuous as well as categorical moderator variables can be included in such models.

### 8.2.2   Theoretical Background

The theoretical background illustrated here can be found in more detail in Viechtbauer (2010).

If $i$=1,...,n independent effect size estimates, each estimating a corresponding true effect size.

$$y_i = \theta_i + \varepsilon_i,$$

where $y_i$ is the observed effect in the *i*-th study, $\theta_i$ is the corresponding (unknown) true effect, $\varepsilon_i$ is the sampling error ($\varepsilon_i \sim N(0, v_i)$). As a result, all the $y_i$'s are assumed to be unbiased and normally distributed estimates of their corresponding true effects. Note that the sampling variances $v_i$ are assumed to be known.

However, variability (or heterogeneity) can be present among true effects. A random effect model is used to account for potential heterogeneity.

In this case, the true effect $\theta_i$ is:

$$\theta_i = \mu + u_i,$$

where $u_i$ follows a normal distribution with mean value $\mu$ and variance $\tau^2$. If $\tau^2$ equals zero, then the true effects are assumed to be homogenous (i.e. $\theta_1 = \theta_2 = ... \theta_n = 0$). Therefore, $\mu$ equals $\theta$ (true effect).

Another way to deal with potential heterogeneity is to conduct a meta-regression. By doing so, the moderators included in the model may account for heterogeneity in the true effects (or for a part of).

In this case, the model is:

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{i\kappa} + u_i,$$

In this equation, $x_{ij}$ is the value of j-th moderator variable in the i-th study. Again, $u_i$ follows a normal distribution with mean value $\mu$ and variance $\tau^2$. It is noted that in meta-regression, $\tau^2$ is the amount of residual heterogeneity among the true effects (the variability among the true effect that cannot be explained by the moderators entered in the meta-regression model).

### 8.2.3   Suggested Software

- **R-studio (open source)** https://www.r-project.org/.
*Package: metafor* (it can be found at: http://CRAN.R-project.org/package=metafor)
*Manual: https://cran.r-project.org/web/packages/metafor/metafor.pdf*

The next Table (Table 8.1) illustrates and compares meta-analysis capabilities of the most well-known packages in R. It is observed that the *metafor* package is the most flexlible and in addition, it is the only package that handle meta-regression analysis.

**Table 8.1** Comparison of the capabilities of the metafor, meta, and rmeta packages for conducting meta-analyses in R. Notes: (1) Only fixed-effects with moderators model. (2) When used together with the copas (Carpenter and Schwarzer 2009) package. Source: Viechtbauer (2010).

| | metafor | meta | rmeta |
|---|---|---|---|
| *Model fitting*: | | | |
| Fixed-effects models | yes | yes | yes |
| Random-effects models | yes | yes | yes |
| Heterogeneity estimators | various | DL | DL |
| Mantel-Haenszel method | yes | yes | yes |
| Peto's method | yes | yes | no |
| *Plotting*: | | | |
| Forest plots | yes | yes | yes |
| Funnel plots | yes | yes | yes |
| Radial plots | yes | yes | no |
| L'Abbé plots | no | yes | no |
| Q-Q normal plots | yes | no | no |
| *Moderator analyses*: | | | |
| Categorical moderators | multiple | single[1] | no |
| Continuous moderators | multiple | no | no |
| Mixed-effects models | yes | no | no |
| *Testing/Confidence Intervals*: | | | |
| Knapp & Hartung adjustment | yes | no | no |
| Likelihood ratio tests | yes | no | no |
| Permutation tests | yes | no | no |
| *Other*: | | | |
| Leave-one-out analysis | yes | yes | no |
| Influence diagnostics | yes | no | no |
| Cumulative meta-analysis | yes | yes | yes |
| Tests for funnel plot asymmetry | yes | yes | no |
| Trim and fill method | yes | yes | no |
| Selection models | no | yes[2] | no |

### 8.2.4 Illustration of the metafor package

**Step 1. Specify the data**

- Firstly, a set of effect size estimates with their corresponding sampling variances have to be obtained (calculated by hand, from another software or from *metafor* package). It is needed that the observed outcomes $y_i$ and their corresponding sampling variances $v_i$ are given (or the square root of sampling variances $se_i$)

**Step 2. Specify a random effects model**

- A random effects meta-analysis is suggested to be calculated first. The aim is to calculate the $\tau^2$. By default, Restricted maximum-likelihood estimation is used. Other estimators can be set as well.

Code in blue:
res1<-rma(yi,vi, data=data)

where yi=observed outcome
vi=sampling variance
data is the dataset

Two important parameters that are calculated are the $\tau^2$ and the $I^2$ ($\tau^2$ is used later in meta-regression). The $I^2$ statistic shows how much of the total variability in the effect size estimates can be attributed to heterogeneity among the true effects (therefore if $\tau^2=0$ then the $I^2=0\%$).

**Step 3. Build the meta-regression (mixed-effects) model**

Study characteristics such as year, location etc. may explain at least part of the heterogeneity. The *metafor* package can handle categorical moderator variables with the appropriate dummy coding.

Let's assume that we want to test the moderator variables *year* and *location*, and test their influence.

Code in blue:
res2<-rma(yi,vi, mods=cbind(year, location), data=data)
or
res2<-rma(yi,vi, mods=~year+ location, data=data)

where yi=observed outcome
vi=sampling variance
data is the dataset
year=year of the study
location=location of the study

The model summary illustrates the estimated amount of residual heterogeneity equals to $\tau^2$ (different than the $\tau^2$ calculated earlier in the random effect meta-analysis model). In this way, we can estimate the total amount of heterogeneity that can be accounted for by accounting the moderator variables in the model.

The model summary provided the estimate, the standard error, the z-value, the p-valued and the confidence intervals for the constant term as well as for each moderator variable.

## Step 4. Funnel Plot for detecting publication bias

In funnel plots when carrying out meta-regression models, the horizontal axis shows the residuals, whilst the vertical axis shows the corresponding standard errors. A vertical line is drawn at zero with a pseudo confidence interval region given by **±1.96*Standard error** (shown in vertical axis).

Code in blue:
funnel(res2, main="Meta-regression model")

where *res2* the fitted meta-regression model (see above) and the argument *main* builds title of the graph

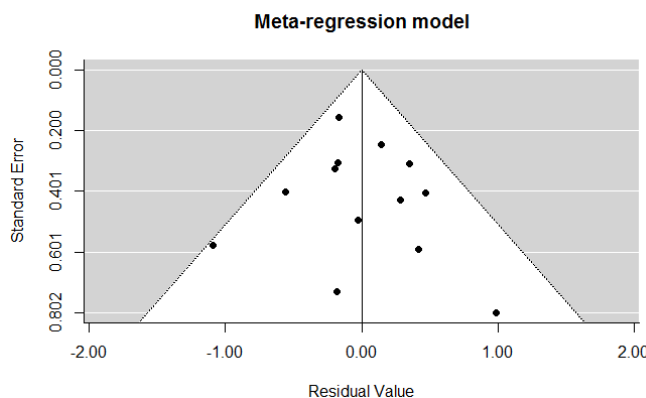Figure 8.2 figure shows an example of a funnel plot for meta-regression models.



**Figure 8.2** Example of funnel plot in meta-regression

## Step 5. Further tests for Funnel Plot asymmetry

Funnel plots are very useful to detect publication bias (e.g. unpublished studies with non-significant findings). In meta-regression models in particular, we can detect asymmetry in funnel plots by testing whether the residuals are related to their corresponding sampling variances, standard errors or even sample sizes (statistically significant results show a correlation).

A regression test is carried out. Code in blue:
regtest(res2, predictor="vi")

Where res2 is the fitted meta-regression model.

One can use:
"vi" is for sampling variance
"sei" is for standard error
"ni" is for sample size
"ninv" is for inverse of the sample size

## Step 6. Model fit statistics

By using the argument *fitstats()* we can calculate model fit statistics (e.g. loglikelihood, AIC, BIC, etc.). Lower values of AIC, BIC Note that these model statistics have sense only for model comparisons of <u>nested models</u> in the <u>same dataset</u>.

Code in blue:
*fitstats*(res2)

## 8.3    CONCLUSION ON META-ANALYSIS

A meta-analysis can help to combine the results from several studies, if these results are produced under comparable conditions. That means if the same method of data-analysis is applied or that the results can be transformed into the same type of outcome. Moreover, it should be clearly reported if the studies differed on important variables. In such a case a sub-group analyses or a meta-regression can indicate how these variables affect the results. It should be noted however, that for both kind of analyses a substantial number of studies is required. For a subgroup analysis there should be at least 10 studies per group (15 is better) and a meta-regression is worth fitting from 20 studies up-wards.

Whether or not it makes sense to summarize the results of several studies, can be tested by assessing the heterogeneity of the study-results. If this test is significant the results should only be summarized and published with great caution. To explain these differences a meta-regression or a subgroup analysis can be used.

# PART 3 – Economic efficiency of countermeasures

# 9 The E³ Calculator

## 9.1 INTRODUCTION

The countermeasures for which the analysis of safety effects has resulted in an estimated reduction of crash occurrence are submitted to an economic efficiency evaluation. A tool for Economic Efficiency Evaluation (E³) of road safety countermeasures will be implemented in the DSS, the methodology is however based on the Excel version of the E³ Calculator. This tool allows to combine the information about the effectiveness of a measure (i.e. the percentage of crashes or casualties prevented) with the costs of measures and the monetary value that is given to the avoidance of crashes and casualties. As input to the E3-calculator, the estimated costs for crashes and casualties of different severity have been collected from all European countries (see Deliverable D3.2 "Crash cost estimates for European countries").

The outcomes are the cost-effectiveness (i.e. the costs for preventing one crash or casualty) for different levels of severity: fatal, serious, light – and damage only crashes. In a cost benefit analysis, outcomes of different severity can be considered jointly by including a monetary valuation of these outcomes.
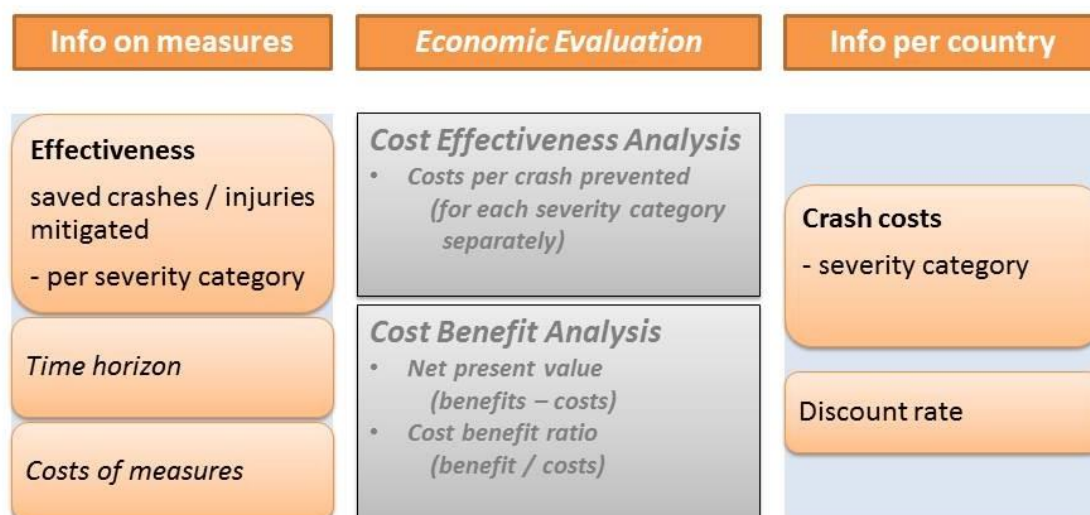


**Figure 9.1** Economic Efficiency Evaluation (E3) in SafetyCube

## 9.2 CRASH AND CASUALTY COSTS

In D3.2, an overview was presented of the components that should be included in crash cost estimates and how each cost component should be determined according to the international guidelines (e.g. Alfaro et al., 1994) and best practices (e.g. ECMT, 1998; Bickel et al, 2004). The component logic was applied to costs per crash and costs per casualty. Second, information on costs of crashes and costs of casualties was collected by means of a survey among all EU countries. Four severity levels were differentiated: fatal, serious injuries, slight injuries, and damage only (with the last category available for crash costs but not casualty costs). Third, for some countries not all information is available or costs are not calculated according to the international guidelines. In those cases, additionally to the country's own estimates (if available) comparable estimates according to

the standard procedure were provided by means of value transfer. In that way, an estimate for the total costs of crashes and casualties in the EU is provided as well.

## 9.3    E3-CALCULATOR

As input to the calculator for the Economic Efficiency Evaluation, the following is needed
- Measure costs
    - Initial costs
    - Annual costs
- Number of crashes / casualties prevented (for each level of severity)
    - Target crashes of countermeasure
    - % reduction
- Time horizon (period considered for E3)

On the basis of this input and the crash  or casualty costs, the calculator adds for each year within the time horizon the present value of all costs and benefits, resulting into the following outputs:
- Number of crashes / casualties prevented (per unit of implementation)
- Cost effectiveness: cost per prevented crash / casualty
    - Costs per prevented fatality / fatal crash
    - Costs per prevented severe injury / severe crash
    - Costs per prevented slight injury / light crash
    - Cost per prevented damage only crash (if applicable)
- Total benefits
- Cost benefit ratio (benefits/costs)
- Net effect (benefits – costs)

By default the SafetyCube analyses are conducted for the EU level and the crash costs used should be the EU-standardized values. If no measure costs are entered, the break-even costs are calculated: the costs of the measure at a benefit-cost ratio of 1. This indicates how much a measure could max-imally cost and still be cost-effective.
The principles underlying an economic efficiency analyses conducted by this E$^3$ calculator are de-scribed in D3.4 "Preliminary guidelines for priority setting between measures" and D3.5 "Guidelines for priority setting between measures". This section is complementary to those deliverables and focusses purely on the practical steps to conduct a benefit-cost analysis with the excel-calculator. The reader is strongly advised to read about the interpretation of the different criteria produced and the effect of parameters like the discount rate or the time horizon in D3.5.

## 9.4    OUTLINE OF THE ECONOMIC EFFICIENCY EVALUATION

A large part of the E3 analyses conducted in SafetyCube is based on existing studies. Often one study has the essential input (effectiveness estimate, target group) and it might provide additional information (side effects, effects on the penetration rate of a measure). These can be enterd into the E3 calculator next to the items listed above, but their provision is optional.

Even if the an E3 analysis is more or less exclusively based on one study, the results are adjusted in SafetyCube, by using the crash costs collected in WP3 and by up-dating all other costs by correcting them for inflation and to the price level of the country for which we are conducting the analysis.

### 9.4.1    Crash costs

One of the main features of cost benefit analyses (CBA) is that the outcome depends heavily on the assumed crash costs. Because of large differences in crash cost estimates between countries, this

means that a measure that is cost-effective in one country can be evaluated as "economically non-efficient" in another. Of course there is realistic variation in what a country can (is willing to) spend on road-safety and that should be rightfully reflected CBA's for different countries. However a lot of variation is also introduced by the estimation method of the crash costs. The most important factor here is the estimation of the human costs. Willingness to pay is the "state of the art", because it is in accordance with economic welfare theory, which is the basis of CBA. Other estimation methods like human capital lead to much lower estimates. In D3.2, we have therefore re-estimated the human costs (and other absent or "deviant" components -- using value transfer from the countries that have estimated the component in question). We suggested that partners use the countries' own official value but use the standard value (based on common estimation methodology) in the sensitivity analysis. For some countries, e.g. for Germany, it makes a big difference which crash costs are used.

### 9.4.2    Up-dating Measure costs

We will often use measure costs from studies, which might not be very recent and/or might come from another country. Because 100 Euro now buys less than it bought one 10 years ago, and because 100 Euro in Spain buy more than 100 Euro in Denmark, it is important to up-date all other costs (apart from the crash costs) to the 2015 price-level of the country for which the analysis is conducted. It requires some look-up from tables included in the E3 calculator and multiplication with two different values. Details are explained in Section 10.2.5.

### 9.4.3    Outcomes and further steps

The Excel tool conducts a cost benefit analysis and a cost effectiveness analysis. The results allow to observe costs and benefits of every single measure as well as to rank measures according to their economic efficiency, e.g. by comparing their benefit-to-cost ratios.

On the basis of the E3 analysis, a CBA synopsis is prepared. This synopsis is a two page PDF-document) that provides information with respect to all sources that have been used for the analysis, the assumptions that were made and the outcomes of the analyses. The steps to be undertaken for the E3 analyses are indicated in Figure 9.2.
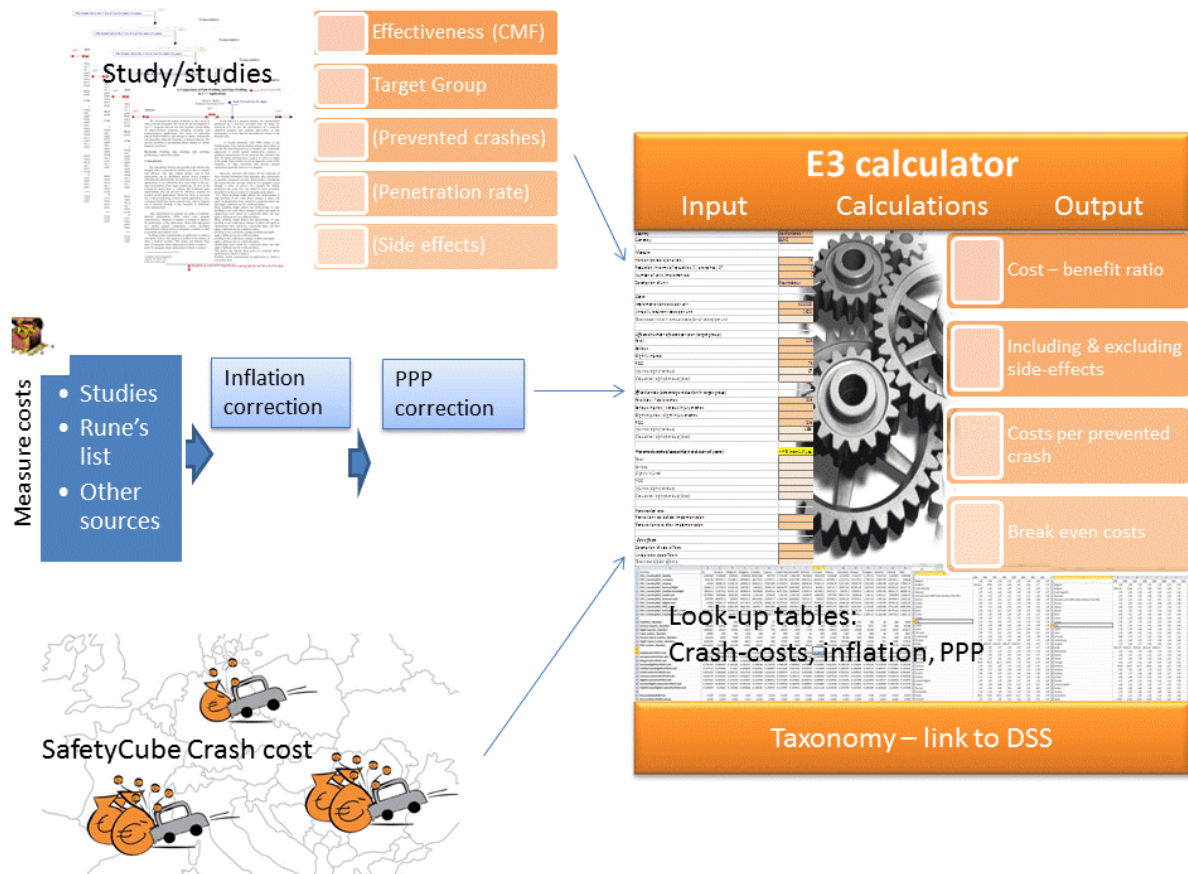
**Figure 9.2** Outline of SafetyCube Economic Efficiency Evaluation ($E^3$)

In following chapters we will describe the input that partners have to enter into the E3 calculator (Chapter 10), the calculations that are conducted with them (Chapter 11), and the output that is generated (Chapter 12).

# 10 Inputs to the E³ Calculator

## 10.1 COUNTRY

Enter the country for which the analysis is done. If most data come from one country (country example), fill in this country's name. Otherwise choose EU.

### 10.1.1 Country-example

If most of the information used for the analysis comes from one specific study, or from different studies but all in the same country, you should keep the analysis in the currency of that country.

| Country | Netherlands |
|---|---|
| Currency | EURO |

- Fill in country's name
- Fill in country's currency

Use also the crash costs from the same country (see Section 10.9.1).

### 10.1.1 General analysis

If the information comes from different studies, in particular if the effectiveness is estimated by means of a meta-analysis, choose EU rather than a single country. Also do this when the information comes from a non-European country. Use the EU-standard value for crash costs (see Section 10.9).

| Country | EU |
|---|---|
| Currency | EURO |

## 10.2 MEASURES AND MEASURE COSTS

| Measure | | |
|---|---|---|
| Horizon (period of analysis) | 25 | |
| Reduction in terms of casualties (1) or crashes (2)? | 2 | **Crashes** |
| Number of units implemented | 1 | |
| Description of unit | Roundabout | |

### 10.2.1 Horizon

The life-time of a countermeasure and therefore the time over which the analysis is conducted. For infrastructure measures this can be up to 25 or 30 years. For other measures, e.g. awareness raising campaigns or educational activities the horizon is typically much shorter. In many cases 1 year can be chosen. 1 year is the smallest unit of analysis in the calculator. This way, the implementation costs of a measure are weighed against the benefits of prevented crashes over the whole lifetime of the measure.

### 10.2.2 Crashes vs. casualties

The countermeasures are evaluated in terms of their ability to prevent crashes or casualties. Some countermeasures are mostly meant to prevent crashes (e.g. pedestrian detection, alcohol checks, rumble strips) while others are meant to mitigate the consequences of crashes (e.g. seatbelt reminders, guard rails) and many are expected to do both (e.g. ABS, speed reduction). Similarly, an E3-analysis can concern the number of *crashes* prevented or the number of *casualties* prevented.

If the analysis is done in terms of casualties, there is nevertheless the option to include effectiveness and target-groups size (see Section 10.3.2) for PDO crashes. Choose zero effectiveness (i.e. 0% reduction) for PDO crashes, if the measure is purely directed to mitigating consequences.

In the remainder of this document we will refer mainly to the reduction of crashes. Unless mentioned explicitly, the analysis in terms of casualties is however conducted strictly analogously.

### 10.2.3  Implementation units

The $E^3$ calculator is based on the assumption that countermeasures will be evaluated in terms of single implemented units. A unit of intervention can be:

- A single infrastructure element, e.g. a roundabout or a pedestrian bridge
- A road stretch where an infrastructural or enforcement measures is implemented, e.g. guardrails per km, road lightning per km or section speed control per km
- A vehicle that is equipped with a safety system , e.g. airbags or an ISA system
- An hour of enforcement on a road section, e.g. speed cameras or drink driving surveillance

In general it is recommended to choose a single implemented unit (e.g. 1 intersection, 1 km of road) as the unit of analysis. This enables presenting results in a meaningful way and furthermore allows to compare measures with each other. If the analysis is based on a particular study, and it is unclear what the unit of analysis was, it is also possible to choose the subject of the study as one unit. An example for this is the CBA by Yannis et al. (2005), where traffic calming of one area is analysed. However this can reduce the possibility to transfer results to other settings afterwards as the units of the measure might not be unequivocally understandable.

As a standard, for the DSS, we evaluate the **cost effectiveness of one unit of intervention**.
  ➔  The "number of units implemented" should therefore by default be 1.

If the information is based on a study you can enter a higher number. Many studies calculate a national effect (e.g. how many crashes could be saved if all vehicles were equipped with a particular measure?). In such a case the target-group (number of affected crashes) can be taken from the national data. The "number of units implemented" should in this case be the size of the country's vehicle fleet.
  ➔  The "number of units implemented" has to match the target group size.

### 10.2.4  Measure costs

| Costs | | | |
|---|---|---|---|
| Implementation costs per unit | 500 000 | | |
| Annually recurrent costs per unit | 2 500 | | |
| *Total costs (initial + annual costs for all years) per unit* | | | |

For the measure costs, you have to differentiate between implementation costs and maintenance costs. **One-time investment costs** are only paid once, before the measure can produce any benefits (**year 0**) while the **recurrent costs** are paid regularly afterwards. Please indicate them **on a yearly basis**. All costs should be converted to the same price level, in the SafetyCube project this is 2015 (see also Section 10.2.5)

If the figures you have are not differentiated between implementation costs and annually recurrent costs, use the field "*Total costs (implementation + annually recurrent costs*".

### 10.2.5  Up-dating measure costs

We will often use measure costs from studies, which might not be very recent and/or might come from another country. It is therefore important that you up-date the measure costs to the year 2015 and to the country of your analysis.

*Correcting for inflation*
Because 100 Euro buys you less than it bought you 10 years ago, costs from earlier years have to be up-dated to the **2015 price-level**. Take Sheet ***"Inflation conversion table"***, select the year in which the costs were established in the top row, select the country in which they were established in the left column. The cell where these two meet indicates the factor by which you have to multiply the measure costs. The table includes values starting from 1995 for each country. Check inflation tables on the website of Eurostat ([http://ec.europa.eu/eurostat/data/database](http://ec.europa.eu/eurostat/data/database)) in the extreme case that only information from before 1995 is available.

*Correcting for price-level*
Because 100 Euro in Norway buys you less than it does in Greece, the price of a particular measure in Norway will most likely be higher than the price for the same measure in Greece. Therefore measure prices have to be up-dated to the price level of the country for which you are conducting the analysis (or to the mean EU-price level, if you are doing a general analysis).
Take Sheet "PPP conversion table" select the country in which the costs were established in the left column, select the country to which you want to transfer in the top row. The cell where these two meet indicates the factor by which you have to multiply the measure costs to correct for differences in price-level and in currency. Watch out: PPP values are expressed in local currency, so if you want for instance to convert a Norwegian value to a UK value, you need to apply the PPP-value 'from Norway to EU' to the price in NOK in order to obtain a value in GBP. Please note that the used currency for the EU-level is the euro.

**Important: correct first for inflation then for price-level.**

## 10.3   TARGET GROUP

The effectiveness of a measure is calculated on the basis of size of the target group per unit of intervention and the effectiveness of a measure: percentage reduction based on the crash modification factor (CMF). As an example, a CMF of 0.8 indicates an effectiveness of 20%. The target group is the number of crashes to which this reduction has to be applied annually.

### 10.3.1  Severity classes

Target group and effectiveness should ideally be determined separately for each severity class. The four severity classes considered in the E3 calculator are based on the categories for which we have crash costs: fatal, serious, slight, PDO (property damage only). Ideally you should have a separate estimate for each severity category, for the target group as well as for the effectiveness. Often this is not the case. You might have estimates only for joint categories or for some categories you might have no estimates at all.

Many studies are conducted on slight, serious, and fatal crashes jointly. In this case, leave the fields for the separate categories empty and fill in those for the joint categories.

***Important: For each severity category, data should be entered only once: either in the separate categories, or in the joint ones --- BUT NOT BOTH!*** (Otherwise those casualties will be counted double in the analysis).

| Affected number of cases per year (target group) | |
|---|---|
| Fatal | 0.01 |
| Serious | |
| Slightly injured | |
| PDO | 14 |
| *Injuries (slight/serious)* | *3* |
| *Casualties (slight/serious/fatal)* | |

Correct!

| Affected number of cases per year (target group) | |
|---|---|
| Fatal | 0.01 |
| Serious | 0.05 |
| Slightly injured | 2.95 |
| PDO | 14 |
| *Injuries (slight/serious)* | *3* |
| *Casualties (slight/serious/fatal)* | *3.01* |

***Wrong!***

### 10.3.2 The size of the target group

The target group of a road safety measure is the number of cases affected per year. This means those crashes that could potentially be prevented by a particular measure. A study that discusses the effectiveness of a certain road safety measure should also define the target group the effectiveness estimate relates to. For example: if a study estimates that head-tail crashes can be reduced by 15% for vehicles with AEB, the target group is the number of head-tail crashes. Another study might have estimated the reduction among all crashes, but that would result in a much lower percentage. Therefore it is important to match the target group to the way that the effectiveness is estimated.

The size of the target group also has to match the number of implemented units. If the number of implemented unit is 1 vehicle, the target group size has to be the expected number of crashes per vehicle (a very small number luckily). If you want to enter the national number of crashes as target group size, use the size of the country's vehicle fleet size as number of implemented units.

For infrastructure measures the target group is often simply the number of crashes that happen at a particular piece of infrastructure. The size of the target group should not be calculated for a specific location, but ideally be based on national statistics (e.g. the average number of crashes at a 4 legged crossing, rather than last year's number of crashes at a crossing that was turned into a roundabout).

### 10.3.3 Empty categories

For the target group size we have foreseen that you might have data on injury crashes, but not on PDO crashes. We also have foreseen the case that you have a target group size in terms of fatal crashes but not for injury crashes. In both cases, a small calculator (see below) makes a suggestion based on the relation between fatal, serious, slight, and PDO crashes in the national data of the selected country.

| Suggested values if you have no target group of injury or damage only crashes | | |
|---|---|---|
| | **Analysis on casualties** | **Analysis on crashes** |
| PDO | **9.990** | **13.292** |
| Injuries (slight/se | **4.258** | **3.444** |

For these estimates the calculator applies the relation between fatalities, injuries, and PDO in the country's crash cost data to the numbers already entered into the template (see Section 10.3.1). In the example shown, the calculator estimates that given the number of casualties entered into the template already, one would expect 3 injuries and 13 PDO crashes. Note that the suggested number of injuries is based on the entered number of fatalities. You will see this suggestion even if you have entered injury data yourself. If you have your own data, you should *not* overrule it with a value suggested by the calculator. The PDO suggestion is based on whatever is filled in already. There is a hierarchy implemented so that the calculator checks cells "injured (serious/slightly)", "casualties (fatal/serious/slightly)", "slight injuries", "serious injuries", "fatalities" in that order. If the analysis is done by casualty numbers, the estimate would be somewhat different, so make sure you choose the correct column.

## 10.4   PERCENTAGE REDUCTION OF CRASHES

The effectiveness (E) of a measure is defined as the percentage reduction (PR) in the target crashes when implementing the measure. It is generally measured by a crash modification factor (CMF)

E = PR= 100* (1-CMF)

Of course the effect of a measure can also be an increase in crash numbers. A CMF of 1.5 indicates an increase of crashes by 50%. In the E3 tool an increase by 50% would be entered as -50% (because the percentages normally indicate the *reduction* of crashes).

The effect of a measure could however differ for crashes of different severity. It is therefore advisable that whenever possible the analysis takes into account separate effectiveness estimates for different severity levels of crashes.

Sometimes we will not have separate effectiveness estimates for each severity class. If the measure is thought to address crash occurrence and it is thought to have no particular effect on crash severity, use the CMF's from the nearest known severity category (e.g. CMF slight injuries for damage only crashes) can be used.

As a rule of the thumb, you should NOT assume equal effectiveness across severity classes if
- Effectiveness estimates for two severity classes differ strongly (e.g. for median guard-rails the effectiveness for severe and slight injury crashes differs strongly. Therefore one should certainly *not* assume that the effectiveness for possibly unknown severity levels would be the same as either that for sever or that for slight injury crashes.)
- Effectiveness concerns largely the mitigation of consequences (Measures that mitigate consequences of crashes reduce the severity of injuries, implying e.g. that serious injuries are prevented but slight injuries are increased.)

If effect on severity is found, use this effect to adjust CMFs for unknown severity categories.

➔ If you have reasons to think that your effectiveness estimate can also apply to other severity classes, you can copy it to those fields where you think it might very well apply as well.

➔ If you think that a measure does not affect a particular severity class (e.g. seatbelts might not affect the number of PDO crashes) enter 0%.

## 10.5 NUMBER OF PREVENTED CRASHES/CASUALTIES: AN ALTERNATIVE WAY TO ENTER CRASH REDUCTION

In some cases, the percentage annual reduction is not known (nor is the target group), but it is known how many crashes have been prevented by a measure over the time-span of the analysis. In this case there is an alternative way of entering this information:

| Prevented crashes/casualties (total over all years) | -- fill in only if you do not have target group & effectivenes | | |
|---|---|---|---|
| Fatal | | | |
| Serious | | | |
| Slightly injured | | | |
| PDO | | | |
| Injuries (slight/serious) | | | |
| Casualties (slight/serious/fatal) | | | |

In the remainder of the analysis, the numbers entered here are divided by the horizon (number of years spanned by analysis) to estimate the annual number of prevented crashes.

The same principles apply as for entering target groups and effectiveness:
- You can use the calculator for suggestions of injury and/or PDO crashes if you do not have data for these categories -- be aware though that adopting the values from the calculator is based on the ***assumption that the effectiveness is the same for all severity categories***.
- For each severity category, data should be entered only once: either in the separate categories, or in the joint ones --- BUT NOT BOTH! (otherwise those casualties will be counted double in the analysis).

| | |
|---|---|
| Fatal | 0.025 |
| Serious | |
| Slightly injured | |
| PDO | 35 |
| Injuries (slight/serious) | 7.5 |
| Casualties (slight/serious/fatal) | |

Correct!

| | |
|---|---|
| Fatal | 0.025 |
| Serious | 0.5 |
| Slightly injured | 7 |
| PDO | 35 |
| Injuries (slight/serious) | 7.5 |
| Casualties (slight/serious/fatal) | 7.575 |

**Wrong!**

## 10.6 WHAT IF INSUFFICIENT INFORMATION IS AVAILABLE ABOUT THE TARGET GROUP?

For the Calculations and the Output the "Affected number of cases per year (target group)" (see Section 10.3) is an essential input. If no information is available about the affected number of cases, you should find values for the 'number of prevented crashes' (see Section 10.5).

The following example illustrates that we need in any case to have some information about the targeted or prevented number of crashes.
Assume that you find a good meta-analysis about the effects of constructing fences along roads in order to prevent collisions with animals. The best estimate in this meta-analysis is a 15% reduction of injury crashes. However it appears that the only included studies are Australian and also that kangaroos were present in most of the reported animal collisions. In such case, it is quite obvious that one cannot just transfer the effectiveness estimate to other countries. It also becomes clear that CMF's in themselves are only meaningful if they are applied to relevant target crash data, in the present case to crashes with animal populations that are believed to be comparable with the ones that

occur in the areas where the studied data are from. A proper specification of the target group of crashes is thus a necessary part of any cost benefit analysis.

Unfortunately, information on target crashes is often not easy to retrieve. For instance in meta-analyses, which are the preferred sources of effect estimates to use for CBA's, typically no specific values for the number of target crashes are presented.
There are a few possible solutions for this:
1)  After having selected the appropriate CMF, check whether relevant crash data are available for the selected country (or for Europe). In many cases this will be possible, it just might require some extra effort. For example, if a measure is believed to target children from 0 to 7, look for national crash data in order to retrieve information on crashes with children.
2)  If it is not possible to find appropriate data that are applicable to the concerned country, a possible workaround is to look for some relevant data elsewhere, e.g. in one of the papers where the meta-analysis is based upon (in principle they should contain some descriptive information about the used data) or in one or another report that contains relevant crash information for the measure under study.

### 10.7   PENETRATION RATE

Some measures do not prevent crashes directly but do so by increasing the likelihood of another measure (or by decreasing the likelihood of a risk factor). The percentage of users that apply the measure, the penetration rate can be entered in the template and is used together with the effectiveness of the measure and the target group to determine the annual number of prevented crashes or casualties.

| Penetration rate | |
| --- | --- |
| Pentration rate before implementation | |
| Penetration rate after implementation | |

*Example Motorcycle helmet*
Suppose that a (set of) measures results in an increase in the use of motorcycle helmets in a certain country and that the penetration rate (i.e. the percentage of motorcyclists wearing a helmet) increased from 80% to 90%. Moreover, in that country, on average 120 motorcyclists die every year as a result of a crash. According to Liu et al (Liu et al., 2008), motorcycle helmets reduce the risk of death with 42%. All this amounts to a prevention of 8 fatalities annually.

The target group for fatal crashes is 120, the effectiveness for fatal crashes is 42%, the penetration rate before is 80% and the penetration rate after is 90%.

### 10.8   SIDE EFFECTS

In a benefit-cost analysis of countermeasures in principle all relevant welfare impacts need to be taken into account, including impacts on travel time, environment, health, etc. Studies have shown that for several measures these impacts may have substantial influence on the outcomes of a benefit-cost analysis. However, in many cases it is quite complex to estimate these impacts and it is beyond the scope of the SafetyCube project to estimate these impacts.

One alternative is to included side impacts qualitatively, that is by (briefly) describing them (for now in the comment field). Also in some existing studies the impacts may have been estimated. These can be entered into the E3 calculator.

| Side effects | |
|---|---|
| Description of side effects | |
| Annual benefit side effects | |
| *Total benefit of side effects* | |

Side effects can either be included as annual benefits or as total benefits over the time-span of the whole analysis (-> horizon).

- Side effects are considered to be benefits. If they are costs, enter them as a negative number.
- If different side effects are considered add all benefits up (taking into account whether they are costs or benefits).
- Up-date the values of the side effects to 2015 and your country's price level (use the same procedure as described for measure costs - see Section 10.2.5).

## 10.9  INPUT FROM SAFETYCUBE CRASH COST ESTIMATES

At the lower side of the input page, there is a section where crash costs and the total number of crashes / casualties per severity class have to be entered.

- Open worksheet CrashCosts.
- Select the country that you are analyzing in the top row (or "EU" if you are doing a general analysis). Copy all the completed values in this column including the country name.
- Paste "text only" into the orange coloured fields in the E3 template.
- Under the last input row, there is a formula (selecting the right type of Discount Rate) please do not overwrite it.

### 10.9.1  Crash costs

For D3.2, we have collected crash costs from all European countries. Not all countries supplied values for each unit. For instance, only some countries have a unit cost for PDO crashes. For those countries that do not have an estimate, we have taken the average of the countries that do have an estimate. Moreover, the method of estimating costs is not the same in all countries. For those countries that have a deviant method for (part of) their costs, we have made an estimate, what their crash costs would be when applying the recommended standard methodology. For each country we have therefore two types of crash costs:

- The countries self-reported crash costs (+ estimates for those components that the country did not include)
- The country's crash costs when applying the common methodology

In the worksheet CrashCosts all countries are first listed with their own crash costs (simply titled with the country's name) and then the estimates applying common methodology.

| Country | Austria_CommonMethod | Belgium_CommonMethod | Bulgaria_CommonMethod |
|---|---|---|---|
| CPU_Country2015_fatality | 3253022.326 | 2726369.197 | 4001889.308 |
| CPU_Country2015_serinjury | 425370.5588 | 337511.827 | 534556.0818 |
| CPU_Country2015_slinjury | 35969.1343 | 32681.45818 | 48350.40626 |
| CPU_Country2015_SeriousSlight | 82056 | 68759 | 105895 |

| Copy from EXCEL SCWP3 CrashCosts for E3.xlsx | |
|---|---|
| Country | Netherlands |
| CPU_Country2015_fatality | 2717621.303 |
| CPU_Country2015_serinjury | 292002.5034 |
| CPU_Country2015_slinjury | 6542.999786 |
| CPU_Country2015_SeriousSlight | 23860.50392 |
| CPU_Country2015_FatalSeriousSlight | 30171.54054 |
| CPU_Country2015_FatalCrash | 3065353.535 |
| CPU_Country2015_SeriousCrash | 336163.7031 |
| CPU_Country2015_SlightCrash | 37069.20488 |
| CPU_Country2015_PDO_Crash | 3663.039244 |
| CPU_Country2015_SeriousSlight_Crash | 58049.18409 |
| CPU_Country2015_FatalSeriousSlight_Crash | 66754.65522 |
| | |
| Fatalities_Number | 720 |
| Serious injuries_Number | 18600 |
| Slight injuries_Number | 288000 |
| Fatal crashes_Number | 669 |
| Serious injury crashes_Number | 16164 |
| Slight injury crashes_Number | 214273 |
| PDO crashes_Number | 1021000 |
| | |
| FatalitiesPerPDOCrash | 0.000705191 |
| SerinjuriesPerPDOCrash | 0.018217434 |
| SlinjuriesPerPDOCrash | 0.282076396 |
| SeriousSlightPerPDOCrash | 0.30029383 |
| FatalSeriousSlightPerPDOCrash | 0.300999021 |
| FatalCrashesPerPDOCrash | 0.00065524 |
| SeriousCrashsesPerPDOCrash | 0.015831538 |
| SlightCrashesPerPDOCrash | 0.209865818 |
| SeriousSlightCrashsesPerPDOCrash | 0.225697356 |
| FatalSeriousSlightCrashesPerPDOCrash | 0.226352595 |
| | |
| Discount rate - risk free | 3% |
| Discount rate - premium | 5% |
| Discount rate selected | 3% |

Use the countries own estimates in the first place. Check how the cost estimates differ when applying common methodology. If the difference is big, consider entering the common methodology values in the yellow fields, next to the costs to use them in the sensitivity analysis (see Section 12.3)

The table below indicates for each country, the relation between the common methodology costs and those reported by the country. A number smaller than one (in blue) indicates that the country estimates are smaller than the common-methodology estimates. In a CBA, this means that measures will be more likely to be judged as "economically *not* efficient" when using the country's own estimate. A red field (ratio >1) indicates that country-estimates are larger than common methodology estimates. #VALUE indicates that the country did not supply its own estimate (these values are filled in by SafetyCube, based on the average of those countries that applied common methodology).

| Ratio: Countrie's own value / common meth. est. value | | | | | | | |
|---|---|---|---|---|---|---|---|
| Country | Fatal crashes | Fatalities | PDO crashes | Serious injuries | Serious injury crashes | Slight injuries | Slight injury crashes |
| Austria | 0.93 | 1.00 | 0.82 | 0.96 | 0.95 | 0.80 | 0.76 |
| Belgium | #VALUE! | 0.80 | #VALUE! | 0.99 | #VALUE! | 0.65 | #VALUE! |
| Bulgaria | 0.55 | 0.60 | #VALUE! | 0.73 | 0.81 | 2.09 | 2.00 |
| Croatia | #VALUE! | 0.82 | #VALUE! | 0.86 | #VALUE! | 0.69 | #VALUE! |
| Cyprus | #VALUE! | 0.45 | #VALUE! | 0.45 | #VALUE! | 0.36 | #VALUE! |
| Czech Rep | #VALUE! | 0.44 | 0.96 | 0.62 | #VALUE! | 0.63 | #VALUE! |
| Denmark | #VALUE! | 0.80 | #VALUE! | 1.11 | #VALUE! | 1.73 | #VALUE! |
| Estonia | #VALUE! | 1.24 | #VALUE! | 3.16 | #VALUE! | 1.34 | #VALUE! |
| Finland | 0.90 | 1.00 | 0.87 | 1.00 | #VALUE! | 0.98 | #VALUE! |
| France | #VALUE! | 1.30 | 1.14 | 1.21 | #VALUE! | 0.51 | #VALUE! |
| Germany | 0.02 | 0.51 | 1.05 | 0.38 | 0.06 | 0.18 | 0.35 |
| Greece | 0.96 | 1.05 | 0.77 | 0.77 | 0.69 | 1.63 | 1.45 |
| Hungary | #VALUE! | 0.95 | 1.44 | 1.65 | #VALUE! | 0.20 | #VALUE! |
| Iceland | 1.04 | 0.87 | 2.09 | 0.86 | 1.00 | 0.83 | 1.52 |
| Ireland | 1.04 | 1.02 | 0.73 | 1.01 | 1.26 | 0.97 | 1.04 |
| Italy | 0.00 | 0.62 | #VALUE! | 0.54 | 0.03 | 0.52 | 0.25 |
| Latvia | #VALUE! | 0.50 | 1.66 | 0.09 | #VALUE! | 0.01 | #VALUE! |
| Lithuania | #VALUE! | 0.44 | #VALUE! | 0.30 | #VALUE! | #VALUE! | #VALUE! |
| Luxembou | #VALUE! | #VALUE! | #VALUE! | #VALUE! | #VALUE! | #VALUE! | #VALUE! |
| Malta | #VALUE! | 0.75 | #VALUE! | 0.79 | #VALUE! | 0.59 | #VALUE! |
| Netherlan | #VALUE! | 1.00 | 1.00 | 1.00 | #VALUE! | 0.24 | #VALUE! |
| Norway | 1.00 | 1.00 | 0.85 | 1.00 | 0.98 | 0.99 | 0.95 |
| Poland | #VALUE! | 0.36 | 4.27 | 3.22 | #VALUE! | 0.42 | #VALUE! |
| Portugal | 0.36 | 0.37 | #VALUE! | 0.45 | 0.48 | 1.29 | 1.28 |
| Romania | #VALUE! | #VALUE! | #VALUE! | #VALUE! | #VALUE! | #VALUE! | #VALUE! |
| Serbia | 0.27 | #VALUE! | #VALUE! | #VALUE! | 0.27 | #VALUE! | 0.23 |
| Slovakia | #VALUE! | 0.30 | 1.06 | 0.42 | #VALUE! | 0.61 | #VALUE! |
| Slovenia | 0.02 | 1.05 | 1.01 | 0.99 | 0.05 | 0.85 | 0.32 |
| Spain | #VALUE! | 0.74 | #VALUE! | 0.91 | #VALUE! | 0.40 | #VALUE! |
| Sweden | #VALUE! | 1.02 | 0.81 | 1.36 | #VALUE! | 1.13 | #VALUE! |
| Switzerlar | #VALUE! | 0.99 | #VALUE! | 0.96 | #VALUE! | 0.54 | #VALUE! |
| UK | 0.88 | 0.88 | 0.69 | 0.86 | 0.85 | 0.62 | 0.71 |

**Figure 10.1** Ratio of crash and casualty costs supplied by each country and SafetyCube estimation based on common methodology. Blue fields indicate that country's estimate is smaller than common methodology estimate, red fields vice versa.

### 10.9.2 Number of crashes and casualties

If the number of crashes or casualties is known for some categories (e.g. all fatal crashes) but not for others (e.g. PDO crashes), SafetyCube suggests values on the basis of the categories that were filled in. This is based on the number of crashes reported in the crash cost collection (D3.2). We apply the relation observed in the crash cost reporting to the target group numbers that are filled in by the user. For example, in the Netherlands in 2015, there were 720 fatalities and 1021000 PDO crashes. So, for every fatality there are 1418 crashes with property damage only. The E3 calculator makes use of these relations to estimate the number of crashes / casualties for those categories for which they are not available.

Next to the input fields for the target group as well as for the number of prevented crashes (to be used alternatively to target group and effectiveness), we have implemented a small calculator, that uses the information that you copy below to suggest values for injury crashes (if you only have fatal crashes) and/or PDO crashes (on the basis of any other filled in category). See Section 10.3.1.

While the crash cost collection in WP3 is a convenient source for the estimated number of casualties and crashes. Especially with respect to crash numbers these are not trivial to derive from national statistics (in particular PDO crashes). It must be indicated though that the numbers are not always very recent. You can overwrite this information if you for instance wish to use more recent figures. Note that the absolute numbers are not relevant for the E3 analysis, but the relation between the numbers. Note also that the relation in the target group might be different from that in the national statistics. If the measure addresses crashes that are very severe, the proportion of severe crashes will be much higher than in the national statistics. It is always better to use data specifically collected for the measure in question.

# 11 Calculations

This section gives back-ground information on how the output is generated. Although not strictly speaking necessary for conducting an E3 analysis, it helps to understand how the output is generated.

## 11.1 CALCULATIONS PER YEAR

### 11.1.1 Horizon

In the second sheet of the E$^3$ calculator, the costs and benefits are calculated per year. The horizon is entered by user (between 1 and 50 years). Line 4 is 1 for each year within the horizon and 0 otherwise. By multiplying all costs and benefits with this value, only those values that fall within the horizon are added up.

### 11.1.2 Costs

The costs (split up by implementation costs in year 0 and exploitation costs in years 1-30) are first given in actual values (this means that the annual running costs entered in the input page is simply repeated each year). This is done for measure costs as well costs of side-effects.

The costs are then brought to present value using the discount rate.

$$present\ value = \frac{actual\ value}{(1 + discount\ rate)^{year}}$$

| Year | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Horizon | | 1 | 1 | 1 | 1 | 1 |
| | | | | | | |
| *Costs* | | | | | | |
| *Actual values* | | | | | | |
| One-time investment costs | 500 000 | | | | | |
| Recurrent costs | | 2 500 | 2 500 | 2 500 | 2 500 | 2 500 |
| Side-effects | | - | - | - | - | - |
| | | | | | | |
| | | | | | | |
| *Present values* | | | | | | |
| Investment costs | 500 000 | | | | | |
| Recurrent costs | | 2 439.02 | 2 379.54 | 2 321.50 | 2 264.88 | 2 209.64 |
| Side-effects | | - | - | - | - | - |
| | | | | | | |

### 11.1.3 Benefits

The benefits are first reported in terms of crashes saved per year. Note that the severity categories should be *exhaustive* and *non-overlapping* (e.g. if there is no value for PDO, than the costs of the PDO crashes will not be taken into account. If there are, for instance, casualties calculated for "Fatal/serious/slight", but also for "Fatal", "Slight", and "Serious", then those categories will be counted double.

Prevented crashes are calculated in 3 alternative ways (depending on the input):

- TargetGroup * Effectiveness
- TotalSavedCrashes / Horizon
- Based on penetration rate, target group, and effectiveness (see formula in Appendix A)

| Prevented casualties/crashes | | | | | | |
|---|---|---|---|---|---|---|
| Fatalities | | 0.0050 | 0.01 | 0.01 | 0.01 | 0.01 |
| Serious injuries | | - | - | - | - | - |
| Slight injuries | | - | - | - | - | - |
| PDO | | 7.50 | 7.50 | 7.50 | 7.50 | 7.50 |
| Serious & slight | | 1.70 | 1.70 | 1.70 | 1.70 | 1.70 |
| Fatal / serious / slight | | - | - | - | - | - |

To calculate the "actual value" of the **benefits,** the prevented crashes are multiplied with the unit costs entered into the input sheet.

| Benefits | | | | | | |
|---|---|---|---|---|---|---|
| Actual values | | | | | | |
| Fatalities | | 15 327 | 15 327 | 15 327 | 15 327 | 15 327 |
| Serious injuries | | - | - | - | - | - |
| Slight injuries | | - | - | - | - | - |
| PDO | | 27 473 | 27 473 | 27 473 | 27 473 | 27 473 |
| Serious & slight | | 98 684 | 98 684 | 98 684 | 98 684 | 98 684 |
| Fatal / serious / slight | | - | - | - | - | - |
| Sum | | 141 483 | 141 483 | 141 483 | 141 483 | 141 483 |

The sum of the actual value of the benefits is transferred into Present value by applying the same formula as for the costs (see Section 11.1.2).

| Benefits Present values (Sum) | | 138 032 | 134 666 | 131 381 | 128 177 | 125 051 |
|---|---|---|---|---|---|---|

# 12 Output

## 12.1 COST BENEFIT ANALYSIS

In the output sheets the results are summed up over the years within the horizon. All costs are reported in present values (i.e. corrected by discount rate) and in the selected country's currency.

The costs are reported separately for one-time investments, recurrent costs throughout the lifetime of the measure. Because, the inclusion of side-effects is not systematic but depends on whether studies have included them, the total costs are given twice: with and without including side-effects. This way the user can still compare results of studies with side-effects to those without side-effects

The benefits are given jointly for all severity classes.

The measures of socio economic return are the net present value and the benefit-to-cost ratio (BCR).

*Net present value = Present value benefits –Present value costs*

*Benefit-to-cost ratio = Present value benefits / present value costs*

A **benefit-to-cost ratio (BCR) > 1** indicates that a measure is economically efficient. A Net present value (NPV) > 0 indicates the same. The two measures can lead to different rankings (please consult D3.4, Chapter 3), but the basic categorization of economically efficient vs. economically not efficient will always be the same for both types of results.

If no measure costs have been filled in, no BCR and NPV can be calculated. The **break-even costs** can, however, still be estimated. They indicate the maximal costs that one unit of a measure can have to still be economically efficient.

| COST-BENEFIT ANALYSIS | |
|---|---:|
| | |
| *Costs (present values)* | |
| One-time investment costs | 500 000 |
| Recurrent costs | 44 879 |
| Total costs | 544 879 |
| | |
| | |
| *Benefits* | |
| Prevented Crashes | 658 382 |
| Side effects | 19 512 |
| Total benefits including side-effects | 677 894 |
| | |
| *Socio-economic return excluding side-effects* | |
| Net present value | 113 503 |
| Benefit-to-cost ratio | 1.2 |
| | |
| *Socio-economic return including side-effects* | |
| Net present value | 133 015 |
| Benefit-to-cost ratio | 1.2 |
| | |
| *Break-even cost for measure (per unit)* | 677 894 |

## 12.2 COST-EFFECTIVENESS ANALYSIS

Cost-effectiveness analysis relates the measure costs to the effect of the measures (here, prevented crashes) but not to the monetary valuation of these effects (no benefits).

| COST-EFFECTIVENESS ANALYSIS | | |
|---|---|---|
| | | |
| *Prevented crashes* | | |
| Fatal | 0.1250 | |
| Serious | 0.0 | |
| Slight | 0.0 | |
| PDO | 187.5 | |
| Serious & slight | 42.5 | |
| Fatal / serious / slight | 0.0 | |
| | | |
| *Costs per prevented crash* | | |
| Fatal | 4 368 488 | EURO |
| Serious | #DIV/0! | EURO |
| Slight | #DIV/0! | EURO |
| PDO | 2 912 | EURO |
| Serious & slight | 12 848 | EURO |
| Fatal / serious / slight | #DIV/0! | EURO |

## 12.3 SENSITIVITY ANALYSIS

A sensitivity analysis should be included in each E3 analysis. In the table below, it is indicated for each input variable whether a lower value would lead to a higher BC-ratio (green arrow up) or a lower one (red arrow down). The effect of a higher input value is of course the opposite.

In the table below, a green arrow upwards (↑) indicate that a value lower/higher than the estimate makes is *more likely* that the measure is evaluated as being economically efficient. A red arrow downwards (↓), indicates that a lower/higher value makes it less likely that the measure is evaluated as being economically efficient.

| *Costs* | | Lower value | Higher value |
|---|---|---|---|
| Implementation costs per unit | | ↑ | ↓ |
| Annually recurrent costs per unit | | ↑ | ↓ |
| *Total costs (initial + annual costs for all years) per unit* | | ↑ | ↓ |
| | | | |
| *Affected number of cases per year (target group)* | | | |
| Fatal | | ↓ | ↑ |
| Serious | | ↓ | ↑ |
| Slightly injured | | ↓ | ↑ |
| PDO | | ↓ | ↑ |
| *Injuries (slight/serious)* | | ↓ | ↑ |
| *Casualties (slight/serious/fatal)* | | ↓ | ↑ |
| | | | |
| *Effectiveness (percentage reduction in target group)* | | | |
| Fatalities / fatal crashes | | ↓ | ↑ |
| Serious injuries / serious injury crashes | | ↓ | ↑ |
| Slight injuries / slight injury crashes | | ↓ | ↑ |
| PDO | | ↓ | ↑ |
| *Injuries (slight/serious)* | | ↓ | ↑ |

| | | | |
|---|---|---|---|
| *Casualties (slight/serious/fatal)* | | ↓ | ↑ |
| | | | |
| ***Prevented crashes/casualties (total over all years)*** | | | |
| *Fatal* | | ↓ | ↑ |
| *Serious* | | ↓ | ↑ |
| *Slightly injured* | | ↓ | ↑ |
| *PDO* | | ↓ | ↑ |
| *Injuries (slight/serious)* | | ↓ | ↑ |
| *Casualties (slight/serious/fatal)* | | ↓ | ↑ |
| | | | |
| ***Penetration rate*** | | | |
| Penetration rate before implementation | | ↑ | ↓ |
| Penetration rate after implementation | | ↓ | ↑ |
| | | | |
| ***Side effects*** | | | |
| Description of side effects | | | |
| Annual benefit of side effects | | ↓ | ↑ |
| *Total benefit of side effects* | | ↓ | ↑ |
| | | | |
| ***Crash costs*** | | | |
| CPU_Country2015_fatality | | ↓ | ↑ |
| CPU_Country2015_serinjury | | ↓ | ↑ |
| CPU_Country2015_slinjury | | ↓ | ↑ |
| CPU_Country2015_SeriousSlight | | ↓ | ↑ |
| CPU_Country2015_FatalSeriousSlight | | ↓ | ↑ |
| CPU_Country2015_FatalCrash | | ↓ | ↑ |
| CPU_Country2015_SeriousCrash | | ↓ | ↑ |
| CPU_Country2015_SlightCrash | | ↓ | ↑ |
| CPU_Country2015_PDO_Crash | | ↓ | ↑ |
| CPU_Country2015_SeriousSlight_Crash | | ↓ | ↑ |
| CPU_Country2015_FatalSeriousSlight_Crash | | ↓ | ↑ |

Each of the input fields is in principle suited for a sensitivity analysis. The most important factors are the crash costs, the costs of countermeasures, and the effectiveness.

For **crash costs** it is advised to enter the country's common-methodology estimate next to its own reported value (see Section 10.9.1), except if working with the EU standardized value (because it is based on the common methodology already).

For the **effectiveness**, the indicators that are used to estimate the number prevented crashes usually come with a 95% confidence interval which should be used to calculate the lower- and upper-threshold for the input values.

For the costs of the **countermeasure**, it is advised to double them and half them. This leads to the following versions of the analysis:

- Common methodology estimate for crash costs

- Low measure effect (lower confidence interval from source study)
- High measure effect (upper confidence interval from source study)
- High measure costs (+ 100%)
- Low measure costs (-50%)

The different versions concerning effectiveness and costs of the countermeasure should be combined into two scenario's:

- Worst case: low effectiveness + high measure costs
- Ideal case: high effectiveness + low measure costs

## 12.4   CONCLUSION

Based on the benefit-to-cost ratio (BCR) a conclusion can be written on the estimated economic efficiency of the measure under study. In principle a BCR above 1 means that the measure is economically efficient, a value below 1 means that the measure is not economically efficient. However, before obtaining conclusions one should thoroughly consider the uncertainty of the underlying estimates. This uncertainty refers to the estimated measure costs, the assumed effects of the measure (and its confidence intervals) and the assumed target group of crashes.

Based on the benefit-cost ratio (CBR) we can give another colour-code:

Green:
CBR > 1
CBR with ↓field values >1
If you have several examples: all (or most) resulting CBRs > 1


Grey:
- Mixed results
- The range of CBR's from sensitivity analysis contains 1
- Result depends on whether side-effects are included or not

Red:
CBR <1
CBR with ↑field values <1
If you have several examples: all (or most) resulting CBRs < 1

# Part 4 – Conclusion and Outlook

# 13 A methodological framework

The SafetyCube project has been dedicated to reviewing literature on Risk factors and counter-measures. From these reviews a Decision Support System has been built, which will enable policy makers to quickly gather information for their road safety strategies, which risks are most important to take into account and which measures are most suitable for their situation.

Delivering this information in a reliable way requires first a thorough review of the existing literature, then registering all relevant articles for each risk factor and measure, and finally summarising the information from literature.

## 13.1 UNDERSTANDING THE LITERATURE

The most difficult point when doing a literature review, is to decide to what extent the results from different studies are comparable. Two examples:

- The MAIDS study of crash causation factors for motorcycle crashes found that 21% of the crashes were speed related. A percentage more or less doubled by several other studies (MOTAC, 36%, COMPAR 45%; HVU, 50%; Vägverket 40%). An important detail of the MAIDS study was that only speed relative to other road-users was considered a possible factor, meaning that all cases where the motorcyclist was riding by him/her self where not taken into consideration. When reviewing the results of speeding as a risk-factor it is thus important to have a close look at the definition of "speeding" in each of the studies.

- When evaluating driver training, bias due to self-selection is an important obstacle: those who find it necessary to follow training are often more conscious of risks and avoiding them than those who do not sign up. Harrington (1972) compared the crash records of drivers who chose to take high-school driver education to those who did not choose to take high-school driver education. While the uncorrected results yielded a large advantage for those who followed the course, the differences were all but vanished when the estimates were corrected for self-selection bias.

These examples show that the true difficulty in summarising the results from different studies is not to add up the effects observed (and weigh them appropriately), but to judge whether the results are comparable. Was the implementation of measures the same? Have the same variables been controlled for? Did the authors apply the same (or comparable) corrections? Over which time-spans were the effects observed? Where the control groups (even if not perfect) more or less the same in different studies? All these questions are vital when having to decide whether effects from different studies can be compared, e.g., in a meta-analysis. Even more so, they need to be taken into account when evaluating whether a risk factor or countermeasure has been studied satisfactorily. For this reason, this methodology starts with some general consideration how risk-factors and countermeasures are studied in road safety. Moreover, a course on study designs, their typical results as well as their weakness and potential pit-falls is given.

## 13.2 CODING AND SUMMARISING RESULTS FROM LITERATURE

Furthermore, in this methodology detailed instructions are given to SafetyCube partners on how to collect scientific evidence for the decision support system. It is described how the literature is searched, and how studies are selected and prioritised for inclusion into the system.

The literature has to be searched in a systematic and transparent way so that the interested user can comprehend why studies are included in the system and why others might not be. Next to applying fixed criteria to include or exclude studies, we will realistically not always be able to code all studies that qualify on the basis of the inclusion criteria. In these cases it is important to indicate by which criteria the studies have been prioritised for coding.

Studies can be included into the DSS by coding them into a template (Excel based), which are then read into a database of coded studies. This database forms the back-end of the decision support system allowing end-users to gain information quickly about the research question of these studies with respect to; the effect of the road-safety measure or the road-risk, about the methods applied, about possible biases of the study, and the important results. The big challenge for the coding template was to make it so flexible that very different types of studies can be entered, preserving the information about study-design and type of information collected, but at the same time allowing to compare the results. The resulting coding template shows therefore a high degree of complexity and requires very precise coding. The information coded in the templates is only valuable when it allows the user to know exactly what the figures mean, and when the input rules are respected so that automatic read-out is possible. A good understanding of the research designs applied and a close study of the coding instructions are therefore a necessary requirement for entering studies into the DSS.

The next step to making the scientific information available for policy makers and other stakeholders is to give an overview of the results, to find a suitable way to summarise the results and to write a synopsis of the results. The DSS is dedicated to supporting such reviewing and summarising activities. The output is in the form of tables of studies (including descriptions of the methodology applied, the most important characteristics of the sampling frame, and the outcome measured), and tables of study-results (including the conditions compared, the effects and their statistical characteristics (confidence intervals or standard errors, possibly results of statistical tests, and p-values)). The overview of study characteristics will help to decide whether the results can be analysed in a meta-analysis, which is only possible if the resulting effect estimates are based on the same effect estimate (e.g. all Odds ratios) and result from comparable study designs. If this is not the case a vote count analysis should be conducted. On the basis of the result overview it can be decided which effects should be summarised jointly and which conditions should be kept separate, so that in the subsequent meta- or vote-count analysis the results are reported per condition.

All these steps lead to a compact synopsis that gives lay-men a good overview of the most important findings, containing the general estimation of how effective a measure is or how risky a particular factor is next to a description which factors influence this general tendency. Next to the quantitative results coded in the studies included to the DSS, the synopsis should also give a short introduction of the mechanisms and, if applicable, the theoretical framework associated with a particular countermeasure or risk factor.

## 13.3  EVALUATING EFFICIENCY IN ECONOMIC TERMS

Based on the results form effectiveness studies of road safety countermeasures, it is important to give decision makers an indication how much value they will get for their money in each countermeasure. An economic efficiency evaluation is implemented as part of the studies that form the SafetyCube Decision Support System. The studies were to some extent based or prior existing benefit-cost studies. In those cases, the costs were up-dated to the year 2015 and the crash costs were replaced by the crash costs that have been collected in the SafetyCube project – in cooperation with the InDev Project. This procedure assured a maximal comparability across benefit-cost studies with respect to method and time of the cost estimates.

# 14 Outlook

More than 1200 studies evaluating risks or measures have been analysed with more than 7000 estimates of risks or measure effects coded. They have been summarised in 210 synopses and more than 35 cost benefit analyses. All these analyses followed the methodological guidelines outlined here and therefore show maximal transparency, comparability, scientific rigour, and faithfulness to the original studies.

All analyses are now available in the SafetyCube DSS (Decision Support System) that is available at the following URL: http://www.roadsafety-dss.eu. Its pilot operation started early 2017; since then the system has been updated continuously and this process will continue until April 2018 (end of the SafetyCube project) and beyond. The framework presented here also forms the basis for such an extension. It will enable coders to apply the same rules as they have been applied throughout the project and will help to up-date the synopses and add new ones.

The DSS is intended to become a major source of information for industry, policy makers and the wider road safety community; it incorporates the knowledge base of crash causation, risks and measures developed in the project and the underlying methodological systems. The DSS has a great potential to further support evidence-based decision making at local, regional, national and international level, aiming to fill in the current gap of comparable measures effectiveness evaluation across Europe and worldwide.

# References

Alfaro, J. L., Chapuis, M., Fabre, F. (Eds.) (1994). Socio-economic cost of road accidents: final report of action COST 313. Commission of the European Community, Brussels.

Ball, K., Ackerman, M. (2011). The older driver (Training and assessment: Knowledge skills and attitudes. Handbook of Driving Simulation for Engineering, Medicine and Psychology, CRC Press.

Bambach, M.R.; Mitchell, R.J.; Grzebieta, R.H.; Olivier, J. (2013). The effectiveness of helmets in bicycle collisions with motor vehicles: A case–control study. Accident Analysis and Prevention 53 (2013) 78– 88. DOI: http://dx.doi.org/10.1016/j.aap.2013.01.005

Bickel, P. et al. (2006). Proposal for harmonised guidelines. EU project HEATCO Deliverable 5. University of Stuttgart, Stuttgart.

Bieber-Tregear, M., Funmilayo, D., Amana, A., Connor, D., & Tregear, S. (2011). Diabetes and commercial motor vehicle safety. Department of Transportation's Federal Motor Carrier Safety Administration.

Blaeij, A.T. de, Florax, R.J.G.M., Rietveld, P. & Verhoef, E. (2003). The value of statistical life in road safety; A meta-analysis. In: Accident Analysis and Prevention, vol. 35, nr. 6, p. 973-986.

Boardman, A.E. et al. 2006. Benefit-cost analysis. Concepts and practice. Third edition. Prentice Hall, Upper Saddle River, New Jersey.

Boele, M., De Craen, S. (2014) Langetermijneffecten van een eendaagse voortgezette opleiding voor motorrijders. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid (SWOV). Den Haag. https://www.swov.nl/rapport/R-2014-22A.pdf; www.swov.nl/rapport/R-2014-22E.pdf

Boyle, L. (2011). Analytical Tools, In: Fisher, D., Rizzo, M., Caird, J., Lee J., Handbook of Driving Simulation for Engineering, Medicine and Psychology, CRC Press.

Caird J. K., Horrey W. J., (2011), Twelve practical and useful questions about driving simulation, Handbook of driving simulation for engineering, medicine and psychology, CRC Press.

Chu X., Guttenplan, M., Baltes M. (2003). Why People Cross Where They Do - The Role of the Street Environment. Transportation Research Record No 1878, pp. 3-10.

Commandeur, J. J. F., Koopman, S. J. (2007). An introduction to state space time series analysis. Oxford, Oxford University Press.

D3.2. (2017) Crash cost estimates for European countries. Deliverable of EC H2020 project SafetyCube.

D3.4. (2017) Preliminary Guidelines for priority setting between measures. Deliverable of EC H2020 project SafetyCube.

D3.5 (2018) Guidelines for priority setting between measures with practical examples. Deliverable of EC H2020 project SafetyCube.

.

D7.3 Costs related to serious road injuries (2017). Deliverable to EC H2020 project SafetyCube.

D8.3 Development of the European Road Safety Decision Support System. Deliverable 8.3 of the SafetyCube project.

DaCoTA (2012). Benefit-cost analysis. Deliverable 4.8d of the EC FP7 project DaCoTA.

Drummer, O. H., Gerastomoulos, J., Batziris, H., Chu, M., Caplehorn, J., Robertson, M. D. & Swann, P. (2004). The involvement of drugs in drivers of motor vehicles killed in Australian road traffic crashes. Accid Anal Prev 36, 239-48.

ECMT (1998). Efficient transport for Europe; Policies for internalisation of external costs. Organisation for Economic Co-operation and Development OECD, Paris.

Eisenberg, D. (2004). The mixed effects of precipitation on traffic crashes. Accident Analysis and Prevention, 36, 637-647. Doi: 10.1016/S0001-4575(03)00085-X

Elvik, R. (2001). Benefit-cost analysis of road safety measures: applicability and controversies. Accident Analysis and Prevention, 33, 9-17.

Elvik, R. (2004). To What Extent Is There Bias by Selection?: Selection for Road Safety Treatment in Norway. Transportation Research Record: Journal of the Transportation Research Board(1897), 200-205. doi:10.3141/1897-26

Elvik, R. (2007). Prosoects for improving road safety in Norway. Report 897. Oslo, Institute of Transport Economics.

Elvik, R. (2008). Comparative analysis of techniques for identifying locations of hazardous roads. Transportation Research Record: Journal of the Transportation Research Board(2083), 72-75. doi:10.3141/2083-08

Elvik, R. (2009). An exploratory analysis of models for estimating the combined effects of road safety measures. Accident Analysis & Prevention, 41(4), 876-880. doi:10.1016/j.aap.2009.05.003

Elvik, R. (2011). Effects on accident risk of using mobile phones: Problems of meta-analysis when studies are few and bad. Transportation Research Record, 2236, 20-26.

Elvik, R. (2012). Analytic choices in road safety evaluation: Exploring second-best approaches. Accident Analysis & Prevention, 45, 173-179. doi:10.1016/j.aap.2011.12.006

Elvik, R. (2014). Problems in determining the optimal use of road safety measures. Research in Transportation Economics, 47, 27-36.

Elvik, R. (2015a). Methodological guidelines for developing accident modification functions. Accident Analysis & Prevention, 80, 26-36. doi:10.1016/j.aap.2015.03.038

Elvik, R., & Amundsen, A. H. (2000). Improving road safety in Sweden. TØI report 490/2000. Oslo: Institute of Transport Economics.

Elvik, R., Bjørnskau, T. 2015. Safety-in-numbers: A systematic review and meta-analysis of evidence. Forthcoming in Safety Science.

Elvik, R., Hoye, A.; Vaa, T., Sorensen, M. (2009). The handbook of road safety measures. 2nd Edition. Emerald, Bingley, UK.

Elvik, R., Rydningen, U. 2002. Effektkatalog for trafkksikkerhetstiltak. Rapport 572. Oslo, Transportøkonomisk institutt.

Elvik, R., Veisten, K. (2005). Barriers to the use of efficiency assessment tools in road safety policy. Report 785. Institute of Transport Economics, Oslo.

ETSC (2016). Ranking EU progress on road safety. 10th Road Safety Performance Index Report.

Ewert, U. (2014). Evaluation der Kampagne „Kopfstützen schützen" – Kampagne des Schweizerischen Versicherungsverbandes (SVV) im Auftrag des Fonds für Verkehrssicherheit (FVS). Bern: BfU.

FHWA (2010). A Guide to Developing Quality Crash Modification Factors. Federal Highway Administration. Report No. FHWA-SA-10-032. Gross F., Persaud B., Lyon C.

Fildes,B; Keall, M; Bos, N.; Lie, A.; Page, Y.; Pastor, C.,Pennisi g, L.; Rizzi, M., Thomas, P.; Tingvall, C. (2015). Effectiveness of low speed autonomous emergency braking in real-world rear-end crashes. Accident Analysis and Prevention 81 (2015) 24–29. DOI:10.1016/j.aap.2015.03.029

Gadegbeku, B., Amoros, E. (2007). Responsibility study: Main illicit psychoactive substances among car drivers involved in fatal road crashes in France. Deliverable 2.3.2 of the EC FP6 Project DRUID.

Gennarelli, T. W. (2008). The Abbreviated Injury Scale 2005. Update 2008. Des Plaines, IL: American Association for Automotive Medicine (AAAM).

Gollier, Christian (2005), "Quel taux d'actualisation pour quell avenir ? », Revue française d'économie, vol. 19, n°4, pp. 59-81.

Grabowski, D. C., Campbell, C. M., & Morrisey, M. A. (2004). Elderly Licensure Laws and Motor Vehicle Fatalities. JAMA, 291(23), 2840-2846.

Haight, F.A., 1970. A crude framework for bypassing exposure. J. Saf. Res. 2, 26–29.

Hakkert, S. & Wesemann, P. eds (2004). The use of efficiency assessment tools. Solutions to barriers. ROSEBUD Report Nr. 5.

Harrington D. M. (1972). The young driver follow-up study: an evaluation of the role of human factors in the first four years of driving. Accident Analysis and Prevention, 4, 191-240.

Harvey, A., & Durbin, J. (1986). The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling. Journal of the Royal Statistical Society, Series A, 149, 187-227.

Hauer, E. (1986). On the estimation of the expected number of accidents. Accident Analysis & Prevention, 18(1), 1-12. doi:10.1016/0001-4575(86)90031-X

Hauer, E. (1997). Observational Before/After Studies in Road Safety. Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Oxford: Pergamon press (Elsevier Science).

Hauer, E. (1997). Observational Before/After Studies in Road Safety. Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Oxford: Pergamon press (Elsevier Science).

Hauer, E., Bonneson, J., Council, F., Srinivasan, R.., Zegeer, C. 2012. Crash modification factors. Foundational issues. Transportation Research Record, 2279, 67-74.

Hauer, E., D. Harwood, F. Council, M. Griffith. Estimating safety by the empirical Bayes method: a tutorial. In Transportation Research Record: Journal of the Transportation Research Board, No. 1784, Transportation Research Board of the National Academics, Washington, D.C., 2002, pp. 126-131.

Hels, T., Bernhoft, I. M., Lyckegaard, A., Houwing, S., Hagenzieker, M., Ugent, S-A., Isalberti, C., Van der Linden, T., Verstraete. A. (2012). Risk of injury by driving with alcohol and other drugs. DRUID Deliverable D2.3.5. Copenhagen, DTU Transport.

Hels, T., Bernhoft, I. M., Lyckegaard, A., Houwing, S., Hagenzieker, M., Ugent, S-A., Isalberti, C., Van der Linden, T., Verstraete. A. (2012). Risk of injury by driving with alcohol and other drugs. DRUID Deliverable D2.3.5. Copenhagen, DTU Transport.

Hermitte, T. (2012) Review of Accident causation models used in Road Accident Research. Deliverable D5.9 of the EC FP7 project DaCoTA.

Hollnagel, E., (1999). Anticipating failures: what should predictions be about? In: Proceedings of the Human Factors and Medicine Panel (HFM) Workshop, HFM, Siena, Italy.

HVU (2009). Motorcycle Accidents. Danish Road Traffic Accident Investigation Board

Jaensch M., Otte D., Pund B., Chiellino U. Hoppe M. (2008). Implementation of ACASS – Accident Causation Analysis with Seven Steps – in In-Depth Accident Study GIDAS. International Symposium on ESAR "Expert Symposium on Accident Research", 3rd, 2008, Hannover, Germany.

Jiang, X., Lyles, R. W. (2010). A review of the validity of the underlying assumptions of quasi-induced exposure. Accident Analysis and Prevention, 42, 1352-1358.

Johannsen, H. (2013). *Unfallmechanik und Unfallrekonstruktion.* Springer Vieweg.

Kröyer, H. R. G. 2015. Accidents between pedestrians, bicyclists and motorised vehicles: Accident risk and injury severity. Doctoral dissertation. Lund University, Department of Technology and Society.

Lachin JM, Matts JP, Wei LJ (1988). "Randomization in clinical trials: conclusions and recommendations". Controlled Clinical Trials 9 (4): 365–74.

Lardelli-Claret, P.; Luna-del-Castillo, J.; Jimenez-Moleon, J. J.; Rueda-Dominguez, T.; Garcia-Martin, M.; Femia-Marzo, P.; Bueno-Cavanillas, A. (2003). Association of main driver-dependent risk factors with the risk of causing a vehicle collision in Spain, 1990-1999. Annals of Epidemiology, 13, 509-517.

Liu, B. C., Ivers, R., Norton, R., Boufous, S., Blows, S., & Lo, S. K. (2008). Helmets for preventing injury in motorcycle riders. Cochrane Database Syst Rev, (1), Cd004333. http://doi.org/10.1002/14651858.CD004333.pub3

Martensen, H. et al. (2016), Preliminary guidelines for priority setting between measures, Deliverable Number 3.4 of the H2020 project SafetyCube.

MAIDS (2004). MAIDS In-Depth investigation of accidents involving powered two wheelers. Final report 2.1. ACEM.

McGwin, G., Sims, R. V., Pulley, L., Roseman, J. M. (2000). Relations among chronic medical conditions, medications, and automobile crashes in the elderly: a population-based case-control study. American Journal of Epidemiology, 152, 424-431.

McLean R., Anderson V. (1984). Applied factorial and fractional designs. Marcel Dekker Inc., New York, USA.

Méndez, A.G., Izquierdo, F.A. (2010). Quasi-induced exposure: The choice of exposure metrics. Accident Analysis and Prevention, 42, 582-588.

Montgomery D. (2000). Design and analysis of experiments. (5th ed.)Wiley and Sons, New York, USA (2000).

MOTAC: Martensen, H. & Roynard, M. (2013) Motorcycle accident causation. Diepteanalyse van zware en dodelijke ongevallen waarin motorfietsers betrokken waren. Brussel, België: Belgisch Instituut voor de Verkeersveiligheid – Kenniscentrum Verkeersveiligheid

Moreno-Altamirano A, López-moreno S, Corcho-berdugo A. Principales medidas en epidemiología. Salud Publica Mex. 2000;42(4).

Naing, C. , Bayer, S., Van Elslande, P., Fouquet, K.. (2007) Which Factors and Situations for Human Functional Failures? Developing Grids for Accident Causation Analysis. Deliverable 5.2 of the EC project TRACE (Traffic Accident Causation in Europe)

NCEC (National Center for Environmental Economics (2010), "Discounting Future Benefits and Costs", in Guidelines for Preparing Economics Analysis, chapter 6, pp. 6-1 -6-20, US Environmental Protection Agency, Office of Policy.

Nilsson, G. (2004). Trafiksäkerhetsåtgärder och efterlevnad. Hastighetsanpassning, användning av bilbälte och nykter som bilförare.

Nordtyp-projektgruppen. (1980). Trafikulykker på vejstrækninger. En sammenstilling af ulykkesfrekvenser for nordiske typesektioner. Rapport. Vejdirektoratet, København.

OECD/ITF (2015), Road Safety Annual Report 2015, OECD Publishing, Paris. http://dx.doi.org/10.1787/irtad-2015-en

OECD-ITF (2011). Reporting on serious road traffic casualties . Combining and using different data sources to improve understanding of non-fatal road traffic crashes. IRTAD Working group on combining police and hospital data. OECD Publishing, Paris.

OECD-ITF (2012). Sharing Road Safety: Developing an International Framework for Crash Modification Functions. OECD Publishing.

Pai, M. Measures of effect and potential impact. McGill University, Montreal, Canada. http://www.teachepi.org/documents/courses/fundamentals/Pai_Lecture4_Measures%20of %20Effect%20and%20Impact.pdf. Downloaded Sep 2015.

Papadimitriou, E., Lassarre, S., Yannis, G. (2016), Introducing human factors in pedestrian crossing behavior models. Transportation Research Part F 36, pp. 69-82.

Papantoniou, P., Papadimitriou, E., Yannis, G. Assessment of driving simulator studies on driver distraction (2015) Advances in Transportation Studies, 35, pp. 129-144.

Persaud, B., Lyon, C., 2007. Empirical Bayes before–after studies: lessons learned from two decades of experience and future directions. Accident Analysis and Prevention 39, 546–555

Porta M. A dictionary of epidemiology. Sixth edit. Press OU, editor. 2014.

Reason, J. (2000). Human error: models and management. BMJ 18 320(7237): 768–770.

Regan, M., Williamson, A., Grzebieta, R., Tao, L. (2012). Naturalistic driving studies: literature review and planning for the Australian Naturalistic Driving Study, A safe system: expanding the reach: Australasian College of Road Safety national conference, Sydney.

Rizzo, M., Jermeland, J., & Severson, J. (2002). Instrumented vehicles and driving simulators. Gerontechnology, 1 (4), 291-296.

Robertson, L. S. (1992). Injury Epidemiology: Oxford University Press, USA.

Rosén E, Sander U. (2009). Pedestrian fatality risk as a function of car impact speed. Accident Analysis & Prevention, 41(3): 536–542.

Ross, H.L. (1982). Deterring the Drinking Driver. Legal policy and Social Control. Lexington Books, D.C. Heath and Company, Lexington; Massachusetts.

Savoleinen P.T., Mannering, F.L., Lord, D., Quddus, M.A., (2011). The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accident Analysis and Prevention 43(5), 1666-76. doi: 10.1016/j.aap.2011.03.025.

Schiaffino A. ¿ Odds ratio o razón de proporciones ? Gac Sanit. 2003;17(1):70–4.

Schulz KF, Grimes DA (2002). "Generation of allocation sequences in randomised trials: chance, not choice". Lancet 359 (9305): 515–9.

Siegrist, S. (2010). Towards a method to forecast the effectiveness of national road safety programmes. Safety Science, 48(9), 1106–1110. http://doi.org/http://dx.doi.org/10.1016/j.ssci.2009.11.007

Stamatiadis, N., Deacon, J. A. (1997). Quasi-induced exposure: Methodology and insight. Accident Analysis and Prevention, 29, 37-52.

Stanley, T., Doucouliagos, H. 2015. Neither fixed nor random: weighted least squares meta-analysis. Statistics in Medicine, 34, 2116–27.

Staten vegvesen (2012). Special Analysis Fatal Motorcycle Accidents 2005-2009. Norwegen Public Roads Administration, Report Nr. 45.

Strandroth, J. (2005). Fatal motorcycle accidents: an analysis of the Swedish Road Administration indepthstudies 2000-2003, Swedish Road Administration.

Tefft B. (2011) Impact speed and a pedestrian's risk of severe injury or death. AAA Foundation for Traffic Safety, Washington.

Tefft, B. C. (2014). Driver license renewal policies and fatal crash involvement rates of older drivers, United States, 1986-2011. Injury Epidemiology 2014, 1:25, Accident Analysis and Prevention 45, 634-638.

Thorpe, J.D., 1967. Calculating relative involvement rates in accidents without determining exposure. Australia Road Res. 2, 25–36.

Trawén, A., Maraste, P. & Persson, U. (2002). International comparison of costs of a fatal casualty of road accidents in 1990 and 1999. Accident Analysis and Prevention, vol. 34, nr. 3, p. 323-332.

Van Elslande P., Fouquet K. (2008) Analyzing human functional failures in road accidents'. Deliverable 5.1 of the EC project TRACE (Traffic Accident Causation in Europe).

COMPAR : Van Elslande, P., Fournier, J.Y., Jaffard, M. (2011). Facteurs d'accidents, défaillances fonctionnelles et configurations accidentelles. IFSTTAR/DSCR.

Van Elslande, P., Fournier, J.Y., Vincensini, M., Roynard, M., Nussbaum, F., Clabaux, N. (2008). Comparative de procédures d'accidents mortels et non-mortels. Rapport R3 du Projet ANR Predit 2RM.

Van Schagen, I., Welsh, R., Backer-Grondahl, A., Hoedemaeker, M., Lotan, T., Morris A., Sagberg, F., Winkelbauer, M. (2011). Towards a large-scale European Naturalistic Driving study: final report of Prologue, Deliverable D4.2.Wijnen, W. & Stipdonk, H.L. Social costs of road crashes: an international analysis. Paper submitted for publication.

Viechtbauer W. (2010). Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software, Volume 36, Issue 3.

Weijermars, W., & Wesemann, P. (2013). Road safety forecasting and ex-ante evaluation of policy in the Netherlands. Transportation Research Part A: Policy and Practice, 52, 64–72.

Wijnen, W., Weijermars, W., Vanden Berghe, E., Schoeters, A., Bauer, R., Carnis, L., Elvik, R., Theofilatos, A., Filtness, A., Reed, S., Perez, C. and Martensen, H. (2017), Crash cost estimates for European countries, D3.2 of the H2020 project SafetyCube.7.4 of the EC FP6 Project TRACE.

Wijnen, W., Wesemann, P. & de Blaeij (2009). Valuation of road safety effects in benefit-cost analysis. Evaluation and Program Planning, 32, pp. 326-331.

World Health Organization WHO. Global Status Report in Road Safety. 2013.

World Health Organization. Basic epidemiology. 2006.

Yannis G., Gitelman V., Papadimitriou E., Hakkert A.S., Winkelbauer M. (2008), Testing a framework for the efficiency assessment of road safety measures, Transport Reviews 28 (3), 281-301.

Zangmeister, T., Kreiss J.P., Schüler, L. (2007). Evaluation of Existing Safety Features. Deliverable 7.4 of the EC FP6 Project TRACE.

Reason J. Human error. New York: Cambridge University Press; 1990.

Reason, J., 1997. Managing the risks of organizational accidents.Ashgate.

Hollnagel, E., 2004. Barriers and Accident Prevention. Aldershot, Ashgate.

Hollnagel, E., (1999). Anticipating failures: what should predictions be about? In: Proceedings of the Human Factors and Medicine Panel (HFM) Workshop, HFM, Siena, Italy.

Hollnagel, E., (1998). Cognitive Reliability and Error Analysis Method (CREAM) . Elsevier.

Hollnagel & Woods, 2005 Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. CRC Press, Boca Raton, Florida, USA.

Ljung, M. (2002). DREAM: Driving Reliability and Error Analysis Method. Master's thesis from University of Linköping's Web site: http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-2033