



Deliverable 2.3
EVALUATION TOOL FOR ROAD SAFETY
CAMPAIGNS
Final version
Public

CAMPAIGNS AND AWARENESS-RAISING
STRATEGIES IN TRAFFIC SAFETY

Contract No TREN-05-FP6TR-S07.59641-019520-CAST

Project Coordinator: Belgian Road Safety Institute
(IBSR/BIVV)

Author(s): Ankatrien Boulanger (IBSR/BIVV), Stijn Daniels (IMOB), Marko Divjak (ULFF), Isabelle Goncalves (INRETS), Annette Meng (DTU), Inger Synnøve Moan (TØI), Teti Nathanail (UTH), Ivanka Orozova-Bekkevold (DTU), Paul Schepers (Min. of Transport NL), Koos Tamis (Min. of Transport NL), Filip Van den Bossche (IMOB), Vlasta Zabukovec (ULFF).

Date: 03 February 2009

Sixth framework programme

Priority sustdev-2004-3.4.1.4.5.
Sustainable surface transport priority

Title	Evaluation tool for road safety campaigns
Authors	Ankatrien Boulanger (IBSR/BIVV), Stijn Daniels (IMOB), Marko Divjak (ULFF), Isabelle Goncalves (INRETS), Annette Meng (DTU), Inger Synnøve Moan (TØI), Teti Nathanail (UTH), Ivanka Orozova-Bekkevold (DTU), Paul Schepers (DGP), Koos Tamis (DGP), Filip Van den Bossche (IMOB), Vlasta Zabukovec (ULFF).
Editor	Ankatrien Boulanger (IBSR/BIVV)
Date	03/02/2009
Report Number	Deliverable 2.3
Version status	Final version
Distribution	Public
QA check	9 January 2009 by K. Hara (Liikenneturva, Finland) 12 January 2009 by T. Hoekstra (SWOV, the Netherlands)

GLOSSARY OF TERMS	5
--------------------------------	----------

EXECUTIVE SUMMARY	7
--------------------------------	----------

1 INTRODUCTION.....	9
----------------------------	----------

1.1 THE CAST PROJECT.....	9
---------------------------	---

1.2 WHY SHOULD I EVALUATE A CAMPAIGN?	10
---	----

1.3 STRUCTURE OF THE DOCUMENT – DESIGN OF THE EVALUATION STUDY	11
--	----

1.3.1 STEPS TO DESIGN, IMPLEMENT AND EVALUATE A ROAD SAFETY CAMPAIGN	11
--	----

1.3.2 STEPS TO MEASURE THE EFFECT OF ROAD SAFETY CAMPAIGNS	12
--	----

2 WHAT I NEED TO KNOW ABOUT THE CAMPAIGN I AM ABOUT TO EVALUATE. 15	
--	--

2.1 TYPOLOGY OF CAMPAIGNS.....	15
--------------------------------	----

2.2 CAMPAIGN ENVIRONMENTAL ASPECTS.....	16
---	----

2.2.1 BUDGET.....	16
-------------------	----

2.2.2 STAKEHOLDERS OF THE EVALUATION	20
--	----

2.2.3 A-PRIORI INFORMATION	23
----------------------------------	----

3 COMPONENTS OF THE EVALUATION STUDY.....	26
--	-----------

3.1 OBJECTIVES OF THE EVALUATION STUDY	26
--	----

3.1.1 EVALUATION QUESTIONS	27
----------------------------------	----

3.2 MEASUREMENT VARIABLES TO DEFINE THE SUCCESS (OR NOT) OF A CAMPAIGN?	28
---	----

3.2.1 METHOD OF ASKING – SELF-REPORTED MEASURES	29
---	----

3.2.2 METHOD OF OBSERVING – OBSERVED BEHAVIOUR.....	44
---	----

3.2.3 METHOD OF DOCUMENT ANALYSIS – ACCIDENT, INJURY AND DEATH STATISTICS	45
--	----

3.2.4 UNINTENDED EFFECTS	47
--------------------------------	----

3.2.5 METHODOLOGICAL ISSUES	47
-----------------------------------	----

3.2.6 PRACTICAL CONSIDERATIONS	49
--------------------------------------	----

3.2.7 ETHICAL CONSIDERATIONS	50
------------------------------------	----

3.3 EVALUATION DESIGN.....	50
----------------------------	----

3.4 MINIMUM STANDARDS FOR THE EVALUATION.....	53
---	----

3.4.1 MINIMUM STANDARDS WITH RESPECT TO MEASUREMENT VARIABLES.....	54
--	----

3.4.2 MINIMUM STANDARDS WITH RESPECT TO EVALUATION DESIGN	55
---	----

3.4.3 MINIMUM STANDARDS WITH RESPECT TO DATA ANALYSIS.....	55
--	----

4 THE CHOICE OF THE PROPER EVALUATION METHODOLOGY.....	56
---	-----------

4.1 HOW TO GET STARTED?	56
-------------------------------	----

4.2 DETERMINATION OF THE EVALUATION COMPONENTS.....	56
---	----

4.2.1 SCOPE.....	57
------------------	----

4.2.2 TARGET GROUP	61
--------------------------	----

4.2.3 OBJECTIVES.....	65
-----------------------	----

4.2.4 ACTIVITIES	72
------------------------	----

5 COLLECTING DATA AND DATA ANALYSIS.....	78
---	-----------

5.1 INTRODUCTION TO SAMPLING	78
------------------------------------	----

5.1.1 SAMPLING TECHNIQUES.....	78
--------------------------------	----

5.1.2	SAMPLE SIZE.....	81
5.1.3	BIAS IN SAMPLING	82
5.1.4	NON-RESPONSE.....	83
5.2	COLLECTING THE DATA	84
5.3	ANALYSING THE DATA.....	85
5.3.1	WHY ARE STATISTICS NECESSARY?	87
5.3.2	DESCRIPTIVE STATISTICS	88
5.3.3	INFERENTIAL STATISTICS	89
5.3.4	MAGNITUDE OF A CAMPAIGN EFFECT.....	93
5.3.5	REPORTING STATISTICS.....	93
6	<u>REPORTING AND DISSEMINATION OF RESULTS.....</u>	<u>95</u>
6.1	MAJOR SECTIONS OF AN EVALUATION REPORT	95
6.2	WRITING STYLE.....	100
6.3	DISSEMINATION OF RESULTS	101
6.4	USE FOR FUTURE CAMPAIGNS	101
	<u>CONCLUSION.....</u>	<u>103</u>
	<u>BIBLIOGRAPHY</u>	<u>105</u>
	<u>ANNEX 1: ELEMENTS OF A DETAILED DESCRIPTION OF A ROAD SAFETY CAMPAIGN: SUMMARY.....</u>	<u>110</u>
	<u>ANNEX 2: ADVANTAGES AND DISADVANTAGES OF DATA COLLECTION TECHNIQUES: OVERVIEW.....</u>	<u>124</u>
	<u>ANNEX 3: IMPLICATIONS OF THE EVALUATION DESIGNS ON THE BUDGET</u>	<u>130</u>
	<u>ANNEX 4: FORMULAS TO MEASURE BEHAVIOURAL, NORMATIVE AND CONTROL BELIEFS.....</u>	<u>133</u>
	<u>ANNEX 5: ASSUMPTIONS IN REGRESSION ANALYSIS</u>	<u>134</u>
	<u>ANNEX 6: AN INDICATIVE CASE STUDY ON STATISTICAL DATA ANALYSIS.....</u>	<u>135</u>

Glossary of terms

Data collection methods:	Data collection method is the general orientation towards data collection (e.g. the method of asking, the observational method and the method of document analysis).
Data collection techniques:	Data collection technique is more concrete than data collection method and specifically refers to the actual means/instruments/procedures for data collection (e.g. within the method of asking there are various options of how to approach to the data gathering, such as by means of questionnaires, interviews, etc.).
Economic evaluation:	Economic evaluation critically reviews the costs and the benefits of a road safety campaign and concludes if the costs for campaign design and implementation are justified either in terms of monetary or other savings (e.g. saved lives).
Evaluation:	Evaluation is the systematic collection and analysis of information about the activities, characteristics and outcomes of a road safety campaign, and the implementation of a methodological strategy to determine its effectiveness with regard to the attainment of pre-set road safety objectives.
Evaluation methodology:	Evaluation methodology is a general term which refers to all relevant elements/components of evaluation such as research design, measurement variables, data collection methods and data collection techniques.
Formative evaluation:	This type of evaluation is designed to collect data while a campaign is being developed or when an existing campaign is modified, with the intention to improve it. It is used to evaluate the opportunities and barriers of a campaign before it really starts, in order to minimise costly mistakes.
Measurement variables:	A measurement variable represents a measure of the success or failure of a road safety campaign, relative to the aim of the campaign.
Method of asking:	A data collection method that implies communication between the investigators and the subject of the study. It is also called surveying.
Method of document analysis:	A data collection method that extracts and organises relevant information from a set of documents in a structured way.
Method of observing:	A data collection method that involves the direct observation of phenomena (behaviour or attitude) in their natural setting without interference of the investigator with the subjects of the study.
Outcome evaluation	See summative evaluation.

Process
evaluation

See formative evaluation.

Research
(evaluation)
design:

Research (evaluation) design is a framework that structures the research process, so that it shows how all its major parts (groups of subjects, methods of assigning subjects to different groups, measurement variables and interventions) are related in answering the research question.

Summative
evaluation:

This type of evaluation measures the effectiveness of an intervention on the target population, and determines whether an implemented initiative has had the intended effect and reached the pre-set objectives.

Executive summary

Road safety campaigns play an integral part in the effort to improve road safety and thus reducing the number and/or severity of road accidents. The overall aim of road safety campaigns is to promote safe road user behaviours and to prevent road traffic accidents from occurring. The evaluation of road safety campaigns is essential to establish whether these campaigns are effective at improving road user safety. Evaluation allows knowing the campaign effectiveness, learning about the campaign, benefitting for future campaigns, justifying the costs, facilitating future fund raising and so forth. It is therefore also necessary to publish the campaign evaluation results and make them public available. The aim of this evaluation tool is to provide a practical tool that can be used to carry out a proper effectiveness evaluation study.

The CAST manual (Delhomme, P. et al, 2009) gives practical recommendations on how to design, implement, and evaluate a road safety campaign. The manual outlines six steps needed to complete the entire campaign process. Campaign evaluation is mentioned for the first time in the third step and thus before the actual implementation of the campaign. Indeed, you need to think about and organise the evaluation study right after the campaign design step. Three types of evaluation studies are distinguished namely: process, outcome, and economic evaluation. This evaluation tool will help you through the decisions you need to make about the outcome evaluation. Every chapter corresponds with and explains certain steps needed to measure the effect in a proper way.

What I need to know about the campaign I am about to evaluate

The choice for the appropriate evaluation methodology is dependent on the specific characteristics of the concerned campaign. The CAST project defined the campaign attributes that will have the most determinative implications for the proper evaluation study (Boulanger, A. et al, 2007b). These attributes are: scope (national vs. local), target group (general vs. specific), objectives (to change behaviour, attitudes, perceived norms, knowledge etc.), and supportive activities (media only vs. integrated campaign). Campaigners should also take into account some factors given or determined outside the campaign such as: the budget (more specifically the evaluation budget), stakeholders of the evaluation (e.g., sponsors), and a-priori information (in-depth assessment of the road safety problem, marketing studies of the target group...).

Determination of the proper evaluation methodology

First, it is important to define the success criteria of the road safety campaign and formulate the right evaluation questions. The goal of the evaluation study is thus inseparable linked with the campaign objectives which need to be very clearly defined. Appropriate and answerable evaluation questions facilitate the correct decision of which variables to assess in order to measure the campaign effect. Typical three types of measurement variables are used in road safety evaluation studies namely: self-report measures, observed behaviour and accident statistics. The choice will influence the data collection method. Simultaneously, you need to choose the proper research for producing the required and reliable evaluation data. The practical translation of these important decisions is given in a separate chapter (see chapter 4). If you don't know which measurement variables, evaluation design,

or data collection technique to apply for your evaluation study, you will find the solution or some recommendations in this part of the evaluation tool.

Minimum standards for the evaluation methodology

The best and most appropriate evaluation methodology is not always feasible in practice. The given recommendations are not rules obligated to follow, you may deviate from them but only if you have a good reason to do so. To ensure still a satisfactory level of evaluation data, the CAST consortium decided on some minimum standards for evaluation. The ultimate purpose of these standards is to raise both the validity of evaluation findings and the comparability of the results throughout Europe (and beyond). If you have the skills, the knowledge and extra budget, it is highly recommended to upgrade the minimum standards with the more sophisticated evaluation methodology.

Collect the needed data and analyse the level of success of the campaign

Once the evaluation methodology has been decided, it is time to collect your data from the selected sample of the target group. Different ways exist to select a sample, but one of the most straightforward techniques is called *simple random sampling*. Moreover, the determination of an adequate sample size is one of the most controversial decisions. It is important that the sample is representative of the population. After you have gathered all the necessary information, the next step is to examine it. The purpose of the data analysis is to answer the evaluation questions. The selection of the data analysis technique should be done carefully notwithstanding the limitation by the way the variables were measured. But without doubt, some inferential statistics are required to test the pre-set hypotheses formally. These statistics permit to test whether the descriptive results are likely to be due to random factors or to a real relationship, which makes it possible to make conclusions about the (possible) campaign effect.

Report and disseminate the results

A campaign evaluation report should provide all the information needed to understand the outcome of the campaign. So it is essential to write a detailed and structured report. To ensure that all information is included, this document provides step by step guidelines on how to write a complete evaluation report. Afterwards it is indispensable to disseminate the campaign results as it might even increase the impact of the campaign. Finally, do not forget to consult these reports before starting designing a new road safety campaign!

1 Introduction

1.1 The CAST project

The CAST project is a research project executed from 2006 until 2009 and supported by the European Commission. The main goal of CAST is to provide practical tools to encourage the proper design and evaluation of road safety campaigns in all EU countries (and beyond). The *design tool* or the manual (Delhomme, P. et al, 2009) contains detailed guidelines for designing and implementing a campaign. The *evaluation tool* aims at guiding the users to the best evaluation practice depending on the characteristics of the road safety campaign to enable a thorough effectiveness evaluation. Finally a *reporting tool* (Boulanger, A. et al, 2009) provides precise guidelines for reporting the results of a (single) campaign and its evaluation in a standardised way. These tools ensure that new campaigns can be planned and executed in a way that they will have the optimal effect.

The research activities regarding campaign evaluation in the CAST project aim at dressing up an inventory of evaluation methodologies for road safety media campaigns and at the development of an evaluation tool for (cost)-effectiveness of a single media campaign. This development has been achieved on the basis of a theoretical and practical analysis of a large number of road safety campaigns, based on an exhaustive literature review. This document covers the CAST evaluation tool.

Two other CAST reports contributed to the realisation of the evaluation tool:

- The first report (Boulanger, A. et al, 2007b) lists an exhaustive typology of road safety campaigns. Road safety campaigns differ among other things with regards to their main focus and goals. The appropriate choice for an evaluation method is dependent, among other things, on the goal of the concerned road safety campaign. To identify these determinative attributes of a campaign, an inventory of different media campaigns and the applied evaluation methodology in (and beyond) the EU have been made. The report answers the question: 'what has been done in practice'.
- The more theoretical aspects of 'what possible evaluation methodologies could be applied' are treated in another CAST report (Boulanger, A. et al, 2007a). This report lists all possible evaluation designs, measurement variables and data collection methods and techniques that are at least theoretically possible to use for a campaign evaluation. These evaluation components are compared, both from a theoretical and practical viewpoint in so far relevant for the CAST evaluation tool, with regard to their merits and weaknesses, to measure the effect of a road safety campaign or to isolate the effect of an integrated media campaign. As a result it is possible to identify the appropriate evaluation methodology.

Besides, a reporting checklist for fieldworkers and professionals has been assembled, enabling them to report all relevant aspects of their campaigns and evaluation studies in a standardised way. Such a standard will (1) increase the accessibility and adequateness of relevant information in the domain of media campaigns and (2) facilitate (scientific) evaluation of campaigns. This tool is called

the reporting tool (Boulanger, A. et al, 2009). These evaluation research results contributed to a considerable extent to the development of the CAST manual (Delhomme, P. et al, 2009).

1.2 Why should I evaluate a campaign?

1. *To know the effectiveness of the campaign:* it is essential to measure if the implemented road safety campaign had a positive effect or not in conformance with the preset objectives.
2. *To learn about the campaign:* an evaluation of a road safety campaign offers the opportunity to the parties involved to learn about the campaign. In particular, it is interesting to know why the campaign worked (or not).
3. *To increase general knowledge about road safety campaigns:* campaign evaluations will increase our knowledge about which campaigns are effective and which campaigns are not. In this way improvements to the campaign can be made as an evaluation pinpoints what kind of campaign does or does not work for specific road safety problems, specific target groups. Campaign evaluations will provide more insight in which theoretical models that are suitable for the development of road safety campaigns.
4. *To involve the exposed people:* evaluation gives the opportunity to the exposed audience to directly express their opinion, approval or disapproval of the campaign. Therefore it is very important to report and communicate the campaign results.
5. *To benefit for future campaigns:* to obtain valuable and scientifically proper information about the effectiveness and efficiency of the design method, implementation and measure proposed by the campaign, which can be used for the design and implementation of (or more) successful campaigns. Even in case of non-significant results, the obtained knowledge can be used for focusing on the improvement or modification or elimination of the components, which generated these results.
6. *To demonstrate accountability:* it's a way to justify the costs, associated to the campaign design, implementation and evaluation by comparing quantitative and qualitative results on its effectiveness and therefore convince and attract stakeholders for greater involvement.
7. *To facilitate future fund raising:* after an evaluation is conducted and thus evidence for the campaign effectiveness can be provided, it will be easier to represent both the accomplishments and the future resource needs of the campaign to the policy makers and to the larger public.
8. *To manage resources:* resources should be in line with the expected outcomes of the campaign. The sooner a discrepancy can be detected, the better the campaign can be adapted in terms of available resources, expected outcome or both.
9. *To check the economical advantage/benefit:* if campaigns are not evaluated it is not possible to know whether or not money used to develop and implement road safety campaigns is well spent. Thus, by evaluating campaigns it is possible to save money.
10. *To improve road safety:* if all road safety campaigns are evaluated, the information obtained from these evaluations can be used to increase road safety. The evaluation results of an adequate number of evaluated campaigns enable

the derivation of vital and statistically significant conclusions for the implementation of techniques for improving road safety or road safety related indicators.

1.3 Structure of the document – design of the evaluation study

The CAST manual (Delhomme, P. et al, 2009) is designed to find easy information about the design of road safety campaigns. The manual outlines six different steps on how to proceed. A very important step is the effectiveness evaluation of the campaign. In this section, the different steps to undertake are discussed as well as for the campaign in general as for the effectiveness evaluation study in detail.

1.3.1 Steps to design, implement and evaluate a road safety campaign

The moment the road safety problem has been identified, it should then be studied if it can be solved by a road safety campaign. When you are going to carry out a road safety communication campaign, you can follow six basic steps (see following paragraphs) that will guide you through the process of designing, implementing, and evaluating your campaign. For more comprehensive information on this six-step process, you can consult the CAST manual (Delhomme, P. et al, 2009). The different steps are presented in a sequential way for the sake of clarity. However effective decisions and optimal work at one stage depend on feedback from a previous stage. This means that a given step may be revisited. All steps are interdependent.

Step I: Getting Started

Before designing the campaign, practitioners need to clearly identify the problem. Ideally, this should be done through the examination of objective data (e.g. crash statistics, observed behaviours, data collected by previous campaigns) as this will provide a precise understanding of the contributing factors, the type of road users involved, and the context in which the problem occurs. To get started, practitioners should: (a) analyse the context, (b) identify and involve partners and stakeholders, (c) draft the budget, (d) gather the campaign participants at a kick-off meeting and finally (e) set up the campaign team.

Step II: Situation Analysis

Once practitioners have precisely identified the problem, they should analyse this problem in greater detail in order to define the specific objectives of the campaign. To do so, the following steps should be considered: (a) in-depth analysis of the problem and possible solutions, (b) segmenting the audience, (c) acting on main motivations and reaching the audience, (d) defining specific campaign objectives, and (e) gathering information from evaluations of past campaigns and other actions.

Step III: Designing the campaign and the evaluation

At this stage, the design of the campaign itself starts. Practitioners deal with questions on how the road safety communication campaign should be done and on the ways its efficiency could be assessed. In summary:

- *develop the campaign strategy*; this includes: (1) defining the strategy, (2) developing the message, (3) choosing the media and (4) developing and pre-testing the message and slogans in their full context;
- *design the campaign evaluation* by (1) defining the objectives of the evaluation, (2) choosing the evaluation design and sample, (3) defining methods and tools to collect data and (4) planning the evaluation. For this step it is advisable to consult this evaluation tool (see further).

Step IV: Implementation of the before period evaluation and implementation of the campaign

Prior to launching the campaign, the measurement for the before period of the evaluation study needs to be implemented. The result of the before measurement shall be used as a baseline measurement for the other phases of the evaluation. This will also be the moment for practitioners to produce the actual campaign materials and launch the campaign. To complete this fourth step, the implementation of the campaign materials should be monitored carefully in order to deal with any problems that may arise.

Step V: Completing the evaluation and drawing clear conclusions

In this step, the other scheduled evaluation measurements should be made and compared with the before measurements. In order to determine the effectiveness of the campaign, the collected data should be analysed data. Regardless of whether there were positive or negative effects, all results must be reported.

Step VI: Writing a final report

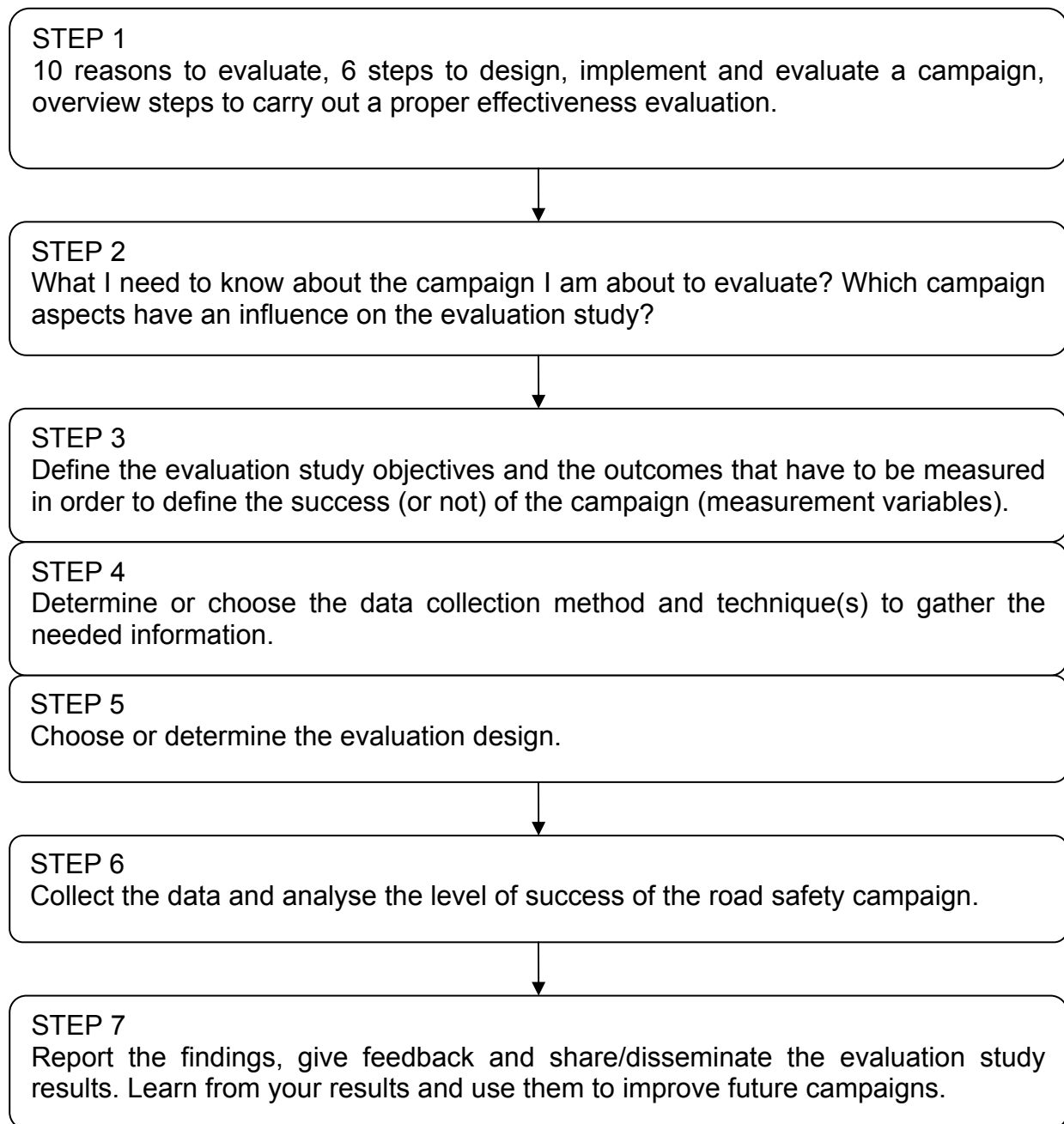
The final report should gather all important information and feedback to provide the campaign partners, stakeholders, and the general public with an in-depth understanding of the campaign results. The improvement of future campaigns depends on the availability of rigorous and easy accessible campaign reports in order to learn from past campaigns.

A clear overview of what the final report should contain can also be found in the reporting tool (Boulanger, A. et al, 2009) of the CAST project.

1.3.2 Steps to measure the effect of road safety campaigns

The CAST project states that implementing a road safety campaign without an evaluation plan is unacceptable. Evaluation is a fundamental part of a campaign. In practice, campaign evaluation is mentioned for the first time in the third step of the CAST manual (Delhomme, P. et al, 2009 or see section 1.3.1). Three types of evaluation are distinguished in the manual namely: process, outcome and economic evaluation. This evaluation tool will help you through the decisions that you need to make about the *outcome evaluation*. As it is advisable to carry out at least two measurements (one measurement before the campaign is implemented), you should start to think about and organise the evaluation right after the campaign design phase. In addition, the utility of measurements gathered during the ongoing campaign reinforces the need for an evaluation process to be continuous and parallel to the implementation of the road safety campaign (Delhomme, P. et al, 2009).

Figure 1.1: Steps to measure the effect of road safety campaign



Regarding the time planning of the evaluation measurements, you should take into account the type of data to collect (see also Box 1.1). This will influence the moments and length of the different evaluation periods. The before measurement should take place before any news about the upcoming campaign has been spread. The after measurement can only start when the campaign has actually ended.

BOX 1.1: A short overview of the timing of the measurements according to the data collection method:

Before measurement

Method of asking: from 2-3 weeks up to 3 months before the start of the campaign.

Method of observing: from 2-3 weeks up to 6 months before the campaign, during at least 2 months.

Method of document analysis: data of at least the last 3 years before the campaign.

After measurement

Method of asking: until one month after the campaign.

Method of observing: during at least 2 months directly after the campaign.

Method of document analysis: data of at least the next 3 years after the campaign.

To offer campaign practitioners a clear overview of the different steps to undertake in order to implement a proper evaluation study (Figure 1.1), we give here an overview and explain the link to the different chapters. Of course some of these steps can and will be taken at the same time.

This first chapter, or step 1, summarises why and how an evaluation should be executed. It also situates the evaluation tool in the wider context of designing and implementing a successful road safety campaign. The next chapter covers step 2 and lists the different campaign characteristics and aspects that influence the choice of the appropriate evaluation methodology. Step 3, 4 and 5 are dealing with the decisions concerning the evaluation components (measurement variables, data collection techniques and evaluation design). These steps are discussed in two chapters, namely 3 and 4. Chapter 3 explains the components of the evaluation study and states the CAST minimum standards for them. Chapter 4 provides practical advice for determining the proper evaluation methodology. After these steps, data will be collected and the level of success will be analysed. More detailed information can be found in chapter 5. Finally, chapter 6 is devoted to reporting and disseminating the evaluation results (step 7).

2 What I need to know about the campaign I am about to evaluate

The choice for the appropriate evaluation methodology is dependent on the specific characteristics of the concerned campaign. Road safety campaigns differ among other things with regards to their main focus (campaign theme) and goals. The CAST project (Boulangier, A. et al, 2007b) defined a typology of campaigns based on relevant attributes and environmental aspects of road safety campaigns that have the most determinative implications for the evaluation study.

Usually an evaluator does not need to start from scratch in defining the components of a campaign description. Quite often, information on the campaign characteristics (such as objectives, the activities....) can be found in existing documents that describe the design of the campaign. A detailed campaign description contains the different components of the campaign and the connections among them (see Annex 1), and can be used to ensure that there is a common understanding of the main aspects of the campaign. A good description of the project guarantees that in the end the right topics are evaluated.

This chapter contains two main sections of elements to take into account when evaluating a particular campaign. Sections 2.1 discuss the different parts of a road safety campaign and their influence on the choice of the evaluation methodology. These campaign attributes are defined as those aspects of a road safety campaign that can have an important influence on the way the evaluation should be done. The second section (2.2) is about the campaign environmental aspects. These aspects are also important in the context of evaluation. Although, they are given or determined outside the campaign, they also have a major influence upon the appropriate evaluation methodology. Factors to consider here are: the evaluation budget, the stakeholders and a priori knowledge.

2.1 Typology of campaigns

Road safety campaigns can be classified along different attributes such as scope, target group, theme, media channels, additional actions.... With regard to scope for example one can distinguish between national and local campaigns (regional, provincial), with regard to theme there are seatbelt campaigns, anti-speeding campaigns, campaigns against drink-driving and so on. In other words, each of these attributes defines a unique typology of road safety campaigns (e.g. typology of campaigns with respect to scope, theme, etc.), while each campaign is characterised by a unique set of characteristics (attributes). These attributes all have different implications for evaluation – some of them are more detrimental while the others are less important for the choice of proper evaluation methodology – evaluation design, measurement variables and data collection techniques. Combining the implications which derive from relevant attributes determines a proper evaluation methodology for that particular campaign.

The CAST project defines scope (national vs. local), target group (general vs. specific; direct vs. indirect), objectives (to change accidents, behaviour, attitudes

etc.) and supportive activities (media only vs. integrated campaign) as those attributes of a road safety campaign that have the strongest implications for evaluation (Boulanger, A. et al, 2007b). The campaign theme and the media plan, on the other hand, are far less influential and thus inferior to the ones listed above. The theme only influences the content and the phrasing of the questions, while the media plan only influences the process evaluation of a campaign (e.g. the subjective exposure to the campaign materials). As we will not discuss the theme and the media plan in detail, a short overview of the consequences for the evaluation methodology is given in the next two paragraphs.

Choosing appropriate and suitable media to communicate and disseminate campaign's message(s) certainly is one of the central issues when designing a road safety campaign. Selection of suitable media plan (combination of various media) should closely follow the campaign's objectives and the relevant characteristics of the target group (e.g. using internet surely is more appropriate when targeting adolescents than when addressing the elderly). Particular combination of media channels could be seen as an important attribute of a campaign. Namely, evaluation questions and measurement variables should to a certain extent be adapted to the media plan. However, used media channels are far less determining for the evaluation methodology than for example the objectives of a campaign.

Generally it is preferable for a campaign to cover one single theme rather than being devoted to several different themes. First of all, the publicity, information and other activities can be more intense and concentrated only upon one specific aspect of road safety. Secondly, it is also clearer for the target audiences what the main message is and what they are expected to change or accomplish. Thirdly, it is easier to find ways to measure the effectiveness of a single-theme campaign (Prigogine, 1975). The theme of a campaign is another important attribute that may influence the evaluation questions although the pre-set objectives are far more determinative (see section 4.2.3).

2.2 Campaign environmental aspects

Campaign designers and evaluators should take into account some factors given or determined outside the campaign as they have a major influence upon the appropriate manner the evaluation can be done. Factors to consider are:

- Evaluation budget
- Campaign stakeholders
- A-priori information

These aspects are called environmental aspects of a campaign and are also important in the context of evaluation. They will be described shortly and their influence on the evaluation components will be explained.

2.2.1 Budget

Regardless of the cost of the campaign itself, every possible evaluation method, even the most basic one, requires the allocation of a certain amount of time and budget resources. This is called the evaluation budget. The evaluation budget should

be clearly distinguished from the campaign budget. Whereas the campaign budget consists of all the costs that are made to design and to implement the campaign, the evaluation budget is limited to those activities that are done to evaluate the effects of the campaign. Consequently, the evaluation budget should be a part of the whole campaign budget. Not only out-of-pocket costs but also for example time efforts (related to the campaign evaluation) that are made by own staff should be included in the evaluation budget.

2.2.1.1 Which elements to take into account in the evaluation budget

The magnitude of the evaluation budget depends mainly on the chosen evaluation design: stronger designs require more measurements, more complex analyses, and higher skilled staff. In general, the needed budget is largely influenced by personnel costs. Other cost factors are the cost of data collection (questionnaires, telephone or internet surveys, observations...) and the cost of hiring external capacity (universities, research institutions, consultants...). From their part, those costs are affected by the sample size and the complexity of the data collection and analysis.

Personnel cost is usually the largest item in a budget; therefore it is of critical importance to pay careful attention to the type and amount of effort that will be required to complete the evaluation. The level of expertise needed to perform particular evaluation tasks will also affect the costs. As a consequence, if the staff needs assistance in conducting the evaluation, external consultants need to be contracted. These consultants can provide specialised expertise and/or different perspectives throughout the activity of evaluation. This need can also be provided by certain stakeholders (see next section). Furthermore, the budget should take into account transportation costs (travel budget for example when interviewing people face-to-face, meetings, feedback workshops, participant involvement, observations, etc.), communication costs (telephone calls, internet access, fax transmissions, postage, etc.), copying and printing costs, and supplies and equipment costs (e.g., computers, packaged software).

2.2.1.2 Need for evaluation budgeting

Too many examples exist of road safety campaigns that were not or improperly evaluated for reasons of costs or time saving. There are at least three explanations for the fact that campaign evaluations are often not carried out or restricted in time and budget: evaluations were until now often the last step in the campaign process, they are considered as 'overhead costs' and the benefits are produced rather on the longer term than on the shorter term. However, those arguments are often incorrect and impede in many cases a rational approach. Moreover those counterarguments should never outweigh the benefits resulting from a proper evaluation as listed in section 1.2.

1. Firstly, the evaluation is the final step of a complete process that includes the planning, design, financing and the execution of a campaign. When designing the campaign, the main focus is placed on the timely implementation of the campaign and not on evaluation. Although a proper evaluation needs to be prepared and even to start (e.g. the before measurement) before the campaign itself is executed, the evaluation is in practice sometimes perceived as a final step that is

still 'comfortably far away' at the moment of the campaign implementation. Another consequence of coming in the last phase of a project is that, in case budget constraints are exceeded in a previous step in the campaign process, it is likely that some savings are strived for in a latter phase, making the evaluation study a likely candidate for budgetary cuts.

2. A second explanation can be found in the fact that evaluation is considered to be an element of overhead or non-direct costs. The evaluation budget is often balanced with the activities budget and since the evaluation does not contribute to the campaign effect directly, the evaluation will often be cut in favour of the campaign activities.

Undoubtedly, this argument is correct to a certain extent. Since an appropriate evaluation needs adequate resources, it could increase the needed campaign budget substantially. However, in that case one should be aware that the simple fact of designing a campaign is costly and time-consuming as well. Whenever one considers setting up a campaign, certainly a mass-media campaign, one must be aware that it will require adequate financial resources. If those resources are not available, the question should perhaps not be whether the campaign should be evaluated or not, but whether the campaign should be executed at all.

3. A third explanation has to do with the expected short term benefits of the campaign. It is only possible to deliver conclusions, and thus revealing possible benefits, after the campaign is finished as the evaluation measurements have to be carried out at least before and after the campaign implementation period. That means that, although the benefits of the evaluation may seem clear to agencies on a longer term, many agencies still fail to allocate sufficient budgetary resources for campaign evaluation.

However, the striking consequence of such a policy is that an institution will never be able to improve future campaigns by the lessons learned from previous evaluations. It must be clear that a purely intuitive approach in evaluating campaigns has serious shortcomings and will not enable sufficient steps forward in tackling some traffic safety problems. Therefore, in the perspective of a long term strategy of quality improvement, a systematic and thorough evaluation of campaigns is highly recommended.

An important element in achieving a useful evaluation is to determine the evaluation methodology, including the budget, from the early beginning of the campaign design. An early allocation of a fixed budget and related human resources for evaluation will also overcome the problem that the evaluation is more or less deliberately 'forgotten' during the entire process.

Sometimes it is advised to build in an evaluation budget that is related to the total campaign budget, for example 10-15% of the total budget (http://www.opportunitystudies.org/repository/File/evaluation/evaluation_budgeting.pdf). However, one must be aware that this is only a rough indication.

In the CAST project (De Dobbeleer, W. et al, 2009) two campaigns were designed, implemented and evaluated according to the CAST guidelines. The evaluation budget used in the Belgian road safety campaign 'Pitstop' was indeed more than 10% of the total campaign budget. On an important side note, the personal costs for preparing the evaluation study, interpreting the results and writing the final report are

not included in this estimate. As far as the Greek road safety campaign “Eyes on the road” is concerned, the evaluation budget slightly exceeded the 10% of the total campaign budget (however, some of the costs are not included in this estimate neither).

Considering the extra efforts and investments required when measuring the isolated effect of an integrated campaign, it might be necessary to deliberate carefully over the gain and if the extra information is worth the costs. Sometimes it is more important for practitioners to know in detail why a campaign did or did not work than knowing how many percent of the effects were due to communication efforts, due to enforcement, and due to other additional efforts. In case an evaluator needs to prove to the decision makers for example that an integrated approach is the most successful, an extra investment should be considered.

2.2.1.3 Implications for the evaluation methodology (Figure 2.1)

- *Introduction*

Being aware of the need for a proper campaign evaluation, an adequate budget needs to be foreseen. The budget comprises those activities that are done to evaluate the effects of the campaign. The evaluation budget is influenced by the evaluation design and by the used measurement variables, data collection methods and techniques.

- *Measurement variables and data collection methods and techniques*

When discussing different data collection techniques and their advantages and disadvantages, it becomes clear that they can complement each other. The table in Annex 2 gives an overview of the advantages and disadvantages of the different data collection techniques. Advantages and disadvantages may differ but prices may differ as well. Using an interviewer will be more expensive, but is likely to get a better response rate than a postal survey. A postal survey, however, will provide for example greater anonymity. In general, existing statistics are a relatively economical and very efficient way of data collecting (for example, possibility to study a trend over time). As mentioned before, a major problem is finding the appropriate units of analysis. On-site observations can be very expensive and time consuming to plan and carry out field trips but is an excellent technique to evaluate actual behaviour.

When choosing a data collection method and technique, it is the intention to reach an optimum between performance (e.g. validity and reliability of the data) and feasibility (e.g. costs and time-efficiency of the technique). Because the costs of the different techniques differ among the EU countries, every researcher should carefully consider their situation.

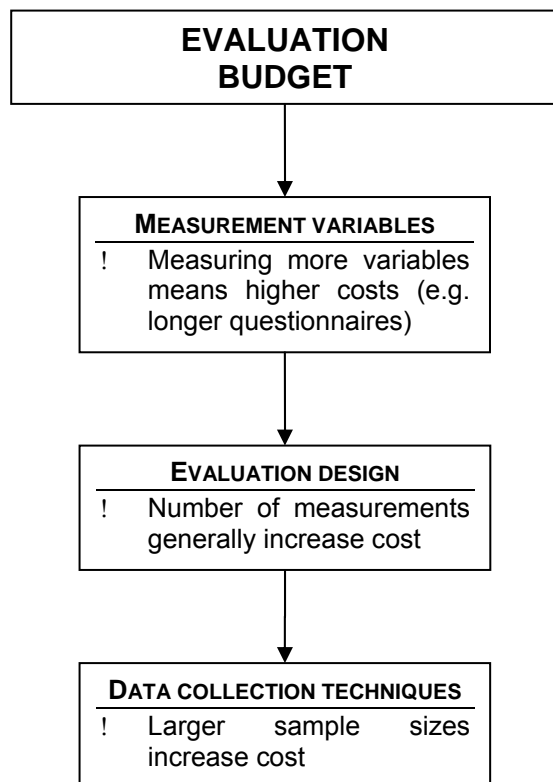
- *Evaluation design*

The cost of the evaluation design generally depends on the number of measurements and on the number of independent groups of subjects. The more measurements and the more independent groups there are, the more expensive the particular evaluation design (see overview implications of the research designs on

the budget in Annex 3). The one-group after-only evaluation design surely is the least expensive design as only one group of subjects is measured after a campaign. The CAST consortium, however, advises against this, as it is the weakest of all designs in terms of internal validity (valid inferences about causes and effects). On the other hand, although single-group and multiple-group designs with more than one measurement (the single-group before-after design, the time series design, and the non-equivalent control group design) have far more internal validity, such designs are much more expensive.

Nevertheless, it needs to be taken into consideration that the cost of a particular evaluation design also depends on how it relates to the measurement variables and/or data collection techniques (see the above subchapter). A combination of for example the single-group before-after design with surveys or accident data that is already at hand certainly is less costly than a combination of the same design with observed behaviour (as two observational studies need to be conducted – before and after a campaign).

Figure 2.1: Budget implications - summary



Note: (!): keep in mind

2.2.2 Stakeholders of the evaluation

For an evaluation to be successful, it is necessary to include stakeholders in the planning, design, implementation and interpretation of the evaluation (CDC 2001). Campaign initiators and other parties involved need to communicate about the evaluation approach. Because different parties can have different objectives and needs, they should be aware of each other's role and position in the evaluation. For

example, if evaluation is not discussed at the beginning of the campaign, proper data gathering might become impossible and evaluation results invalid (Elliot 1989).

2.2.2.1 *Why involving stakeholders?*

Parties that are not involved in an evaluation may spread doubt and suspicion, which can decrease the value of the evaluation. Stakeholders are much more likely to support an evaluation if they are involved in the whole evaluation process. Considering different perspectives towards the campaign evaluation may also enhance its usefulness.

Also, stakeholder engagement creates a market for the evaluation results, and guarantees that the evaluation results will be used to make a difference in future campaigns.

To optimise the involvement of stakeholders in the campaign evaluation process, it is necessary to understand the stakeholders' interests in or expectations of the evaluation. Box 2.1 gives examples of questions that can help to determine their expectations.

BOX 2.1: Questions that help in determining what each stakeholder will expect from the campaign evaluation.

- What do they want to know from the evaluation?
- How rigorous do they expect the results to be?
- How can you meet their information needs?
- Can you prioritise stakeholder needs due to budget limitations?

Source: Christchurch City Council, 2004

2.2.2.2 *Which stakeholders should be involved?*

The stakeholders who are involved in the campaign evaluation can be the same as those who were involved in the campaign design and implementation. However, it might be recommendable to reconsider their involvement since the scope of both campaign design and evaluation is different and some specific skills (e.g., methodological, statistical) might be required for the evaluation.

To optimise the involvement of stakeholders in the campaign evaluation process, it is necessary to understand the stakeholders' interests in or expectations of the evaluation. The choice of which stakeholders to involve is a function of the purpose of the evaluation and who will use the results (CDC 2001). Sometimes there is no choice, for example when politicians ask for increased attention for a certain road safety problem because of their political agenda. Different stakeholders can have different, yet equally important roles in the campaign and its evaluation.

Evaluation results might be enhanced if priority is given to stakeholders that correspond to the criteria in the following table (note, however, that this list is not exhaustive).

1. Stakeholders that increase the credibility of the evaluation;

2. Stakeholders that are responsible for day-to-day implementation of the campaign activities;
3. Stakeholders who can and will authorise changes to the campaign that are recommended by the evaluation;
4. Stakeholders who will fund or authorise the continuation or expansion of the campaign.

As an example, consider the “Bob” designated driver campaign in Belgium. This campaign is supported by the Belgian brewers' federation and the Belgian federation of insurance companies. Especially the support of the Belgian Brewers' federation in the campaign increases the credibility, as it excludes possible opposite forces that may reduce the effect of the message. Their involvement as stakeholder could be enlarged. In the same vein, the advertising agency that is responsible for the distribution of the message in various media channels (TV spots, posters, etc.) could also be involved to provide information on the number and timing of messages distributed (process information). A third group of stakeholders could include research institutes who are able to estimate the effects of the campaign, and to identify its strong and weak aspects. If they can draw clear conclusions from the evaluation results, these will be accepted more easily since they are made by independent researchers, based on objective evaluation methods. Finally, some campaign sponsors and/or the government (Ministry of Transport) can be stakeholders in evaluation activities. Based on the results of the evaluation, they will decide on their future involvement in the campaign.

2.2.2.3 Implications for the evaluation methodology (Figure 2.2)

- *Introduction*

Campaign initiators and possible involved parties confer with each other about the evaluation approach. Different parties can have different objectives and needs. This might affect the used evaluation methodology to a certain extent.

- *Measurement variables and data collection methods and techniques*

Because of their knowledge, stakeholders can facilitate the evaluation process. They can bring their own expertise (e.g., by ensuring the right questions are asked) and may provide evaluators with a broad range of knowledge (e.g., cultural understanding see also Christchurch City Council 2004). Of course, the input of the stakeholders depends on their interests or expectations of the evaluation. Accordingly, it is possible that the stakeholders will advocate the use of a certain data collection method or technique.

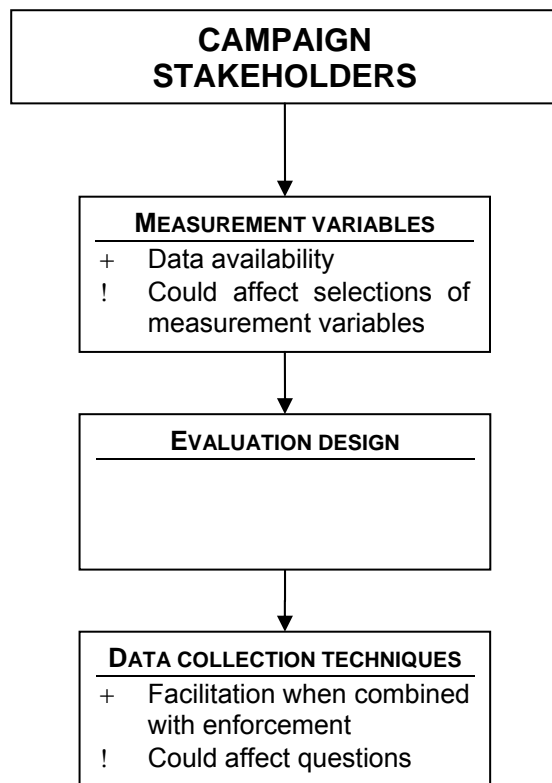
In case of an integrated campaign, it can be especially advantageous to ensure the involvement of stakeholders related to the respective supporting activity (e.g., enforcement).

- *Evaluation design*

Stakeholders' involvement will not have special consequences for the evaluation design, unless they would like to measure the isolated effect of an integrated

campaign. In that case, this will imply the application of stronger designs, more measurements and thus higher costs.

Figure 2.2: Stakeholders implications - summary



Note: (+): positive consequence for the evaluation component, (!): keep in mind

2.2.3 A-priori information

When considering implementing a campaign, it is necessary to execute a profound analysis on beforehand to identify the problematic behaviour and the target group that is most at risk (see also Boulanger, A. et al (2007b) and Delhomme, P. et al, (2009)). This relevant background information can be obtained either by available information such as sources of crash and casualty data and findings of the previous campaigns (problem recognition) or by executing specific knowledge, attitudinal or behavioural research (situational analysis). If there is a lack of existing research, it is advisable to study the target group extensively. It is essential for an effective campaign that its design (e.g. construction of the message/main ideas, the choice of suitable media channels) is approached in a scientific manner.

Example:

Study made by the Société de l'Assurance Automobile du Québec on the evolution of the speed attitudes and behaviours in Quebec between 1991 and 1992.

A study was made in Canada on speed and showed that people believed that the other drivers drive faster than themselves. Another result of this study was that most of the drivers don't respect the speed limits. For several drivers, the allowed speed is not the official speed limit, but the official speed limit added to the police tolerance

(Brault et Letendre, 2003). The design of the campaign was based on this pre-campaign study. According to these results, we can see on which elements the speed campaign should focus.

2.2.3.1 Implications for the evaluation methodology (Figure 2.3)

The stage of development of a campaign is important to set a realistic evaluation focus. Indeed, a campaign that has just been planned will be evaluated differently from a campaign that has been repeated for several years, mainly because the evaluation data that is already available and the experience that has been built up will be completely different. These data can also be seen as the result of a pre-campaign study or a-priori knowledge.

- *Measurement variables and data collection methods and techniques*

The advantage of a-priori knowledge is that these data can be used as baseline information (see next paragraph about evaluation design). In order to compare the before and after measurements, the choice of the after-data collection method is limited to the used method for the baseline data. In theory, for an optimal comparison, the same questions should be asked to the respondents, the same behaviour should be observed on similar locations, and so on. Of course, new measurement variables and more evaluation questions can be added as needed.

The whole procedure can be less time-consuming if detailed information about the evaluation methodology already exists. In such cases, the methodology has been assessed before and improvements can be made.

- *Evaluation design*

This section refers to the existence of data from former campaigns that can be used as before measurements in evaluation of particular current campaigns. In case of campaigns that are part of the long term strategy (e.g., a speeding campaign is run each year for 5 subsequent years while the target group remains the same) an after measurement of the first campaign may serve as a before measurement of the second campaign. Thus, to evaluate these particular campaigns (e.g., the second in the row), only an after measurement is required because pre-campaign data is already at hand. So, when it is possible to use existent data, it is recommended to use evaluation designs with multiple measurements (at least before-after designs) – either single or multiple groups design (depending on the nature of available information). An extra drawback of using after measurements of a campaign in year 1 as before measurements for a campaign in year 2, is that the history threat enlarges when the time period between the before and after measurement becomes larger.

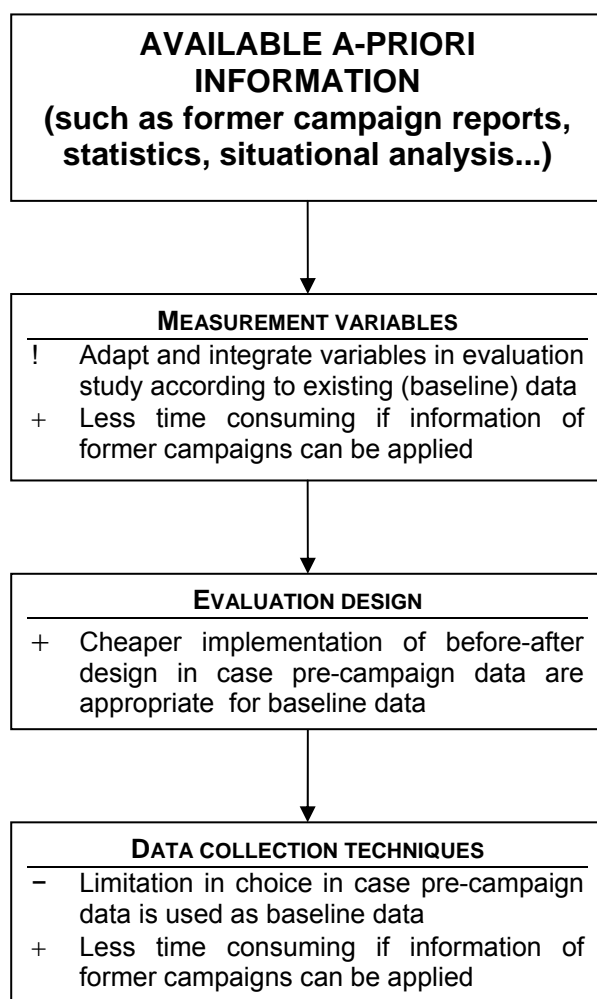
If pre-campaign data exist both for the intervention and the control group, the separate pre-post samples quasi-experimental design may be used to evaluate the current campaign (of course, other attributes of the current campaign should be considered as well). If data are only available for the intervention group and/or the current campaign does not allow the control group to be included in the research design, it is suggested to use the one-group before-after or the one-group before-

during-after research designs (having in mind that the groups of subjects that are measured on each occasion are not the same).

Thus, to summarise the main limitations and opportunities:

- The limitation of exploiting existent data is related to the fact that the (different) groups that are measured on various occasions might be substantially different, which means that observed (in)effectiveness of particular campaign is more the consequence of these differences than the consequence of the campaign itself.
- + History threat can be reduced by randomly selecting subjects in the study sample. If both samples (the one that generated the existent pre-campaign data and the one used to assess the current campaign afterwards) are representative of the target population then there is higher probability that all random fluctuations are ruled out and that the two samples are comparable.

Figure 2.3: A-priori knowledge implications – summary



Note: (+): positive consequence for the evaluation component, (-): negative consequence for the evaluation component, (!): keep in mind

3 Components of the evaluation study

The implementation of the evaluation study itself needs to take into account some important steps. First, it is important to define the success criteria of the road safety campaign and formulate the proper evaluation questions (section 3.1). Together with this step the variables to be measured are defined. Section 3.2 gives an overview of the typical three groups of measurement variables that are used in evaluation studies of road safety campaigns. In order to achieve a reliable evaluation study, you need to choose the proper research design for producing the needed evaluation data. This component is discussed in section 3.3. The CAST consortium decided on some minimum standards for evaluation to ensure some satisfactory level of evaluation data, as it is not always feasible in practice to carry out the best and most ideal evaluation study. These CAST standards are given in section 3.4.

3.1 Objectives of the evaluation study

According to Annex 1 the results of a campaign can be considered from two points of view. An evaluator considers the *measurable outcomes* of a campaign as the main study topic, while a campaign designer focuses on *objectives and goals*. Although in the end, these two views match perfectly with each other. For that reason, we will keep it simple in this evaluation tool by not making any distinctions between different types of campaign objectives.

The objectives of the evaluation study are inseparably linked with the objectives of the road safety campaign. Objectives of the campaign address specific aspects of the problem and thus specify in detail what a campaign is expected to achieve. Hypotheses on the expected campaign results will be formulated and analysed in the evaluation study.

Having clearly defined campaign objectives is very important for evaluating the success of the campaign. It is important to specify beforehand what could be considered as a success or a failure. At least in theory objectives should be clearly defined in terms of (a) desired communication or behavioural effect, (b) the public who is suppose to reach the effect, (c) the expected level of attainment or accomplishment (e.g., 20% increase in awareness) and (d) the time period (e.g. campaign period) when the effect is expected to occur. But if an objective is declared purely descriptive (without incorporating the success criteria) as for example 'increase the seat belt wearing rates' then it is hard to say at the end what kind of an increase is sufficient for a campaign to be considered successful. The lack of clearly defined objectives seems to be one of the major weaknesses of the collected and analysed evaluation reports (see reference D2.1). In other words, a proper evaluation study starts with a clear definition of the campaign objectives in order to formulate the exact objectives of the effectiveness study. Nevertheless, if behavioural change is not stated as a direct campaign objective, it should always be measured.

However, it is reasonable to assume that the final effect of a campaign will be a combination of several factors that can have an impact on human behaviour. Also, intermediate outcome measures (e.g., exposure of the campaign) are important to find improvements in case that the goals are not reached. These outcomes will allow distinguishing between a campaign that was not set up properly, and a well designed campaign that was indeed not effective.

Finally it is advisable to ensure that the stakeholders have the opportunity to provide input to defining evaluation objectives or formulating evaluation questions.

3.1.1 Evaluation questions

Evaluation studies are undertaken to answer questions about the effectiveness of the executed campaign for example. So it is important to understand the research problem. Therefore it is necessary to know the goal of the campaign (in detail) and what questions we hope to answer.

What makes a good evaluation question? (Rossi, 2004)

The form that evaluation questions should take is shaped by the functions they must perform. Their principal role is to focus the evaluation on the objectives of the campaign evaluation and the areas at issue for key decision makers and stakeholders. Good evaluation questions must first of all be reasonable and appropriate. Furthermore, evaluation questions must be answerable: they must involve performance dimensions that are sufficiently specific, concrete, practical, and measurable that meaningful information can be obtained about their status. To deal with appropriate dimensions, the evaluator can work with relevant stakeholders to scale down and focus the evaluation questions.

There are two complementary ways for an evaluator to assess how appropriate and realistic a candidate evaluation questions is. The first is to examine the questions in the context of the actual campaign activities related to it. The second way to assess whether candidate evaluation questions are reasonable and appropriate is to analyse them in relationship to the findings reported in applicable and relevant literature or former campaign reports. Furthermore it is very easy to formulate an unanswerable evaluation question without realising it. This may occur because the terms used in the question are actually ambiguous and vague, or some questions may invoke issues for which there are so few observable indicators, some questions lack sufficient indication of the relevant criteria to permit a meaningful answer (is this program successful?), and finally some questions may be only answerable with more expertise, data, time, or resources than available to the evaluation.

There are some elements that can be checked to affirm that the question is an answerable one:

1. define the target group;
2. identify the specific measurable characteristics and the cut-off values that represent the greatest need;
3. give an example of the evaluation finding that might result;
4. stipulate the evaluative criteria (at least 90% should...);
5. have the agreement of the evaluation sponsors and other stakeholder on the proposed success criteria.

If these conditions can be met and, in addition, the resources are available to collect, analyse, and report the applicable data, then the evaluation question can be considered as answerable.

3.2 Measurement variables to define the success (or not) of a campaign?

A measurement variable represents a measure (related to the outputs of a campaign) of the success or failure of a road safety campaign, relative to the aim of the campaign. It is thus crucial to consider the objectives of a road safety campaign when deciding which variables to assess in order to measure the effect of campaigns. Moreover, if a campaign is designed to affect/target variables in a particular model (e.g., the Theory of Planned Behaviour, see also Box 3.1), one needs to assess the variables specified in the model. The choice of measurement variables will among others depend on the scope of the campaign and the resources available. Nevertheless, it is important to decide which variables to assess before actually implementing the campaign. In this section guidelines for choosing measurement variables when evaluating the effect of road safety campaigns will be provided. Since the most frequently applied measurement variables were defined and described in Boulanger et al (2007a, see chapter 3), this information will only be repeated in this chapter when necessary. Some additional variables will be described here.

Typically three groups of measurement variables are used in evaluation studies of road safety campaigns, and in the literature concerning road safety: self-report measures, observed behaviour and accident statistics.

These three groups correspond to the classification of the data collection methods. There are several ways to classify these methods but we choose to focus on the fact that data can be achieved by asking, observing or document analysis. When we set out to collect information, it is important to keep in mind the questions we hope to answer on the basis of the resulting data (Peck, 2005). The evaluator's choice of data collection method should be driven by the goals of the evaluation and the availability of resources (Logan, 2006). Whenever possible, more than one source of data should be pursued. In any case, the evaluator should make it clear what types of data are needed and what evaluation questions they can address.

The approaches used for each type of evaluation will vary (GRPS, 2007). Both qualitative and quantitative approaches can be used within the design of an evaluation.

Qualitative research tends to involve detailed, verbal description of characteristics, cases and settings to explain reasons underlying various behavioural patterns. Certain facets of a campaign are difficult to capture and quantify – cross-cultural misunderstandings and target group perceptions for example (Logan, 2006). Qualitative research has certain advantages (Wimmer, 1994): viewing behaviour in a natural setting without artificiality, increasing understanding of the phenomenon under investigation and examination of complex phenomena without relying on structured data collection necessitated by quantitative design. There are, however, some disadvantages associated with these methods as well, for example small

sample sizes and low reliability of the data. Finally, it is crucial that qualitative research is carefully planned to ensure that the focus will be on key issues.

Quantitative research is equally often concerned with road users evaluations of various aspects, for example their attitude towards seat belt wearing etcetera, and to what extent road users drive within speed limits, wear seat belts or not (never – always). Quantitative research generally uses numbers to communicate this amount. One advantage of quantitative research is that the use of numbers allows greater precision in reporting results (Wimmer, 1994). Another advantage is the use of powerful methods of statistical analysis.

To sum up, a data collection method states what type of data are obtained (which group of measurement variables), how they will be collected (data collection techniques), the consequent analysis and the interpretation of the results. Hereunder the three methods together with the respective measurement variables (data) and possible data collection techniques will be illustrated.

3.2.1 Method of asking – self-reported measures

The first method, namely the method of asking (or also called surveying) implies communication between the investigators and the subject of study. Survey researchers sample respondents who answer the same questions. Surveys are descriptive methods to gain information regarding the subjects. A survey may focus on opinions or self reported factual information depending on its purpose. The choice of a method of asking is useful when studying behaviour(s), opinions, attitude etcetera, which cannot be directly observed and when the purpose is to investigate a large number of subjects.

Several data collection techniques are used such as questionnaires, interviews, focus groups etcetera. These techniques imply an (active) communication between the investigator and the studied subjects. In this section we will introduce shortly these data collection techniques. All these techniques generate information fairly quick and simple but of course each medium (face-to-face, post, telephone, and internet) has a different price tag. An overview of the advantages and disadvantages can be found in Annex 2 but more detailed information is provided in Boulanger, A. et al (2007a).

Questionnaires are the most widely used data gathering techniques in many research fields. A questionnaire is a research instrument consisting of a series of questions and other prompts for the purpose of gathering information from respondents. Questionnaires can be paper or computer based. Mostly the questions are administrated by the respondent themselves. But of course an interviewer, on the telephone or in person, can also administer the questionnaire.

Questionnaires, in an evaluation context, are systematic ways of collecting primary data – quantitative, qualitative or both – concerning a campaign and from individuals (or from other sources) associated with the intervention. This is useful in an outcome evaluation to collect baseline data on knowledge, attitudes and behaviour of the target group (and the control group, when used in the evaluation). Questionnaires are also useful for gathering demographic information about participants such as age, gender and ethnicity.

Interviews are question and answer sessions with an individual person. It is a conversation between two or more people (the interviewer and the interviewee). The general goal is to elicit rich, detailed material that can be used in analysis (Lofland & Lofland, 1995). Depending on the degree of structure, an interview yield quantitative or more qualitative data.

Example: Filter-questions

These 2 questions are possible filter-questions meaning in case of 'no' the interview can be stopped.

Do you have a full car driving licence (or permit)?	Yes (continue) No (stop interview)
Have you driven a car in the last 12 months?	Yes (continue) No (stop interview)

Generally, in-depth interviews can be used at any stage of the evaluation process. Interviews are useful at the beginning of a study (pre-campaign research) to explore the issues and develop survey materials. But they are also carried out with individuals to explore their views of the campaign. Follow-up interviews are also valuable tools when studying the impact of a program over a period of time. It is an excellent approach for obtaining time-sequenced data and for following-up on users' feelings and experiences.

Focus group interviews represent a productive and time-efficient variant of in-depth interviewing that combines elements of both interviewing and participant observation. They work on the same basis as interviews but ask open-ended questions to a group of people. They multiply the number of respondents and opinions of several people are received at the same time. The technique inherently allows observation of group dynamics, discussion, and firsthand insights into the respondents' behaviours, attitudes, languages, opinions of groups, etcetera. A special case of focus groups are the online focus groups. They use an Internet chat room as the venue for the discussion.

Focus groups are practical in exploratory research or to generate new ideas for hypotheses, questionnaire items and the interpretation of results. First of all, focus group interviews are very useful as a pre-testing tool in the concept development stage of the communication process. They provide insights into target audience beliefs on an issue allow campaign planners to obtain perceptions of message concepts, help trigger the creative thinking of communication professionals and come in handy during the formative phases when campaign planners are beginning to formulate program goals and objective. Secondly, focus groups are opportune during the needs of assessment. Similarly, they can be convened during a process evaluation to discover the strengths and weaknesses of a campaign. They could also be used to collect data on project outcomes and impact during the summative phase of evaluation.

A key informant is a person (or a group of persons called advisory groups) who has unique skills or professional background related to the intervention being evaluated, is knowledgeable about the project participants, or has access to other information of

interest to the evaluator. Key informants can be surveyed or interviewed individually or through focus groups.

Essentially, this technique consists of asking experts in a given subject area for their opinions on specific evaluation issues. Evaluators can use this information to determine, predict and assess campaign outcomes.

Self-reported measures (i.e., reach, recognition, recall, social cognition variables and behaviour) are possibly gained by all these data collection techniques, classified under the term method of asking. Two main categories of self-reported measurement can be distinguished: (1) comprehension and likeability of campaigns and (2) social cognitive variables and behaviour. Background information of the respondent can be seen as a third category.

Likeability and comprehension of campaigns

A basic premise for a road safety campaign to be effective is that the relevant target group actually has been exposed to the campaign. Results concerning the reach of the campaign are often reported in terms of percentages of the target group reached. Moreover, if campaigns are positively evaluated by the target group it is likely to have a stronger impact. In other words, if the campaign is perceived as interesting, meaningful and personally relevant the message is likely to be more thoroughly elaborated and thus the likelihood of permanent attitude change increases. However, measures of awareness etcetera do not provide sufficient information to fully understand effects on road safety as they do not represent actual changes in behaviour or the antecedents of behaviour (e.g., attitudes, intentions).

Advise:

In order to know whether or not road safety campaigns have reached the relevant target group it is necessary to include measures indicating the actual reach of the campaign when evaluating campaigns, for example “Have you heard about road safety campaign X” (Yes/No), “Have you seen campaign X on television” (Yes/No).

To determine whether the campaign was evaluated positively or not by the target group, it is necessary to assess the likeability of road safety campaigns in evaluation studies. Likeability of campaigns can for example be measured using one question “What do you think about the ad”. It is important to capture several aspects of the target groups’ evaluation of the campaign (likeability, personal identification etc.). Thus, it is advisable to use a number of 7-point scales to capture these issues (see underneath examples)

Example: 7-point scale questions to assess the likeability

I disliked the ad: _____: _____: _____: _____: _____: _____: _____: I liked the ad

The ad is definitely _____ The ad is definitely
not for people like me : _____: _____: _____: _____: _____: _____: _____: for people like me
meaningless: _____: _____: _____: _____: _____: _____: _____: meaningful
not at all interesting: _____: _____: _____: _____: _____: _____: _____: very interesting

The goal of the message execution strategy is to develop messages that will capture the attention of the target audience and lead them to adopt the safe behaviour. In order to achieve this goal the message needs to be; believed by the audience (credible), possible to achieve and honest (trustworthy), used repeatedly (consistent), easy to understand (clear), able to generate change (persuasive), relevant to the person (relevance) and it needs to be appealing (attractive). In order to assess these aspects, the below questions (7-point scale) can be used.

Example: 7-point scale questions about the campaign message

What do you think about the main message of the campaign?

- highly credible : ____ : ____ : ____ : ____ : ____ : ____ : not credible
- clear message : ____ : ____ : ____ : ____ : ____ : ____ : unclear message
- trustworthy : ____ : ____ : ____ : ____ : ____ : ____ : not trustworthy
- consistent : ____ : ____ : ____ : ____ : ____ : ____ : inconsistent
- informative : ____ : ____ : ____ : ____ : ____ : ____ : not informative
- annoying : ____ : ____ : ____ : ____ : ____ : ____ : not annoying

What do you think about the design of the campaign?

- ugly : ____ : ____ : ____ : ____ : ____ : ____ : beautiful
- eye-catching : ____ : ____ : ____ : ____ : ____ : ____ : not eye-catching
- appealing : ____ : ____ : ____ : ____ : ____ : ____ : not appealing

Do you remember to have noticed this seat belt campaign in the past 3 months (after measurement)?

- Yes, I am sure of that
- Yes, but I am not sure
- No
- I don't know

By which means have you noticed the seat belt campaign in the past 3 months (after measurement)?

- TV-commercial
- Radio-commercial
- Billboards alongside national roads and highways
- Billboards alongside local and provincial roads
- Internet (sites and banners)
- Flyer
- Posters at schools
- Articles in newspapers and magazines
- Items in TV- and radio programs
- Other...

Social cognitive variables and behaviour

A number of psychological mechanisms are important predictors of risky road traffic behaviour (e.g., Conner, Smith & McMillan, 2003; Evans & Norman, 2003; Marcil,

Bergeron & Audet, 2001). Thus, by influencing those mechanisms there is the potential to change road users' behaviour in traffic (Parker et al., 1996). The aim of most mass media campaigns addressing road safety is to change road users' risky behaviour in traffic and/or factors which have been found to predict such behaviours. A wide variety of psychological concepts are available to evaluate the effect of road safety mass media campaigns (see Fylan, Hempel, Grunfeld, Conner & Lawton, 2006, for a review). Existing reviews suggest that three theoretical models have the best practical value: the Protection Motivation Theory (Rogers, 1983), the Social Cognitive Theory (Bandura, 1986), and the Theory of Planned Behaviour (Ajzen, 1991).

Advise:

Based on existing reviews of the literature concerning the evaluation of road safety campaigns, and empirical evidence related to the above mentioned theoretical models, it is suggested to measure the following variables¹: attitudes, subjective and descriptive norms, perceived behavioural control and intentions along with behaviour (Fylan et al., 2006)². Attitude, subjective norm, perceived behavioural control, and intention are usually assessed directly by means of standard scaling procedures. When developing the scales, it is important that all variables (including possible extension variables) are directly compatible with the behaviour in terms of *action* (e.g. exceeding the speed limit while driving), *target* (e.g., car), *context* (e.g., in an urban area) and *time* (e.g., during the next month) are directly compatible with the target behaviour (see Ajzen, 2002, for further details). It is also preferable to describe a scenario concerning the target behaviour, perhaps even with a picture, so all the respondents are confronted with the same situation.

The above variables do of course not represent all possible variables that might influence road user behaviour. Moreover, it has been shown that the above models might benefit from being extended with a number of variables. Particularly, it has been shown that past behaviour and personal/moral norm have made useful contributions to the prediction of intention and behaviour, beyond the impact of the components specified in the TPB (see Conner & Armitage, 1998, for review). Thus, guidelines on how to measure these variables will be provided along with the guidelines on how to measure the TPB constructs. Furthermore, beliefs play a central role in the theory of planned behaviour (Ajzen, 1991, 2002). By measuring beliefs, we can, theoretically, gain insight into the underlying cognitive foundation of certain attitudes, subjective norms and perceptions of behavioural control. This information can be valuable for designing road safety campaigns. Therefore, guidelines on how to measure beliefs will also be provided.

Attitude is a central construct in all models developed to explain and predict human health and risk-related behaviour, and is an important indicator of road users' future behaviour (Armitage & Conner, 2001). Many road safety campaigns also aim to change road users' attitudes towards speeding, seatbelt use, drink driving etcetera.

¹ Which measurement variables that are assessed in a campaign evaluation will of course depend on the aim of the campaign, i.e., whether it is designed to change attitudes, knowledge or behaviour.

² Ajzen (2002) has provided detailed guidelines on how to measure the above mentioned variables – see website <http://www.people.umass.edu/ajzen>

Example: Bipolar items

Attitudes, for example towards speeding can be measured with several bipolar items using the stem “During the next month, exceeding speed limits when driving a car in an urban area will for me be...”:

bad : ____ : ____ : ____ : ____ : ____ : ____ : ____ : good

not useful : ____ : ____ : ____ : ____ : ____ : ____ : ____ : useful

harmful: : ____ : ____ : ____ : ____ : ____ : ____ : ____ : beneficial

foolish: : ____ : ____ : ____ : ____ : ____ : ____ : ____ : wise

unpleasant: : ____ : ____ : ____ : ____ : ____ : ____ : ____ : pleasant

un-enjoyable: ____ : ____ : ____ : ____ : ____ : ____ : ____ : enjoyable

Social influence, i.e., in terms of social pressure to perform or not to perform certain road traffic behaviours (cf. subjective norms) or through observing what significant others are doing (cf. descriptive norms) have been shown to be important predictors of road users’ intentions to perform risky road behaviours (Moan & Ulleberg, 2007; Ravis & Sheeran, 2004). Thus, in instances where the aim of the campaign is to influence road users’ perceptions of social pressure it is important to assess these variables in order to determine the success or failure of the campaign. People who are significant to them can include people older than themselves, same age and sex, same age opposite sex, parents/children, spouse/partner but also the church, the police and other drivers.

Example: Illustration of the format measures of subjective norm can take

Most people who are important to me think that

I should: ____ : ____ : ____ : ____ : ____ : ____ : ____ : I should not
exceed speed limits while driving a car in an urban area during the next month

It is expected of me that I exceed speed limits while driving a car in an urban area
during the next month
extremely likely : ____ : ____ : ____ : ____ : ____ : ____ : ____ : extremely unlikely

The people in my life whose opinions I value would
approve : ____ : ____ : ____ : ____ : ____ : ____ : ____ : disapprove
of me exceeding speed limits when driving a car in an urban area during the next
month

Example: How to measure descriptive norms

Most people who are important to me will exceed speed limits when driving a car in an urban area the next month

completely true :___:___:___:___:___:___:___: completely false

Many people like me keep within speed limits when driving a car in an urban area during the next month

extremely likely :___:___:___:___:___:___:___: extremely unlikely

The people in my life whose opinions I value

keep within speed limits :___:___:___:___:___:___:___: exceed speed limits when driving a car in an urban area during the next month

In addition to the distinction between subjective and descriptive norms, Cialdini, Reno and Kallgren (1990) argued for the inclusion of a third type of norm, i.e., a personal norm (or moral norm) which can be defined as an individual's internalized moral standards. In other words moral norms constitute an individual's perception about moral correctness of performing a specific behaviour (Sparks, 1994). In the context of the TPB moral norms have been included as an additional predictor of intentions with some success across a wide range of behaviours (e.g., Conner & Armitage, 1998; Manstead, 2000, for reviews), including road safety behavior (e.g., Parker et al., 1995). Conner and Armitage (1998) found that in nine out of ten studies, moral norm was a significant predictor of intentions after the TPB variables were controlled for.

Example: How to measure personal or moral norms

It is morally wrong of me to exceed the speed limits while driving a car in an urban area during the next month

completely disagree :___:___:___:___:___:___:___: completely agree

I feel guilty if I exceed the speed limits when driving a car in an urban area during the next month

completely disagree :___:___:___:___:___:___:___: completely agree

I get a bad conscience if I exceed the speed limits when driving a car in an urban area during the next month

completely disagree :___:___:___:___:___:___:___: completely agree

Whether or not the individual feels that s/he is capable of performing a certain behaviour and whether performing the particular behaviour is entirely under his or her control (e.g., keeping within speed limits) is strongly related to both the motivation to perform the behaviour and the actual performance of that behaviour (Armitage & Conner, 2001; Moan & Ulleberg, 2007). Thus, if road safety campaigns succeed in influencing road users' perceptions of control, they might also lead to significant changes in behaviour.

Example: Items to capture the respondents' perceived capability of performing the behaviour

How much control do you believe you have over the speed when driving a car in an urban area during the next month?

no control : ____ : ____ : ____ : ____ : ____ : ____ : complete control

It is mostly up to me whether or not I exceed the speed limits when driving a car in an urban area the next month

strongly agree : ____ : ____ : ____ : ____ : ____ : ____ : strongly disagree

Source: Ajzen, 2002

Example: Typical items used to tap controllability

How much control do you believe you have over the speed when driving a car during the next month?

no control : ____ : ____ : ____ : ____ : ____ : ____ : complete control

It is mostly up to me whether or not I exceed the speed limits when driving a car the next month

strongly agree : ____ : ____ : ____ : ____ : ____ : ____ : strongly disagree

Accessible or salient beliefs.

Beliefs play a central role in the theory of planned behaviour. Although beliefs are not specified in all theoretical models (only the Theory of Planned Behaviour), they are very useful to measure. The reason is that by measuring beliefs, we can, theoretically, gain insight into the underlying cognitive foundation, i.e., we can explore why people hold certain attitudes, subjective norms and perceptions of behavioural control. This information can be valuable for designing road safety campaigns. It is important to note, however, that this explanatory function is assumed only for *salient beliefs*, i.e., beliefs that are readily accessible in memory. Pilot work is required to identify accessible behavioural, normative, and control beliefs. *Behavioural beliefs* are assumed to provide cognitive and affective foundation for attitudes, *normative beliefs* provide the foundation for subjective norms, and *control beliefs* the foundation for perceptions of control (Ajzen, 1991, 2002).

Eliciting salient behavioural beliefs.

Through pilot work (e.g. situational analysis) respondents are given a description of the behaviour and are asked a series of questions in order to identify salient behavioural beliefs (i.e. beliefs underlying attitudes):

- What do you believe are the advantages of you exceeding the speed limits when driving a car in an urban area during the next month?
- What do you believe are the disadvantages of you exceeding the speed limits when driving a car in an urban area during the next month?
- Is there anything else you associate with you exceeding speed limits when driving a car in an urban area during the next month?

Measuring behavioural beliefs.

Two questions should be asked with respect to each of the outcomes generated.

Example: How to measure behavioural beliefs:

Assumed that one of the disadvantages elicited in the pilot study is that the behaviour (i.e., exceeding the speed limit) can increase the risk of having the drivers licence withdrawn. Belief strength and outcome evaluation are then assessed as follows:

- *Behavioural belief strength (b):*

Exceeding the speed limits while driving a car in an urban area during the next month will increase the likelihood of having my drivers' licence withdrawn
extremely unlikely :____:____:____:____:____:____:____: extremely likely

- *Outcome evaluation (e):*

Having my drivers' licence withdrawn is
extremely bad :____:____:____:____:____:____:____: extremely good

The belief strengths and outcome evaluations for the different accessible beliefs provide substantive information about the attitudinal considerations that guide people's decisions to engage or not to engage in the behaviour under consideration (see Ajzen, 2002, for further details).

Eliciting salient normative beliefs.

The following questions can be asked to elicit the identity of relevant referent individuals and groups that are readily accessible in memory:

- Are there any individuals or groups who would approve of you exceeding the speed limits when driving a car in an urban area during the next month?
- Are there any individuals that would disapprove of you exceeding the speed limits when driving a car in an urban area during the next month?
- Are there any other individuals or groups who come to mind when you think about exceeding the speed limits when driving a car in an urban area during the next month?

Measuring normative beliefs.

Assessing normative beliefs (i.e. beliefs underlying subjective norms) follows the same logic involved in measuring behavioural beliefs. Two questions should be asked with respect to each referent generated.

Example: How to measure normative beliefs

Assume that *my family* is one of the accessible referents.

- *Normative belief strength (n):*

My family thinks that
I should: ____:____:____:____:____:____:____: I should not
exceed the speed limits while driving a car in an urban area during the next month

- *Motivation to comply (m):*

When it comes to exceeding the speed limits while driving a car in an urban area during the next month, how much do you want to do what your family thinks you should do?

not at all :____:____:____:____:____:____:____: very much

Measures of normative belief strength and motivation to comply with respect to each accessible referent offer a snap shot of perceived normative pressures in a given population (see Ajzen, 2002, for further details).

Eliciting salient control beliefs.

To generate a list of factors that might facilitate or impede performance of the behaviour (i.e. perception of control), the following questions can be asked:

- What factors or circumstances would enable you to exceed the speed limits when driving a car in an urban area during the next month?
- What factors or circumstances would make it difficult or impossible for you to exceed the speed limits when driving a car in an urban area during the next month?
- Are there any other issues that come to mind when you think about the difficulty of exceeding the speed limits when driving a car in an urban area during the next month?

Measuring control beliefs.

Two questions are asked with respect to each accessible control factor generated.

Example: How to measure control beliefs

Assume that one of the accessible control factors has to do with saving time getting to work in the morning.

- *Control belief strength (c):*

I expect that I will be short of time in the morning when driving to work during the next month

strongly disagree :____:____:____:____:____:____:____: strongly agree

- *Control belief power (p):*

Being short of time in the morning when driving to work would make it

much more difficult :____:____:____:____:____:____:____: much easier

for me to exceed the speed limits when driving a car in an urban area during the next month

Examination of the average strength and power of different control beliefs provides a picture of the factors that are viewed as facilitating or impeding on performance of the behaviour (see Ajzen, 2002, for further details).

Behavioural intentions mark the end of a motivational phase of decision making that many social cognition models focus upon, and intentions typically emerge as the

strongest predictor of future behaviour, including road safety behaviour (Armitage & Conner, 2001; Moan & Ulleberg, 2007). Moreover, recent evidence suggests that changes in intentions do in fact lead to behavioural changes (Webb & Sheeran, 2006). The results presented in this review thus suggest that behavioural intentions are important measurement variables when evaluating the effect of mass media campaigns designed to affect road users' behaviours. Several items are used to measure behavioural intentions, and they are typically measured on 7-point scales (cf. Ajzen, 2002).

Example: How to measure behavioural intentions

I plan to exceed the speed limits when driving a car in an urban area during the next month

strongly disagree : ____ : ____ : ____ : ____ : ____ : ____ : strongly agree

During the next month, I intend to exceed the speed limits when driving a car in an urban area

extremely unlikely : ____ : ____ : ____ : ____ : ____ : ____ : extremely likely

I will exceed the speed limits when driving a car in an urban area during the next month

definitively true : ____ : ____ : ____ : ____ : ____ : ____ : definitively false

In instances where the researcher wants to assess behaviour in a large sample, and when the aim is to understand the mechanisms that influence behaviour or behavioural change (underlying attitudes or motivation) in relation to road safety campaigns, self-report measures is often the only option. To obtain a reliable self-report measure of behaviour, it is desirable to use more than one question. One possible approach of evaluating the effect of road safety campaigns is to combine exact numerical reports, rough numerical estimates, and rating scales and estimate the internal consistency of these three measures (see Ajzen, 2002, for further details). The response scale could take an exact numerical format, where participants are asked to indicate the number of days on which they performed the behaviour in question.

Example: Precise estimation of self-reported behaviour

On how many days in the course of the past month have you exceeded the speed limits when driving a car in an urban area? ____

Example: Less precise estimation of self-reported behaviour

In the course of the past month, how often have you exceeded the speed limits when driving a car in an urban area?

- every day
- almost every day
- most days
- on about half the days

- a number of times, but less than half
- a fewer times
- never

In the course of the past month, how often have you exceeded the speed limit when driving a car in an urban area?

never : ____ : ____ : ____ : ____ : ____ : ____ : nearly all the time

I have exceeded within the speed limit while driving a car in an urban area the past month

strongly disagree : ____ : ____ : ____ : ____ : ____ : ____ : strongly agree

Past behaviour has been shown to predict intention and subsequent behaviour, after accounting for the variables theoretical models such as the theory of planned behaviour (see Conner & Armitage, 1998, for review), including risky road behaviours (e.g., Conner et al., 2003; Elliot et al., 2003). Thus, what road users have been doing in the past might give a good indication of what they will do in the future.

The aim of a large fraction of campaign evaluations is to examine changes in behaviour, and mostly this is done by examining changes in self-reported behaviour. This requires a before and after measurement of behaviour as a minimum. The first measurement of behaviour thus represents a measure of past behaviour and should be measured in the same way as behaviour at the second wave of the data collection (see examples above).

Behaviour can also be a matter of habit, i.e., in some instances behaviour may be performed without a great deal of reasoning and in a more automatic fashion (see Verplanken & Orbell, 2003, for further details). Verplanken and Orbell (2003) developed a 12-item index in order to measure habit and all items should preferably be assessed in order to tap the important aspects of this construct. Some of the items that can be used to assess habitual behaviours are listed below.

Example: Items to assess habitual behaviour

The behaviour (e.g., exceeding speed limits while driving a car in an urban area) is something I do automatically:

completely disagree : ____ : ____ : ____ : ____ : ____ : ____ : completely agree

The behaviour (e.g., exceeding speed limits while driving a car in an urban area) is something I find hard *not* to do:

completely disagree : ____ : ____ : ____ : ____ : ____ : ____ : completely agree

The behaviour (e.g., exceeding speed limits while driving a car in an urban area) is something I do without really thinking about it:

completely disagree : ____ : ____ : ____ : ____ : ____ : ____ : completely agree

BOX 3.1: The development of a TPB-based road safety campaign in practice.

Sutton (2002) suggested that if the goal is e.g., to make young male drivers keep within the speed limits while driving a car in an urban area in the following month, one should do as follows when developing and evaluating a road safety campaign:

1. Conduct an elicitation study in a sample of the target group to identify the *modal salient beliefs* which are relevant for the behaviour and the given population. Those beliefs that are elicited first in response to open-ended questions such as “What do you see as the advantage of keeping within speed limits while driving a car in an urban area during the next month?” are assumed to be salient for the individual. The beliefs elicited most frequently are regarded as the modal salient beliefs.
2. Conduct a study in a new sample of young male drivers who in which all the TPB variables, including modal salient beliefs, should be assessed as described by Ajzen (1991, 2002).
Based on the data from this study, intention is regressed on attitude, subjective norm and PBC in order to estimate the relative contribution of the three determinants. The findings are used to decide whether to target the attitudinal component only, or all three components
3. The same data should be used to identify the beliefs that best discriminate between intenders and non-intenders (or those who subsequently acted or did not act).
4. The researcher develops an intervention designed to change the key beliefs identified. To evaluate the intervention a TPB-study in the target population needs to be conducted.

Background information

One type of background variables are the socio-demographic variables. Socio-demographic information about the respondents in an evaluation study should be gathered in order to determine whether the data was in fact collected among the target group (and control group), and in order to make analyses of subgroups possible (e.g., comparison of men and women).

As a minimum requirement, information about age and gender should be gathered. However, depending on the aim of the campaign and the aim of the evaluation, it might also be useful to collect data concerning the educational level and income among the participants. Although road safety campaigns might be developed to influence the behaviour of all road users in a particular country or region, there will always be some groups of road users that represent a greater risk than others (e.g., young male drivers). Thus, data related to socio-demographic characteristics will enable researchers to examine whether road safety campaigns are more effective in influencing some groups than others.

Other background information referred to as personal contextual data could also be gathered. Topics as exposure to traffic, driving experience, involvement in accidents might be important when determining the effectiveness of road safety campaigns, and whether campaigns have more influence on some road user groups than others. Thus, such information enables researchers to better judge and interpret the results in an evaluation study.

Example: Commonly used questions to assess relevant background information

I am a

- Male
- Female

What was your age on your last birthday? _____

In total about how many kilometres have you driven in the last 12 months? ____ kilometres

In the last 3 years, how many accidents have you been involved in, as the driver of a vehicle, in which someone, including yourself, was injured and received medical attention? ____ accidents

In the last 3 years, how many damage only accidents have you been involved in, as the driver of the vehicle? ____ accidents

How many years car driving experience have you had? ____ years

Is the vehicle you normally drive owned by...?

- yourself
- another member of your family
- your employer/ or employed by your employer
- a friend
- a hire or leasing company

What is your occupation?

- self employed
 - farmer, fisherman
 - professional lawyer, accountant etc.
 - business-owner of shop, craftsman, proprietor
- employed
 - manual worker
 - white collar, office worker
 - middle management, trainee
 - executive, top management, director
- not employed
 - retired
 - housewife, not otherwise employed
 - student, military service
 - unemployed

What is your gross yearly income?

- 0 – 12500 Euro
- 12500 – 25000 Euro
- 25000 – 37500 Euro
- 37500 – 50000 Euro

- 50000 – 62500 Euro
- 62500 – 75000 Euro
- 75000 – 100000 Euro
- More than 100000 Euro

Which of the following applies best to you at the moment?

- single
- living as married
- married
- separated or divorced
- widowed

What level of education did you achieve?

- primary school
- secondary school
- further education
- none

How would you describe the area in which you live?

- rural/village
- small town
- suburban/city outskirts
- urban/city/large town

Source: Cauzard, 2004

Note: The question about Gross Yearly Income was included as an additional measure beyond the questions suggested by Cauzard (2004).

Example: How to acquaint information about the exposure to traffic

At this moment, do you have a driving licence or a provisional driving licence for driving a car? yes /no

How often do you drive?

OR

How frequent do you use your car?

- | | |
|--|---|
| <ul style="list-style-type: none"> ○ every day ○ 3-5 times a week ○ 2-5 times a week ○ once a week ○ less often | <ul style="list-style-type: none"> ○ I use the car on daily basis ○ I regularly use a car ○ I use a car once in a while ○ I never use a car |
|--|---|

How many kilometres do you usually drive per year?

- less than 10 000 km
- between 10 000 and 20 000 km
- between 20 000 and 30 000 km
- more than 30 000 km
- (I don't know)

In the last 3 years, have you been fined, or punished in any other way for not wearing your seat belt?

- no

- yes, only fined
- yes, fined and/or other penalty

On a typical journey, how likely is it that you will be checked for alcohol by the police?

very unlikely :___:___:___:___:___:___:___: very likely

In the past 3 years, how many times were you checked by the police for alcohol?

- never
- only once
- more than once

3.2.2 Method of observing – observed behaviour

The second data collection method is called observation. Observational research (or field research) involves the direct observation of phenomena in their natural setting and is widely used both in quantitative and qualitative research. The main characteristic of this approach is that the investigator does not interfere with the subject(s) under study. It is used to carefully observe and record a given phenomenon, for example speeding behaviour, often during a long period of time.

Examples of data collection techniques used in observational methods are: on-site observing, checklists with well defined behaviour, video recording, participating observation and so forth. In all techniques the investigator records actual behaviour of the studied subjects but people may or may not know that they are being observed. The observer may also engage in roles varying from a non-participant to a complete participant. The technique of on-site observations is a common used data collection tool in the road safety field. The technique involves on-site visits to locations to observe various activities and the researcher takes field notes of the behaviour and activities of individuals. In these field notes, the researcher records, in an unstructured or semi-structured way, activities at the research site but also the context in which events occur.

Several studies have demonstrated that there is a strong correlation between observed behaviour, for example speeding, driving under the influence of alcohol, lack of seat belt use, and injuries (cf., Cummings, Wells & Rivara, 2003; Joint Transport Research Centre, 2006; Kelly, Darke & Ross, 2004). One way to determine the effect of road safety campaigns is to examine the proportion of road users who wear seat belts, the fraction of drivers who exceed speed limits, and the rates of drivers who are intoxicated by alcohol. In relation to seat belt use, counts may be made at a number of sites (e.g., when cars stop at a red light) before and after road safety campaigns have been implemented. Speed (within, below or above the speed limit) might be measured when cars drive passed a certain point in a road section. And alcohol blood concentration might be assessed by stopping drivers. However the latter is mostly only possible with the help of the police. The observations of road user behaviour before and after the campaign period may thus be compared to assess the effect of the road safety campaign.

However, comparing changes in absolute number of drivers wearing seat belts etcetera, before and after a campaign is not useful. The absolute numbers may change due to an increase or decrease in the number of drivers and thus it is important to calculate the rates. Denominators may include number of drivers, registered cars, or kilometres travelled. For example, the injury rate may be the number of injured drivers per licensed drivers, or number of injuries per 100 000 km driven. For seat belt use, the appropriate rate would be the proportion of people wearing seat belts over total drivers observed.

Measuring behaviour in observational studies is generally regarded as a more reliable method than self-reporting. When there are observable products and outcomes, it is a good technique to collect data by means of observation. Observation methods can be used to assess prevalence of a given (actual) behaviour in traffic, for instance wearing of a bicycle helmet or seat belt, or the use of day-time running lights, in a given period of time. Observations can be useful during both formative and summative phases of evaluation. Observation is in fact useful to assess larger samples of participants for instance to execute national observation of the current seat belt use, a representative set of locations should be chosen. Of course, large-scale observations are expensive (in terms of money and staff). However, when the aim is to understand the mechanisms that influence behaviour or behavioural change in relation to road safety campaigns, self-report measures is often the only option.

3.2.3 Method of document analysis – accident, injury and death statistics

Document analysis has been described in various ways, depending on the context in which it was used. However, in short one can say that document analysis is the extraction of relevant information from a set of documents in a structured way and the organization of this information. Many types of information about the social world have been collected and are available to the researcher. Documents are ‘any written or recorded material’ not prepared for the purposes of the evaluation. In general, documents can assist the evaluator in the development of evaluation questions. For example, a search may identify specific methodological difficulties and may uncover specific techniques and procedures for coping with them. Furthermore, a file review can discover pre-existing data that can be used in the evaluation.

Two major groups of public documents can be distinguished. Firstly, some information is in the form of statistical documents that contain numerical information. It is rarely collected for purposes directly related to a specific research question. Secondly, public documents in the form of published compilations (literature) available in a library or on computerised records. A literature search enables the evaluator to make the best use of previous work in the field under investigation, and hence to learn from the experiences, findings and mistakes of those who have previously carried out similar or related work.

Document analysis is the method to achieve the research/evaluation results by content analysis, statistics analysis, secondary analysis, and so on. of existing documents. The choice for one of these analysis techniques determines also the way of data collecting.

Content analysis of existing documents is a technique for gathering and analyzing the content of documents. The content refers to words, meanings, pictures, symbols, ideas, themes, or any message that can be communicated.

In case of document analysis, statistical analysis makes use of existing statistical data (not collected by means of surveying or observation within the scope of the campaign evaluation). Existing statistical research is appropriate when a researcher wants to test hypotheses involving variables that are also in official reports of social, economic, and political conditions. Patterns in the data may be modelled accounting for variations and uncertainty in the observations, and then used to draw inferences about the population being studied (the so called inferential statistics).

Secondary analysis is a special case of existing statistics; it is the reanalysis of previously collected survey data (originally gathered by others) or other information.

The most used analysis technique in road safety and in the evaluation of road safety campaigns is the analysis of existing statistics. An enormous volume and variety of statistical information exists. There are many mathematical tools used for applied statistics, for example regression analysis (linear, non-linear, logistic), multilevel and multivariate analysis, correlation, analysis of variance (ANOVA), different significance tests (chi-square, Pearson etc.). These analyses can establish a cause-effect relationship (in case certain types of designs are used) and/or a correlation between the campaign characteristics and chosen output variable (for example, the number of accidents, fatalities, injuries, saved money due to less fatalities and injuries etc.).

Because the overall aim of most road safety campaigns is to reduce the number of accidents and subsequently the number of injuries and deaths in road traffic, the ideal measurement variables in campaign evaluations could be measures of the number of accidents, injuries or deaths in road traffic.

In spite of the fact that the number of accidents with material damage gives us very important information about the situation for a certain area and an exact period, most of the countries do not register this kind of accident information in a systematic way. However, most of the countries do register the number of accidents with injuries/casualties. Police reports are often used to draw up such statistics (Elvik & Vaa, 2003). It has been suggested that one need to compare accident data over 3 to 5 years before and after the implementation of a road safety campaign in order to obtain a sufficient amount of accidents for statistical analysis (cf. Delhomme et al., 1999). Consequently, in many instances the frequency of injuries in a given month or year is too low clearly to show the effect of the campaign. There is also a risk of a change in the registration rate of road traffic accidents, i.e., the numbers in the police registration divided by real numbers of accidents (SWOV, 2007). An evaluation based on the number of in-patients³ will in this instance lead to biased conclusions. The registration rate may develop differently in different regions. Thus, as the quality in accidents statistics may be variable, it is of utter importance that the completeness and accuracy of these data sources is carefully checked before use.

To conclude, estimating the effect of a road safety campaign through accident analysis is a very delicate task because of numerous and complex influences on the

³ This is a patient who is residing in the hospital where (s)he is being treated.

accident number. Think about the meteorological effect, seasonal effect, the quality of the registration by the police services and so on.

In addition to this data, also further information as referred to as contextual data can be gathered. This data covers topics of interest such as: population characteristics, road type (in case of a regional campaign), legislation changes, enforcement (especially in case of an integrated campaign combined with enforcement activities)... The respective analysis will provide information for interpreting the road fatalities data.

3.2.4 Unintended effects

Unintended effects of road safety campaigns are difficult to anticipate, and thus, difficult to measure. For example, it is possible that a road safety campaign successfully decreases one type of injury but increases another type. This increase could occur in the same group of road users or it could involve another group using the same road section. Safer conditions, for example for car drivers might mean more danger for bicyclists. The basic principle in measuring unintended outcomes is to include measurement variables apart from the ones most directly related to the road safety campaign.

Example: Unintended effects of a road safety campaign

A particular seat belt campaign might for instance increase seat belt use among car drivers. However, an unintended effect of such a campaign could be that the speed of the car drivers also increased because the seat belt made them feel safer. Finally, these changes could have an impact on bicyclists and pedestrians using the same road. By including a measure of speeding in the evaluation, one could determine whether this actually did happen or not.

3.2.5 Methodological issues

The campaign effect is shown as the difference between the outcome that actually occurred and the outcome that would have occurred in the absence of the campaign.

Once the relevant outcomes (or campaign objectives) have been chosen and a full description of each outcome is available, the evaluator must next face the issue of how to measure them. Some campaign outcomes have to do with relatively simple and easily observed variables that are virtually one-dimensional (example: drivers wear their seat belts in the car or not). Many other campaign outcomes, however, are not as simple. In the case of intoxicated driving offences, the evaluation might use measures of number of drivers under influence alongside the road, number of accidents caused by drinking under influence, different type of accidents (e.g., mortal or not), offence severity (level of BAC), offence frequency (recidivism), number of deaths etcetera. To fully represent an outcome, it may be necessary to view it as a multi-dimensional concept and differentiate multiple aspects of it that are relevant to the effects the campaign is attempting to produce. This means that the outcomes have various facets or elements that the evaluator may need to take into account to ensure that no important dimensions are overlooked. This does not mean that all

must receive equal attention or even that all must be included in the coverage of the outcome measures selected. The point is, rather, that the evaluator should consider the full range of potentially relevant dimensions before determining the final measurement variables to be used and evaluation questions to be asked.

One implication of the multiple dimensions of campaign outcomes is that a single measurement variable may not be sufficient to represent their full character. Diversifying measures can also safeguard against the possibility that poorly performed variables will under represent outcomes. Furthermore, by not measuring the aspects of the outcome that a campaign most affects, makes the campaign look less effective than it actually is. Multiple measurements of important outcomes thus can provide for broader coverage of the concept and allow strengths of one measure to compensate for the weaknesses of another. It may also be possible to statistically combine multiple measures into a single, more robust and valid composite measure that it is better than any of the individual measures taken alone. Generally it is recommended to use several items to measure for example attitudes and behaviour (Ajzen, 2002).

The measurement of many outcome variables in evaluation uses procedures and instruments that are already established and accepted for those purposes in the respective campaign areas. For measurement variables such as attitude, cognitive abilities, behavioural intentions, etcetera, various ready-made measurement instruments or procedures may be available, but with little consensus about which are most appropriate for evaluation purposes. For other outcomes of interest, the evaluator must develop these measures. It is often not easy to construct such measures so that they assess what they are supposed to in a consistent way. Unfortunately, there is rarely sufficient time and resources to do this properly. But several standard ways of measuring such variables do exist, such as the guidelines provided by Ajzen (2002). When ad hoc measures must be developed for an evaluation without the opportunity for that development to be done in a systematic and technically proper manner, it is especially important that their basis measurement properties be checked before the weight is put on them in an evaluation. There are two measurement properties of particular concern: reliability and validity.

Reliability refers to the degree of accuracy and stability of a measuring instrument. Generally, reliability can be improved by taking multiple measurements and using the average as the data point. When measuring socio cognitive variables such as attitudes, one way to obtain reliable measures is to use indexes consisting of several items. When principal component analysis suggest that these items constitute one factor (or one construct, i.e., attitudes), then Cronbach's alpha can be used as a test of the internal reliability or consistency of the measurement variable (Cortina, 1993). Cronbach's alpha is thus used to measure to what extent the participant responds similarly from question to question (assuming that the questions are assessing similar things, like attitudes towards speeding). Nunnally (1978) suggested that Cronbach's alpha of 0.70 reflects a satisfactory level of internal consistency. In relation to accidents statistics it is important to be aware that changes in the registration rate might affect the reliability of measures.

A scale is said to be valid if it measures what it claims to measure. In relation to socio cognitive predictors a key issue is whether it is possible to discriminate

between constructs, for example, are attitudes and perceived behavioural control in fact two different measurement variables. Although there exist some evidence that the socio cognitive measurement variables presented in this chapter do comprise some degree of discriminant and convergent validity (cf. Moan, 2005), researchers should determine the discriminant and convergent validity of measurement variables. This can for example be done by using principal component analysis, which is a method used to determine whether questions or items (e.g., items used to assess attitudes and items used to assess intentions) are in fact distinct.

We recommend that evaluators invest the necessary time and resources to develop and test appropriate outcome measures. A poorly conceptualised outcome measure may not properly represent the goals and objectives of the campaign being evaluated, leading to questions about the validity of the measure. An unreliable or insufficiently sensitive outcome measure is likely to underestimate the effectiveness of a campaign and could lead to incorrect inferences about the campaign's impact (Rossi (2004).

3.2.6 Practical considerations

The feasibility of measurement variables also need to be considered. What is the cost of taking these measurements in terms of time and material resources? How much disruption of the work process is involved when choosing various measurement variables? Does the research team possess the necessary expertise to carry out the measurements properly?

The surveying can be done in a number of ways (e.g. by post, telephone, internet, face-to-face). In questionnaires the researchers can ask people about their beliefs, opinions, characteristics, expectations, self-classification, knowledge, and past or present (self-reported) behaviour. This way of data collecting is one of the most frequently used techniques in campaign evaluations. In-depth (personal) interviews give rich data and can provide a large range of views on the subject. Collecting expert opinions is a fast technique that gives useful results. It gives points of views from 'insiders' which is often very valuable but expert opinions should not be the only data source in an evaluation. Focus group interviews may generate new ideas and views in a relatively inexpensive way. Focus groups are a good means of getting evaluation data on opinions and attitudes.

On-site observations serve the purpose of clarifying road user behaviour (if observable) before and after the campaign, ideally in a case-control set up. Examples: use of seat belt, child restraint systems, use of safety helmet, speeding behaviour, use of day-time running lights etcetera. Observations are ideal in topics that do not regard personal attitudes.

Content analysis is not so relevant in relation to road safety campaign evaluation, since it is suited in the development of evaluation questions and research questions involving variables that are messages, symbols etcetera. (for example how newspapers cover a road safety campaign). However, it is suited in document changes in for example attitude as a function of a profound change (over a longer period), such as a change in rules or codes governing vehicle advertising.

Using existing statistics is very common as an enormous volume and variety of statistical information exists and usually easy to access. Nevertheless, it is difficult to determine the accuracy of data (such as used methodology to collect data, equivalence reliability in case of police controls, etc.)

In Annex 2 you find a complete overview of the advantages and disadvantages of the different data collection techniques.

3.2.7 Ethical considerations

The majority of evaluation research involves gathering information of human beings – asking them questions or observing what they have done. Since human beings have certain rights, the researchers must ensure that the rights of the participants in a study are not violated. This requires a consideration of ethical issues. For example, when using personal data (e.g., health records and other sensitive information) it is actually required to obtain written consent from participants.

Survey researchers can intrude into respondent's privacy by asking about intimate actions and personal beliefs. People have a right to privacy. Researchers should treat all respondents with dignity and reduce anxiety or discomfort. They are also responsible for protecting the confidentiality of data. Participation in a survey, interview, or focus group is always voluntary, and any form of coercion is unacceptable. As respondents are free to hang up the phone or to throw away the questionnaire, voluntary participation is a less pressing ethical issue in mail and telephone questionnaires. Nonetheless, researchers have the responsibility to inform potential subjects or respondents of all features of the study that can reasonably be expected to influence participation (length of the questionnaire, follow-up interview yes or no, guaranteed confidentiality of results, anonymity, naming sponsoring organisations...).

Also possible is that people demand answers from surveys that surveys cannot provide or may not understand a survey's limitations. Those who designs and prepare surveys may lack sufficient training to conduct a legitimate survey. Mass media reporting of survey results and the quality of surveys being reported permits abuse. Researchers should include details about the survey to reduce the misuse of survey research and increase questions about surveys that lack such information. Currently there are no quality control standards to regulate the opinion polls or surveys reported in the media.

Behavioural observations may seem to be inappropriate (at least) without consent of those being observed. Behavioural observations in the field of road safety are not ethically questionable as these observations are superficial and not in-depth observations.

3.3 Evaluation design

Research (evaluation) design is usually considered as a framework which structures (integrates) the tasks of implementing and evaluating a campaign so that it shows how all the major elements of the research process are related in answering the research question(s). These are:

- intervention(s),
- groups of subjects,
- group-assignment techniques (how subjects are sampled and allocated to various groups),
- measurement variables/data collection techniques,
- time sequence

To build an evaluation design is a complex process in which both the attributes of a campaign (e.g., scope, target group, objectives, activities) and other issues regarding evaluation framework (measurement variables, data collection method, data collection technique) need to be carefully considered and taken into account. Table 3.1 indicates the relationship between a particular research design and other evaluation components (data collection method, data collection technique, measurement variable) by summarising the combinations that appear to be the most suitable, useful and convenient both in theory and practice. Detailed explanation and argumentation of various combinations is provided in another CAST project report (Boulanger, A. et al, 2007a - chapter 6.1).

An idea about how the campaign will be evaluated and what kind of research design is required to answer the research question should thus be considered and elaborated far before the implementation of a campaign actually starts.

Several types of evaluation designs exist. Some of them (both experimental and quasi-experimental) are listed and thoroughly described in Boulanger et al (2007a). Although the list is not intended to constitute a complete state-of-the-art, it gives the most common and representative designs of each type (which are at least theoretically possible to be used for evaluating road safety campaigns). In general, various evaluation designs can be constructed simply by expanding over the basic evaluation design of the causal research which is the one-group after-only evaluation design. Such a design can be upgraded and improved by expanding it across (Trochim, 2006):

- time (adding extra measurements; such as before-after research design(s), time-series design)
- interventions (running different campaigns, either integrated or alone, resulting in factorial designs)
- measurement variables (e.g., using observed and self-reported behaviour in order to assess convergent and discriminant validity of measures)
- groups of subjects (adding another intervention group(s) and/or control group(s))

Evaluation designs can be classified in two broad categories: experimental and quasi-experimental designs. In experimental designs subjects are randomly allocated to different groups while in quasi-experimental designs they are not. In the classification of evaluation designs provided in Boulanger et al (2007a) the term 'quasi-experimental design' refers either to control-group designs without random assignment or to single-group designs with multiple measurements (e.g., the one-group before-after design, the time series design). Which group of evaluation designs to use? In theory, experimental designs are certainly preferred over quasi-experimental designs although it is often hard to implement them in practice (particularly in case of evaluating large-scale interventions such as road safety

campaigns). Usually some degree of trading between rigour and feasibility is required. Consider the following:

- Control group designs are generally considered more powerful in terms of ensuring strong internal validity than single-group evaluation designs (e.g., control-group quasi-experiments are preferred over single-group quasi-experiments), because they provide the means for comparing the effects of an intervention (e.g., road safety media campaign) observed in the intervention group with the effects observed in the control group where no intervention was given.
- Experimental designs ensure stronger internal validity than quasi-experimental designs, because they totally rely upon random selection and assignment of subjects or groups of subjects to different treatment conditions (either receiving or not receiving an intervention). Hence it is reasonable to assume that the only difference between different treatment conditions is the presence or the withdrawal of a particular intervention (all other potentially differentiating variables are ruled out)

Internal validity refers to the researcher's ability to formulate accurate conclusions about the cause-effect relationships(s). The higher the internal validity the more accurate are the causal inferences.

There are several potential threats that may undermine the internal validity of a particular research design and threaten the accuracy of the causal inferences regarding the effectiveness of a road safety campaign. Different categories of evaluation designs are vulnerable to specific validity threats. Some of the most important ones are described below, however, more information about validity threats can be found in Boulanger et al (2007a, chapter 4.1.1):

- Single-group designs
 - o History: a specific external event/condition (e.g., death of a celebrity in the traffic accident) which is not part of the campaign takes place between the two measurements.
 - o Maturation: subjects change spontaneously (with respect to a particular road safety issue) between the two measurements due to the accumulated experience or the process of maturation.
- Multiple-group designs (e.g., control group designs)
 - o Selection: subjects in the intervention group differ substantially from those in the control group from the very start that is before a campaign is launched (more common in quasi-experimental designs).
 - o Selection-interaction: specific validity threats (such as history, maturation, attrition/drop-out of subjects) influence various groups of subjects differently.
 - o Contamination: control group subjects become contaminated (due to mobility and social processes) with influences/information that were only intended for the intervention group subjects.
- Multiple-measurements design (e.g., before-after)
 - o Testing: the final score (after measurement) is influenced by taking the same test (or questionnaire) before the intervention (before measurement).

- Testing-intervention interaction: before-measurement sensitises subjects so that they respond to the forthcoming intervention differently as they would without it.

Testing and testing-intervention interaction threats only arise when subject-related data collection techniques are used (e.g., questionnaires, interviews) measuring self-reported data.

Table 3.1: Relationship between research designs and measurement variables

Data collection method	Asking					Observing	Document analysis
Data collection technique	Questionnaires, interviews (structured, half structured, etc.)					On-site observations etc.	Statistical databases
Measurement variables	Reach, recognition, recall	Likeability, comprehension	Attitudes	Behavioural intentions	Self-reported behaviour	Observed behaviour	Accident statistics (number of accidents, fatalities, injuries, etc.)
Experimental designs							
The two-group after-only randomised experiment			x	x	x	x	
Factorial design			x	x	x	x	
The two-group before-after randomised experiment			x	x	x	x	x
Multiple time series randomised experiment							x
The Solomon four-group design			x	x	x		
The switching replications randomised experiment			x	x	x	x	
Quasi-experimental designs							
Time series						x	x
Equivalent time samples						x	x
Non-equivalent control group			x	x	x	x	
Proxy pre-test			x	x	x	x	x
Separate pre-post samples			x	x	x		
Switching replications			x	x	x	x	
Counterbalanced						x	x

3.4 Minimum standards for the evaluation

The CAST consortium decided on the minimum standards for evaluation with respect to measurement variables, evaluation design and data analysis. The purpose of minimum standards is to ensure some satisfactory level of evaluation procedures throughout Europe and beyond and hence to raise both the validity of evaluation findings and the comparability of the results. Minimum standards are the least one should do when evaluating a campaign. Of course, it is always highly recommended to do more than what is defined by minimum standards. If you have skills, knowledge and extra budget we advice to elaborate and upgrade the minimum standards with the more sophisticated evaluation methodology (according to specific campaign attributes and environmental characteristics of course) that enables more valid causal inferences about the true effectiveness of a campaign.

3.4.1 Minimum standards with respect to measurement variables

1. Always measure the exposure of the campaign (reach, recall, likeability...)
2. Always measure objectives and behaviour (self-reported and/or observed) as behaviour change is the ultimate goal of all campaigns. Even when behaviour change is not specified as a direct objective of a campaign it might still appear to be one of its most important side (unintended) effects. A minimum requirement would be to conduct a before and after measurement of the behaviour in focus. The following outcome measures should be prioritized:
 - Measurement variable(s) should represent a measure corresponding with the aim of the campaign.
 - In order to understand behavioural change and to detect changes in the predictors of behaviour, variables that are common to the above mentioned theoretical models should be assessed (also when the campaign is not based on a theoretical model): attitude, subjective and descriptive norms, perceived behavioural control and behavioural intentions.
 - To get at more thorough understanding of the socio-cognitive processes that are involved in behavioural change, and in order to provide useful information for future road safety campaigns, it is useful to measure a number of extension variables and salient beliefs related to the behaviour in focus: behavioural beliefs, normative beliefs, control beliefs, personal/moral norm and past behaviour.
 - If a campaign is based on a theoretical model, measure all the main (most relevant) concepts.

Generally, it is suggested to use reliable measures. Thus, quality is preferred before quantity. See chapter 3.2.1.5 above for further details on issues concerning reliability.

Example: Minimum standards with respect to measurement variables

Case 1:

If a campaign aims to increase the knowledge about a new law, both knowledge

and behaviour (e.g., how many representatives of the target group behave in accordance with the new law) should be measured in the evaluation study.

Case 2:

If a road safety campaign is designed to influence some or all of the variables specified in theoretical models such as the TPB, the evaluation should assess the variables in that model accordingly: attitudes, subjective norms, perceived behavioural control, intentions and behaviour.

3.4.2 Minimum standards with respect to evaluation design

- The one-group after-only evaluation design should be avoided.
- Always use at least the single-group before-after evaluation design. However, be careful with conclusions as it could have been something else but a campaign that has influenced the results.
- If you have a control group design also use before and after measurements in each group.
- In case of regional/local (non-national) campaigns it is recommended to include a control/comparison group.

3.4.3 Minimum standards with respect to data analysis

- Describe the target group in terms of the relevant demographic variables (age, gender, education, etc.) and/or other background information (e.g., car driving frequency).
- Provide relevant descriptive statistics.
- Provide basic inferential statistics (significance testing and effect sizes) - explain if the difference between the two measurements is statistically significant (is it possible to generalise?) and what is the size of the effect.

Example: Minimum standards with respect to data analysis

Suppose you want to evaluate how the campaign influenced self-reported bicycle helmet use. Self-reported data was gathered before and after the campaign. When analysing data you should at least describe the respondents, provide the average helmet wearing rate for each measurement and say if the difference between the two rates is statistically significant (use appropriate statistical test, for example t-test). If the campaign significantly increased the helmet wearing rate also report on the size of the effect (e.g., large, medium or small – according to the interpretation of the Cohen's D values).

4 The choice of the proper evaluation methodology

This chapter is the translation of evaluation issues to the practical context of road safety campaigns. If you have a problem or if you don't know which measurement variables, evaluation design or data collection techniques to use to evaluate a campaign in accordance with its specific attributes (such as scope, target group, objectives and activities of the campaign) you can find the solution or some recommendations in this chapter.

The guidelines on how to work with the given information are first clearly described (section 4.1) which is then followed by the main section of suggestions and recommendations (section 4.2) about how to evaluate a particular campaign.

4.1 How to get started?

The appropriate evaluation methodology depends on different aspects (typology) of a road safety campaign. To classify road safety campaigns for the purpose of evaluation we need to consider the relevant attributes that influence the choice of the evaluation methodology. Each campaign consists of a unique combination of attributes and environmental aspects. Based on the results of the CAST project (Boulanger, A. et al, 2007a and 2007b), it was decided to classify road safety campaigns according to the relevant attributes and hence different typologies were obtained. The different typologies to consider correspond with the attributes and the environment aspects (see section 2.2) of the concerned campaign.

After all relevant attributes of a particular campaign (see section 4.2) that will be evaluated are determined; the process to find the appropriate evaluation methodology can start.

The idea behind this chapter is that you construct your own set of recommendations or guidelines to determine the appropriate evaluation methodology for that specific campaign. This implies that for each attribute you receive some tips or recommendations. At the end, the combination of all these guidelines will result in a 'best practice' evaluation methodology based on the unique characteristics of the concerned road safety campaign.

4.2 Determination of the evaluation components

Every campaign is characterised by a set of attributes. These elements are defined during the design and implementation phase of the campaign (see Boulanger, A. et al, 2007b and Delhomme, P. et al, 2009). When we speak of evaluation components this means you need to choose the appropriate measurement variables, evaluation design, data collection method, and data collection technique. The choice of these components is strongly influenced by the campaign attributes, namely the campaign scope, target group, objectives and activities of the campaign.

Thus, before you start reading about the campaign attributes you first need information about the following attributes of a particular campaign you want to evaluate:

1. the scope;
2. description of the target group;
3. clearly defined campaign objectives;
4. and details about the supportive activities in case of an integrated campaign.

In the introduction of each section, more details about what you need to know exactly can be found. Different types/values of each attribute influence the selection of the proper evaluation methodology. In every section, the corresponding guidelines and recommendations you find for each type/value of a particular attribute (typology) regarding the measurement variables, data collection methods, data collection techniques, and evaluation designs you can use if you want to evaluate your campaign properly.

For example, the scope is one of the relevant attributes of a campaign that strongly affects the evaluation methodology. Guidelines and recommendations are then specified for different scopes separately (e.g., for national and for regional campaigns).

4.2.1 Scope

4.2.1.1 Introduction

The scope can be defined as the coverage of the campaign (see Box 4.1). Campaign activities can be implemented on different levels. According to GADGET (Delhomme, et al, 1999), we can define 4 types of campaigns with respect to scope: national, provincial, local and city campaigns. Furthermore, regional or state-wide campaigns exist as well. In generally, two main levels can be distinguished: (1) national and (2) regional, provincial, local etcetera (see also Figure 4.1). The last 3 terms refer to different criteria and definitions used in subdividing national territory and can thus be treated as one homogenous category. The guidelines regarding evaluation design, measurement variables and data collection methods and techniques are the same when a campaign is regional or local.

BOX 4.1: Campaign scopes.

- *National scope*: the road safety campaign is implemented in the entire nation or country. As a consequence, the whole population is exposed to the campaign activities.
- *Regional – local scope*: the campaign is implemented in one or more states, regions, provinces, cities or municipalities. As a result, a subgroup of the entire population is exposed to the campaign materials and/or activities.

In case, a national campaign is combined with for example local actions, other guidelines or recommendations will be applicable to evaluate these local activities than for the evaluation of the nation-wide activities (see section 4.2.4 Activities, Integrated campaigns).

4.2.1.2 Measurement variables and data collection methods and techniques

In theory, an evaluation of a national road safety campaign results in the need to examine every member of the population. But restrictions on available time or money usually prohibit observation of an entire population. The usual procedure in these instances is to take a sample from the population. As a consequence, sample data are selected with the intention of generalising results to the population from which the sample units were drawn. It is important, therefore, that the sample is representative of the population (Peck, 2005), whether it concerns the population of a country or the population of a target region. This goes for all data collection methods. Sample size may have a dramatic effect on an analysis (Wholey, J., 2004). When a sample size is very small, certain methods lose their statistical power. Therefore, researchers should be cautious when generalising results of qualitative studies. This disadvantage can be compensated for, however, for example by in-depth interviews, which elicit richer and more detailed responses.

When a regional or local campaign is implemented, the evaluator should take into account the geographical limits of the study area when sampling. Normally, it is not a problem to conduct questionnaires, interviews, focus groups or observations. When analysing statistics, a researcher should, theoretically, only use data that cover the study area. Besides, in many instances the frequency of injuries in a given month or year is too low to show an effect of the campaign. Consequently, it will be a problem to obtain a sufficient amount of (types of) accidents in that region or municipality. Caution is necessary in case the size of a country such as Luxembourg (0.5 million habitants) is comparable to a region. Moreover, a regional campaign in for example the German state Bavaria (12 million inhabitants) is comparable to a national campaign.

4.2.1.3 Evaluation design

- *National campaigns*

Because (almost) everyone is exposed to the national campaign it is difficult to find and formulate an uncontaminated control group. To evaluate the effectiveness of a national campaign we suggest using one of the single-group multiple-measurements evaluation designs listed below with *the one-group before-after evaluation design being the minimum standard as recommended by the CAST consortium*. While considering different options both the possible budget restrictions and some methodological issues (see Box 4.2) should be taken into account.

- *The one-group before-after evaluation design* gives an indication of the progress/improvement that *might be* the result of the implemented campaign. However, be aware that simple causal inferences might be confounded by not considering the potential factors that could have occurred during campaign's running period (e.g., history & maturation threats, see also Box 4.3). Be careful with the interpretation of the results!
- *The one-group before-during-after evaluation design* enables the researcher to inspect the pattern of changes from the baseline to the follow-up period. However, the above remark about being careful with causal inferences still remains.

- *The time series evaluation design* includes several waves of measurements both in the baseline and the follow-up period and thus enables to (a) assess the level of confounding due to history and maturation threats and to adjust the overall effectiveness estimation accordingly and to (b) assess the sustainability of campaign's effects over time (investigating both the short-term immediate effects as well as the long term persisting effects).

BOX 4.2: Methodological issues to consider when using one-group multiple-measurements evaluation designs:

- The observational evaluation method and the method of document analysis are preferred over the method of asking (especially when observational and statistical data are representative for the (national) population).
- When measuring self-reported measurement variables (e.g. knowledge, awareness, and attitude) by means of telephone polls or internet questionnaires (asking method) it is very difficult to survey the same respondents twice (or even more often). However, if different groups of subjects are measured on each occasion (independent groups designs), this is not considered as affecting the results.

BOX 4.3: History and maturation threats.

- *History threat* refers to the effect of an incident that occurs during the intervention (however not interrelated with the intervention) and may thus influence its effectiveness. The longer the before-after period, the greater the possibility is of such an effect to occur. For example, a car accident with many fatalities may confound (overestimate) the assessed effectiveness of a speed campaign, running at that period.
- *Maturation threat* refers to the changes in the subjects' behaviour (attitude, knowledge etc.) during the time period of the campaign. Such changes may be attributed to the re-evaluation of principles and beliefs that affect the subjects' perception which consequently also affect their behaviour towards the campaign (and its messages/ideas).

- *Regional/local campaigns*

Unlike national campaigns aiming to reach wide audiences, the scope of the regional and/or local campaigns is much more limited. Such campaigns are usually tailored to specific geographical regions (e.g., communities) so they might potentially be evaluated by means of control group evaluation designs. Because a campaign only reaches one segment of the whole (national) population (e.g., due to budget restrictions, initiative of the local authorities, road safety issues that are distinct from the national average) it is possible to expand the single-group multiple-measurements evaluation designs across groups by adding a control group. Hence, another similar region, community or segment of the population not being exposed to a campaign may formulate the control group and serve for comparison.

If you only pick out one region/community to serve as a control group then this is a quasi experimental design. But if you randomly allocate all units (e.g., all communities in a given geographical area) to either experimental or control condition then you have an experimental design. As experimental designs are rarely applicable in practice (e.g., due to budget restrictions) we suggest evaluating regional/local campaigns by means of quasi experimental control group designs (see Box 4.4).

BOX 4.4: Quasi experimental designs.

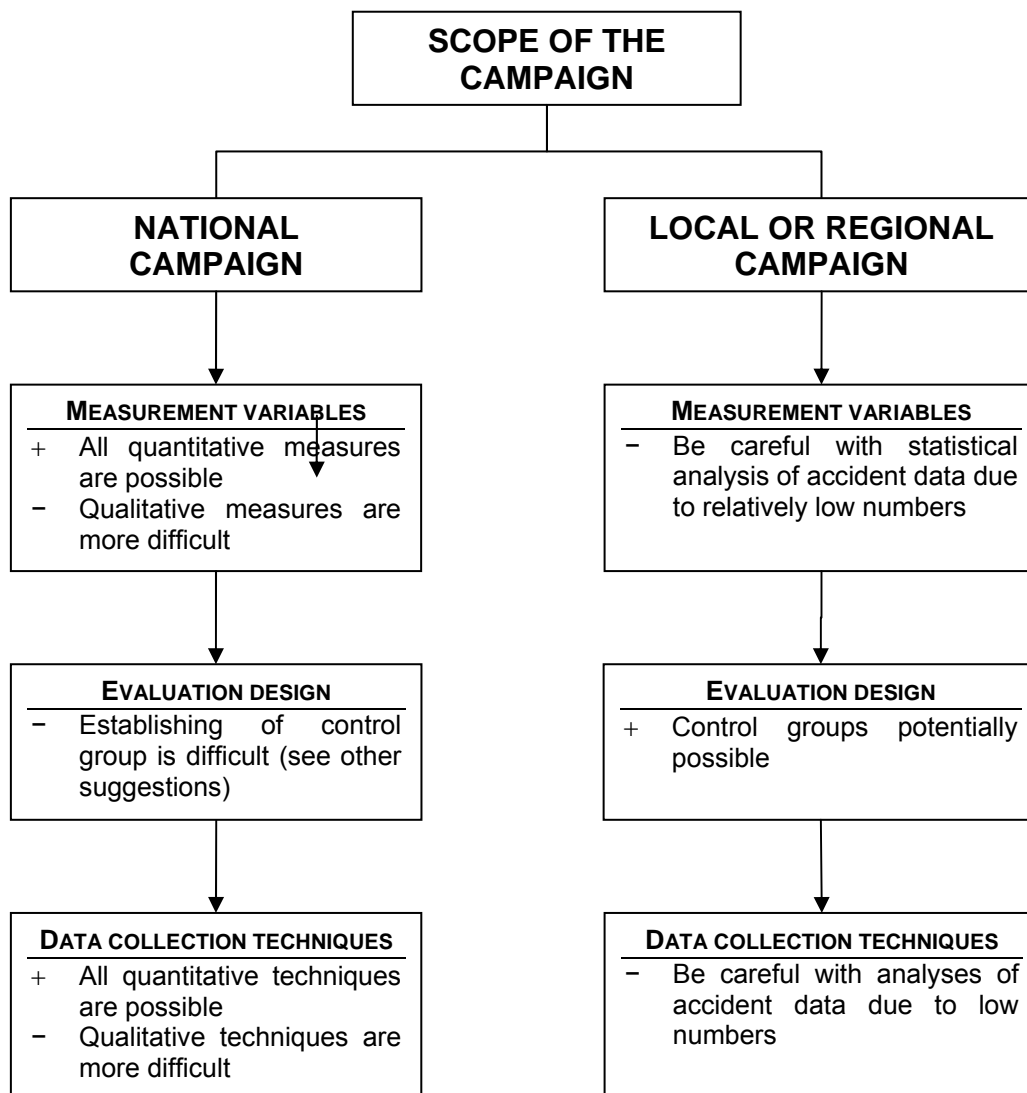
- *Non-equivalent control group quasi-experimental design* (the two-group before-after quasi-experimental design) requires the establishment of a control group, apart from the intervention group. There is bigger selection in measurement variables, since surveys may be conducted, for the assessment of attitude, intentions, reach, recognition/recall, and self reported behaviour, as well as observational studies, and document analysis. Measurements have to be available before and after the intervention, however, in case of document analysis, no time restriction for data collection applies (as in the case of proxy pre-test quasi-experimental design).
- *Proxy pre-test quasi-experimental design* resembles the non-equivalent control group quasi-experimental design, however, enables both before and after measurements (regardless of the measurement method) to be taken after the completion of the intervention. In the case that survey method is selected, special attention to the formulation of the questions should be paid, and a hypothetical phrasing should be used. The low cost of this design is counterbalanced by the reduced reliability of the measurements, due to the proxy variable used.
- *Separate pre-post samples quasi-experimental design* is more complicated, since it required the establishment of 4 groups of subjects, and four measurements, with different participants in the baseline and follow-up period. Although, it increases the randomization factor, as compared to other quasi designs, the groups cannot be considered equivalent, even if they are selected from the same wider population (e.g., same community). Surveying is the most suitable method for measuring the effect of the intervention, and results from a previous survey, may be used as the before measurement of the current research.

Quasi-experimental control group designs (with the first one being preferred) minimise the chance of wrong causal inferences due to history and maturation threats (that is the potential changes within subjects that may spontaneous arise during the subsequent measurements, see also Box 4.5). Quasi-experimental control group designs enable the evaluator to track changes in each of the groups. If there is no difference between the two measurements in the control group while a significant difference is observed in the intervention group, then we can conclude that a campaign probably has been efficient.

BOX 4.5: When evaluating a campaign by means of quasi-experimental designs the following issues need to be considered when interpreting the results:

- Are the respondents in each condition (intervention and control group) similar enough (as for example in case of very distal communities)? Otherwise it may be the prior differences between the groups (not the exposure to a campaign) that might have contributed to different effects in each condition.
- People from the control-region (community) may travel to the intervention-region, resulting in a certain degree of contamination (people from the control-region see/hear/read about a campaign when they visit intervention region). When the two regions are very close to each other it is reasonable to believe that there will be more contamination. In practice, evaluators can ask the respondents if they have seen the road safety campaign.

Figure 4.1: Summary - scope and evaluation components



Note: (+): positive consequence for the evaluation component, (-): negative consequence for the evaluation component

4.2.2 Target Group

4.2.2.1 Introduction

The target group of a campaign is obviously closely linked to the theme/road safety issue being addressed. The target audience is constituted of people who behave in an inappropriate (dangerous) way and whose road safety behaviour has to be changed. To run a successful campaign it is advisable that the target group is clearly defined in order to prepare a campaign according to their specific needs and characteristics. Target groups can be described by defining the road user type, the age category and/or the profile/profession. Another dimension of describing the target groups is if they were targeted directly or indirectly by the campaign.

The *direct target group* is the target audience that is directly concerned by the problem (e.g., road users prone to exhibit dangerous/unsafe behaviour). But, it is often more effective to target other people who can influence the concerned high-risk road users. Examples of those 'significant others' are: parents, partners or peers. They are called the *indirect target group*. The indirect target group is a group of people that may influence the attitudes or behaviour of the direct target group. It can be particularly useful to target a group indirectly when the direct target group has rigid attitudes and is unlikely/unwilling to change their risky behaviour (Elliott, 1989). Some campaigns target both types of target groups.

Example: Direct and indirect target groups

In Great Britain, the campaign "For my girlfriend" (launched by Think!), highlighted the danger of driving too fast. The campaign is based on the fact that more young female drivers die as passengers than as drivers. . The objectives of the campaign were: (1) to encourage young male drivers to slow down, and to make young female passengers aware of the particular danger of the situation, and (2) to influence female passengers (as indirect targets) so that they would try to prevent their male drivers (direct targets) from risky driving

Source: retrieved 18/07/2007 from <http://www.fmg.org.uk/>

For the evaluation of a campaign, all target groups should eventually be taken into consideration when collecting information. But, the fact that a group is either directly or indirectly targeted is less determining than a very precise description of the characteristics of that particular target group(s) (see also Figure 4.2).

The target can further be described in 2 different ways: (1) non-specific and (2) specific.

- A non-specific target group means that all drivers (all holders of a driving license) or all vulnerable road users are targeted.

Example: Non-specific target group

A central city pedestrian safety campaign was launched in 2003 in New Zealand. It was an integrated campaign that combined visual images and messages with activities and rewards. The objectives of the campaign were: encouraging pedestrians to use the crossings, crossing in case of green light and not when the

crossing light is red, and encouraging drivers to give way to pedestrians when turning left. In this campaign we can see that the target audiences are non-specific, all pedestrians and drivers are targeted.

Source: Harré & Wrapson, 2004

- A specific target group is a well-defined target audience. The description can be very precise by defining target group with a combination of road user type, age category or a certain profile (profession, gender, etc.).

Example: Specific target group

A drink drive festive campaign (Don't risk it) in Scotland in 2003 targeted in particular young male drivers aged 17-29. This campaign targets thus a well defined group. The evaluation research aimed to test awareness and understanding by drivers of the campaign and assess the impact. There were two phases: a total of 1000 in-home interviews were conducted with car drivers (quantitative research), and a series of 14 focus groups and 10 in-depth interviews were conducted (qualitative research).

Posters and radio advertising were specifically targeted at young male drivers. By using the "Crimestoppers" hotline any member of the public could report a drink driver to the Police and if a successful conviction followed, the informant could be paid up to 500£ (nearly 800 €). The campaign was evaluated on the target audience, which was compared to the whole population. One of the results is that the radio commercials were remembered by 20% of all respondents with awareness higher amongst the 17-2+ year olds (31%).

Source: *Scottish executive evaluation of 2003-2004 festive drink drive campaign.* (2004). Retrieved 14 January 2009 from www.scotland.gov.uk/socialresearch

4.2.2.2 Measurement variables and data collection methods and techniques

When a target group is specified, a sample should be selected from this specific group as the sample information will be used to gain insight into characteristics of this target group. The more specific the target group, the more difficult it will be to accomplish a representative sample size. This is valid when collecting the necessary data by the method of asking. More difficulties arise in case observations or counts are made at a number of sites. This data collection technique involves on-site visits to locations to observe activities and the researcher takes notes on the behaviour and activities of individuals. A special issue regarding specific target groups is that the individuals to include in the study are defined beforehand. In reality, it will be hard to observe the right persons. Consequently, it will be even more difficult to attain the needed amount of observations. Only if detailed socio-demographic information is available for the related crash and casualty statistics will it be possible to select the needed data for the targeted group.

In case of a non-specific target group, all the above mentioned difficulties related to sampling are not applicable anymore. All three of data collection methods can be used.

4.2.2.3 Evaluation design

- *Specific target group(s)*

Having specifically designed target groups not only means that the audiences are well-defined and described in terms of their age, gender and profile, but also that a campaign is specifically designed to address particular segments of the population (in terms of its characteristics, like media channels that are used). Namely, if a campaign aims to influence well-defined, narrow target groups then members of another (non-targeted) group of road users can formulate the control group.

Thus, for evaluating campaigns targeting specific groups of road users we suggest using one of the control group quasi-experimental evaluation designs (listed above) with targeted subjects being part of an intervention group and particular non-targeted subjects being part of the control group. Of course, it is highly recommended that the two groups of subjects are as similar as possible.

Experimental designs are not an option here because subjects can not be randomly allocated to each of the groups. Even if subjects are randomly drawn from the two populations (targeted and non-targeted) this is still quasi-experiment because random allocation to each of the groups is missing.

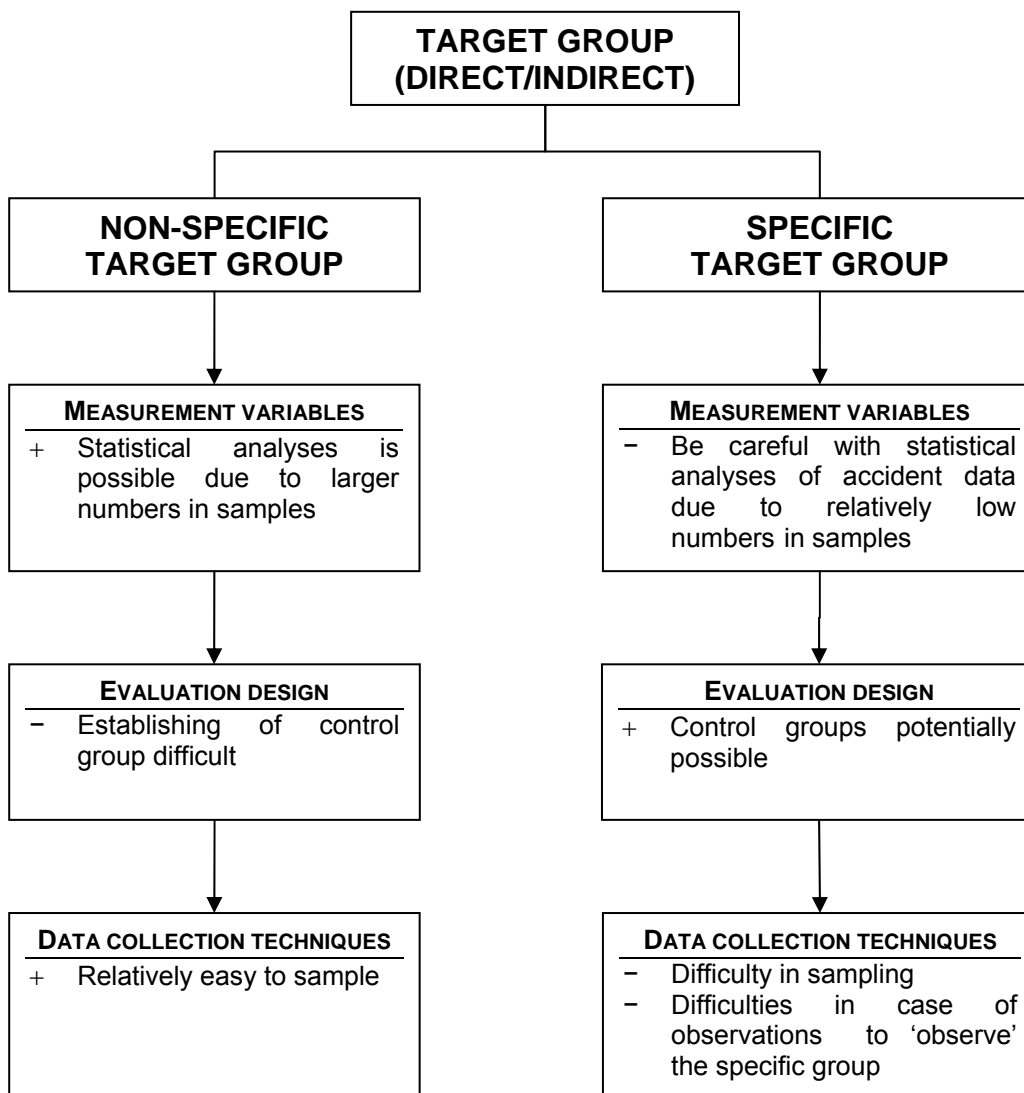
Example: Intervention and control group

- *Intervention group*: children who still need to sit in the child safety seats while travelling in the car (clearly specified by age – although the age interval for obligatory use of child safety seats differs across countries).
- *Control group*: older children who need to wear normal seat belt while travelling in the car.

- *Non-specific target group(s)*

If a campaign targets non-specified, wide audiences (e.g. all road users, all car drivers) and it is consequently impossible to establish a control group from non-targeted population we suggest using one of the single-group multiple measurements evaluation designs outlined above (Section 3.3.1.4). The single-group before-after evaluation design is recommended by CAST consortium as a minimal standard.

Figure 4.2: Summary – Target group and evaluation components



Note: (+): positive consequence for the evaluation component, (-): negative consequence for the evaluation component

4.2.3 Objectives

4.2.3.1 Introduction

While the goal of a campaign is more general, its objectives address specific aspects of the problem. At least in theory objectives should be clearly defined in terms of (a) desired communication or behavioural effect, (b) the public who is suppose to reach the effect, (c) the expected level of attainment or accomplishment (e.g. 20% increase in awareness) and (d) the time period (e.g. campaign period) when the effect is expected to occur. Objectives should be specific and measurable, as they will help in monitoring the progress towards achieving the pre-set goals. In many sources, objectives are considered well-defined if they are SMART (MacDonald, Starr et al. 2001): Specific, Measurable, Achievable and Ambitious, Relevant and Time-bound (see also Annex 1). Having well-defined and measurable objectives is also very important for assessing campaign’s effectiveness (outcome evaluation) (see also

Figure 4.3). A campaign is generally considered effective (successful) if it reaches the pre-set objectives.

A campaign may influence the target audiences at various different levels. For example, it may be designed to increase/change knowledge or awareness of a specific road safety issue, shape positive attitudes, promote safe behaviour or lower the number of road accidents (fatalities, injuries) (or a combination of all these goals). Objectives of a campaign thus commonly aim at influencing knowledge, awareness, attitude, behaviour or accidents.

Example: Possible campaign objectives

- *Knowledge*: to increase the knowledge about a new law concerning the limit for blood alcohol concentration among car drivers in a specific country (the knowledge about the exact legal limit is of interest).
- *Awareness*: a campaign against speed on roads wants to increase awareness of the population about possible accidents that speed may cause (e.g. Irish campaign called “The Faster the Speed, the Bigger the Mess”). A drug driving campaign was specifically aimed to raise awareness of the dangers of drug driving among clubbers (primarily males aged 17-34 years) during the Christmas and New Year period.
- *Social cognitive variables such as attitude*: to increase the positive evaluation among the new legal limit for blood alcohol concentration among car drivers in a given country from 50% to 75% (evaluations might both concern cognitive/rational and affective/emotional aspects).
- *Behaviour*: to increase the percentage of young adults (18 – 25 years old) who report to wear a seat belt all the time from 80% to 95%.
- *Accidents*: to decrease of the number of alcohol related accidents in a specific country with 30%.

4.2.3.2 Measurement variables and data collection methods and techniques

In case a road safety campaign aims to increase the knowledge and awareness of for example existing regulations or the danger connected to particular traffic behaviour, the only possible data collection method is the method of asking. Particularly short interviews (by telephone or online) can be easily conducted to provide insights in people’s knowledge and awareness. Generally, the costs are lower and the response rates are higher than questionnaires.

If a campaign is designed to increase the knowledge concerning issues related to drunk driving in the Netherlands, a good way to determine the effect of the campaign is to measure knowledge before and after the campaign is implemented. If possible it is also recommended to assess knowledge during the campaign period. Finally, it is preferable to measure knowledge directly after the end of the campaign period. Knowledge can be measured using one question: “Do you know what the legal limit for alcohol blood concentration among car drivers is in the Netherlands?” In this

instance one can provide respondents with various response alternatives (0.0, 0.1, 0.2, 0.3 and so on up to 1.0) and make them select one of the alternatives.

In instances when the campaign is designed to increase the awareness of an issue, for example a new law that is introduced in Norway and the legal alcohol blood concentration has been reduced from 0.5 to 0.2, one might ask respondents if they are aware of this new law. In this case one would not be interested in the facts (i.e., the exact limit of the alcohol blood concentration), but rather the awareness that there has been a change in the law. As with knowledge it is necessary to measure awareness before and after the campaign, and if possible it is recommended to also assess awareness during the campaign period. The after-measure should be taken right after the end of the campaign period.

When a road safety campaign concerns the persons' own attitudes related to the desired or safety behaviour, it is better to use self-report through questionnaires. In relation to issues like alcohol and drugs, respondents might also be more honest and feel better if they can sit by themselves when filling out a questionnaire.

Attitudes (and other socio-cognitive variables) towards a new law concerning the legal alcohol blood concentration for car drivers might be assessed before and right after the end of the campaign period. Taking measures during the campaign period as well will result in more robust measures. It is important to use at least three or four items when measuring socio-cognitive variables like attitudes in order to ensure reliable measures. Reliability might be assessed using Cronbach's alpha, after principal component analysis have shown that the items do in fact constitute one factor or one variable. It is recommended to use the guideline provided by Ajzen (2002) when measuring attitudes and other socio-cognitive variables⁴. In the above example, attitudes might be assessed by the following four items, using the stem "In my opinion, lowering the legal limit of alcohol blood concentration from 0.5 to 0.2 for car drivers in Norway, would be": bad (-3) – good (3), not useful (-3) – useful (+3), harmful (-3) – beneficial (+3), foolish (-3) – wise (+3).

In general, the method of asking is useful to collect (baseline) data on knowledge, awareness, and attitudes. It is namely impossible to observe along the road side people's knowledge, awareness, and attitude. For example: a person can be aware that speeding is dangerous but still he infringes the speed limit. The method of document analysis is possible in case former knowledge or attitudinal measurements were investigated. These existing studies can be used as baseline data or as an example on how to formulate the evaluation questions (see section 3.3.1).

Most campaigns are focussing to change behaviour of the road user. To measure the effectiveness of these road safety campaigns different data collection methods can be applied.

In case the campaign combines different objectives (knowledge, awareness, or attitude) it is advisable to uptake behavioural questions in the questionnaire. Critics find the method of asking suspect because of their reliance on self-reported behaviour. This method is very sensitive to deception and slip of the memory. The

⁴ The same is recommended in instances when road safety campaigns are designed to change subjective norms, descriptive norms, perceived behavioural control and intentions.

data inaccuracy can be really large. Surveys also cannot tap into the contextual element. Proponents of surveys counter that properly constructed with built in checks and balances can overcome these problems and provide highly credible data. For example, by using several ways of assessing behaviour through self-report (exact numerical report, rough numerical estimate, and rating scale) one will obtain more reliable measures (Ajzen, 2002). It is also suggested to compare self-reported behaviour with observed behaviour. To determine changes in self-reported behaviour it is necessary to conduct measures before and after the campaign, and if possible it is also recommended to take measures during the campaign period. Moreover it is desirable to measure behaviour directly after the end of the campaign period to determine short-term effect and a few months after the end of the campaign period to determine whether the changes are stable or not. In order to draw more certain conclusions regarding the effect of road safety campaigns on behaviour, a control group (which has not been exposed to the campaign) is necessary.

A potential threat to the validity of self-report measures is social desirability, which is the tendency to give *right* answers. Although some studies suggest that this does not resemble a large problem, this will of course depend on the theme of the campaign ('alcohol- and drug use in road traffic' versus 'safety distance'), and the cultural context where the study is conducted. Thus, particularly when road safety campaigns address sensitive themes such as alcohol- and drug use among road users, it is recommended to apply a social desirability scale and subsequently control for its effect (see Crowne & Marlowe, 1960; Rudmin, 1999, for further details).

When there are observable outcomes (certain traffic behaviour), on-site observations is a good technique to collect data. Observation methods can be used to assess prevalence of a given (actual) behaviour in traffic. These methods are widely used in connection to road safety campaigns, especially campaigns concerning use of safety devices like bicycle, moped or motorbike helmet and seat restraint systems, keeping safe distances and reducing the number of speeding violations. It is an excellent technique to discover, observe, and evaluate actual behaviour of road users in a natural setting but it can be expensive and time consuming to plan and carry out field trips to collect representative data.

If the object of a road safety campaign is to decrease drunk driving in a country or a region of a country, the level of blood alcohol concentration among car drivers may be measured by stopping drivers at a road site. However, researchers can not do so without the help from the police. When the police agree to cooperate it is necessary to conduct such measures before and after the campaign period. A control group will enable more firm conclusions regarding the actual effects of the campaign. Finally, to determine short- and long-term effects of a campaign it is recommended to assess the alcohol blood concentration directly after the end of the campaign period as well as some months after the campaign period.

Because the overall aim of most road safety campaigns is to reduce the number of accidents and subsequently the number of injuries and deaths in road traffic, the ideal data collection method in campaign evaluations is document analysis, more specific the analysis of statistics. Document analysis is a relatively economical and

efficient way of collecting relevant data. Especially data from existing statistics are an unobtrusive source of information, and the analysis is relatively straightforward. In spite of the fact that the number of accidents with material damage gives us very important information about the situation for a certain area and an exact period, most of the countries do not register this kind of accident information in a systematic way. There is also a risk of a change in the registration rate of road traffic accidents, i.e., the numbers in the police registration divided by real numbers of accidents (SWOV, 2007). Thus, as the quality in accidents statistics may be variable, completeness and accuracy of these data sources should be carefully checked before use.

The method of asking also gives the opportunity to measure information about accidents. These self-reported accidents are of course less reliable than statistics because of the social desirability.

In case a campaign aims to reduce accidents caused by drink-driving, data regarding this type of accidents should be available and analysed. Again, it is necessary to use data from before and after the campaign period, and if possible data assessed during the campaign period. However, even if a campaign is intended to reduce the number of accidents it will not always be possible to have a reliable accident evaluation. The number of accidents may be too low for a statistical analysis with enough discriminative power, or in the case of alcohol accidents the reliability of the alcohol registration may not be good enough to allow a good accident analysis (Delhomme et al, 1999). By comparing accident data over 3-5 years before and after an intervention it may be possible to select enough accidents to allow a statistical analysis. Unfortunately there is a risk of a change in the registration rate (the numbers in the police registration divided by real numbers of accidents). As a result, reliability and validity of the data should always be conscientiously assessed.

The manner in which the effectiveness of a campaign is measured will influence the magnitude of the effect (Delaney, Lough, Whelan & Cameron, 2004). Measures of effect such as awareness of a campaign or campaign message are likely to change the most as a result of a road safety mass media campaign. In contrast, changes in objective measures of behaviour, such as seat-belt use or the frequency or severity of crashes, are likely to be smaller in size. Although it is important to determine the reach of the campaign, the measures of for example awareness do not provide sufficient information to fully understand effects on road safety as they do not represent actual changes in behaviour nor consequent reductions in crashes.

The base level of the measure of effect will also influence the degree to which a campaign is successful. In general, the higher the base level of the relevant measures of effect, the smaller the expected impact of the mass media campaign. This is consistent with expectations as the target group on which the campaign can act is smaller in size.

4.2.3.3 Evaluation design

To observe if a campaign reached the pre-set objectives it is necessary to use one of the multiple-measurements evaluation designs. Beside the single-group before-after evaluation design which is considered a minimal standard for evaluation one can also use the single-group before-during-after and the time series design (see section Scope/National campaigns). These designs only differ in the number of

measurements which highly depend on the measurement variables and the sampling procedure.

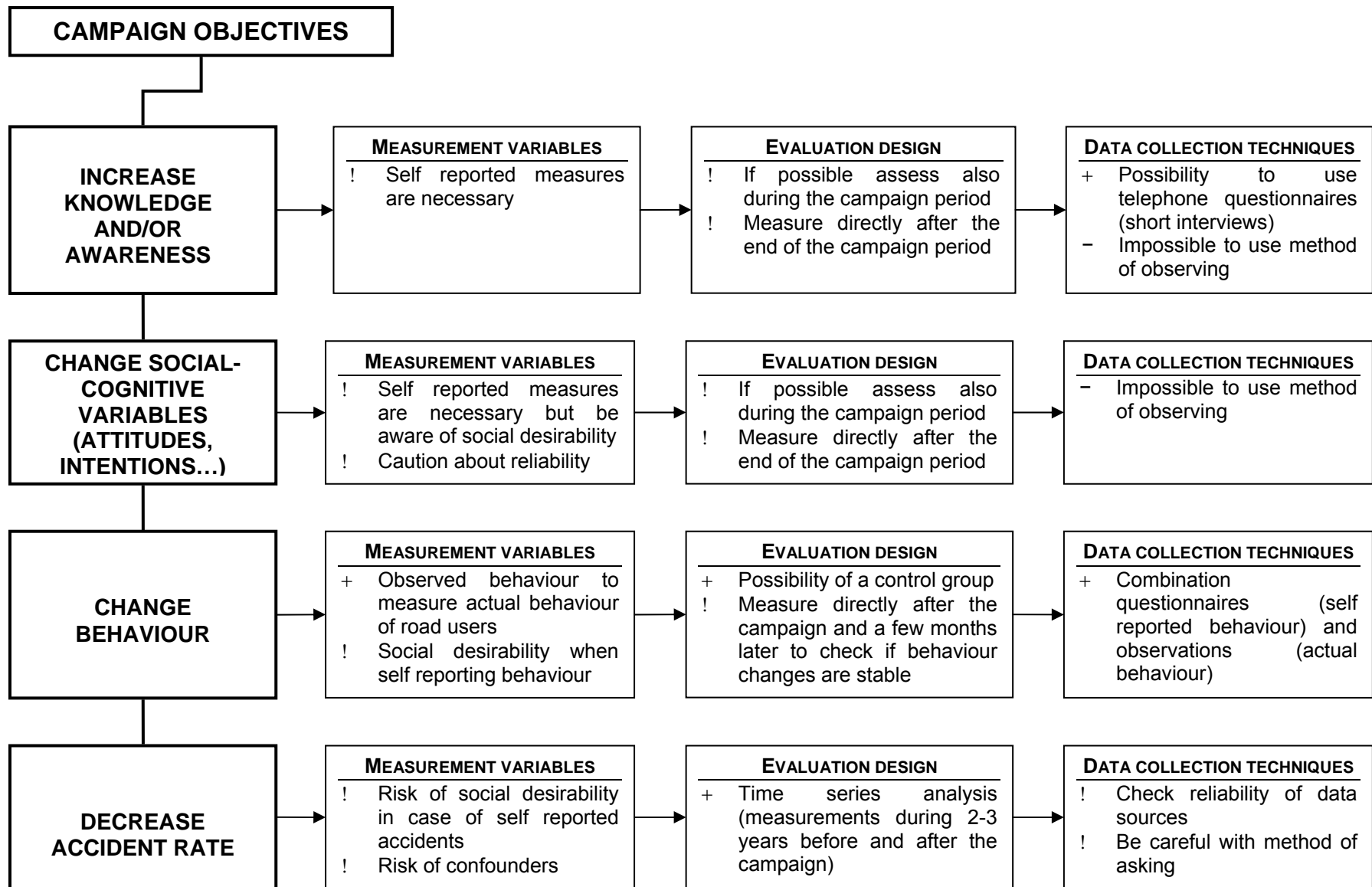
When a campaign is evaluated in terms of observed behaviour and/or accident statistics it is recommended to use several waves of measurements (in order to strengthen the internal validity of the design and to assess both the short-term and the long-term effects) as these two measurement variables do not interfere with the subjects (data is gathered independently of the subjects).

But when self-reported measurement variables are used (knowledge, awareness, attitude, self-reported behaviour etc.) then the number of measurements depend on the sampling procedure (see Box 4.6).

BOX 4.6: Practical considerations about self-reported measurement variables:

- Because self-reported variables can only be assessed by asking participant their opinions/estimates it is not recommended to use evaluation designs with more than two waves of measurement. Participants would probably refuse to fill out the same questionnaires over and over again, which might result in high drop-out rates.
- Independent-groups evaluation designs may solve this problem. Instead of measuring the same subjects on different occasions (e.g., both before and after) independent samples may be drawn from the target population with each sample being measured on one particular occasion (e.g., before the campaign). However, the samples must be representative of the target population in order to ensure sufficient level of comparability (otherwise the observed changes may be attributed to the initial differences between the samples rather than to the campaign itself). For more details, you can refer to the separate pre-post samples design, Boulanger et al (2007a) chapter 4.
- Self-reported variables are recommended to be assessed directly at the end of the campaign's running period.

Figure 4.3: Summary – Campaign objectives and evaluation components



Note: (+): positive consequence for the evaluation component, (-): negative consequence for the evaluation component, (!) keep in mind.

4.2.4 Activities

4.2.4.1 Introduction

Media campaigns rarely stand alone. Other activities are often integrated with a media campaign in order to strengthen and support campaign's message/ideas. A pile of research shows that integrated campaigns are far more effective than media campaigns alone (only publicity and communication) especially in terms of changing behaviour (see Delhomme et al, 1999). Because of this greater effect on road user behaviour, there is a need for evaluation research to focus on the (separated) effect of the other supporting activities (see also Figure 4.4).

There are several actions/activities that could be integrated with media campaigns, such as:

- Police enforcement
- Educational program
- Actions in driving schools
- Rewards or incentives (with or without commitment)
- Technical improvements
- New legislation
- Some specific infrastructural measures
-

The so called media campaigns are carried out alone, meaning only by implementing media activities. The goal is to address the defined target group and to reach as many people as possible to ensure that all persons of the target group are covered. Different media channels can be used such as:

- Internet
- Television
- Radio
- Press
- Billboards
- Posters, brochures, flyers, stickers...
- Gadgets
-

Example: Media campaigns

Campaign 1:

A media campaign in London (Great Britain) wanted to convince drivers to slow down. The used media channel was posters and no other actions supported this campaign.

Source: Simmonds, 1981

Campaign 2:

The National Highway Traffic Safety Administration (NHTSA) launched in 2002 a seat belt used campaign called 'Buckle up now! (=click it or ticket)'. This national seat belt campaign was a good example of a campaign that used several media

channels. The enforcement activity was associated with two different kinds of media: the 'earned' media and 'paid media'.

- *Earned media* is where program details and results are developed into newsworthy events that are circulated into the public by broadcasters and newspapers. Earned media generally began one week before paid media, two weeks before enforcement and continued until the end of the campaign.
- *Paid media* usually lasts two weeks. Radio and TV advertisements aired extensively. Generally paid media began one week before enforcement and continued airing during the first week of enforcement.

Three groups of areas implemented the campaign differently. The *full implementation states* combined paid media, earned media and intensive enforcement activities. The *other implementation states* used only some paid media. The *comparison states* avoided the use of direct paid advertisement placement. Comparisons were made between these 3 types of groups. The results confirmed that intensive, short term and well-publicised enforcement produced large gains in seat belt use.

Source: Solomon, Ulmer & Preusser, 2002

4.2.4.2 *Measurement variables and data collection methods and techniques*

On first sight the supported activities of a campaign have no influence on which method should be used to collect data. Of course, when formulating evaluation questions, the evaluator should take into consideration the several activities integrated in the campaign. Furthermore, it is recommended to gather information about the supported activities themselves. This information is often missing in the current evaluation reports in the EU and beyond. For example, surveillance intensity and the results of the police controls in case of enforcement. Information on surveillance intensity is important because if police enforcement is very intense (e.g., frequent and spread all over the country) and publicised in advance greater effects can be achieved (in terms of safe behaviour of road users) than in case of weak and irregular enforcement activities. In this way, the chance of being caught will increase which may influence the road user behaviour.

To uptake detailed information about the different activities, it is necessary to think about this research on beforehand.

If the objective of a campaign is to reduce the number of incidents of drunk driving, and both a mass media campaign is designed and implemented along with increased control activities (taking measures of the blood alcohol concentration along road sites), then the evaluation should take into account the effect of both the mass media campaign and the control activities. In such instances it is necessary with before- and after-measures which assess all important aspects of both activities, such as reach, awareness, and so forth.

4.2.4.3 *Evaluation design*

- *Media campaign alone*

When media campaigns stand alone, both the scope and the target group should be investigated in detail to determine the proper evaluation. It is also very important to know which media channels are used. In case only high-reach communication channels are used such as TV, radio, billboards and websites it will be impossible to find a control group. But then again, when a media campaign is based on brochures, pamphlets, leaflets, postcards, napkins, magnets, and so on, it is feasible to deliver certain materials only to a (randomly) chosen intervention group, which makes it possible to use one of the control group designs. However, such campaigns are rarely used.

- *Integrated campaigns*

When a campaign is integrated with supportive actions (e.g. educational programs in schools/kindergartens/companies, special actions for driving schools, enforcement etc.) it is possible to use either quasi-experimental (especially the non-equivalent control groups design, the proxy pre-test design or the separate pre-post samples design, see section 3.3) or experimental evaluation designs (often only on the level of the supporting activity, see Box 4.7). Participating institutions are then selected and assigned to distinct conditions either by means of (group) randomization (experiments) or otherwise (quasi-experiments). Members of institutions that were assigned to the intervention condition thus receive a campaign integrated with additional actions (e.g. education), while members of the 'control condition' receive media campaign alone. With such a design it is possible to compare the effectiveness of a media campaign which stands alone with the effectiveness of an integrated media campaign.

Here are some examples of how experimental designs can be applied on the level of additional actions (in case when a media campaign is implemented nationally and practically everybody is exposed to it):

- Campaign integrated with education in schools/driving schools/companies
 - o identify all schools in the country (or in the particular region);
 - o randomly select a particular sample of schools out of the whole population (of schools) and allocate them randomly to the two equal groups;
 - o deliver the program to the intervention group (make sure that the program implementation coincides with the campaign's running period in order to exert accumulated effects);
 - o observe what goes on in the control group (potential changes may be attributed only to the campaign itself);
 - o compare the effectiveness of a media campaign alone with the effectiveness of an integrated campaign (media & educational program).
- Campaign integrated with enforcement
 - o identify all communities/municipalities (or other geographically defined locations) in a given country (or in the particular region);
 - o identify road sections of the same type with similar traffic patterns in each community;
 - o randomly select a particular number of communities and allocate them randomly to the two equal groups;
 - o implement enforcement activities in the intervention communities (if possible during the campaign's running period) while withholding enforcement of the law(s) in the control communities;

- compare the effectiveness of a media campaign alone with the effectiveness of an integrated campaign (media & enforcement).

BOX 4.7: Practical considerations about experimental designs.

- when group randomisation is used (the procedure of sampling/allocating groups of subjects such as schools/communities instead of individuals is called group randomisation) it is suggested to have at least 5 different units (e.g. schools, communities) in both the intervention and the control group (to achieve sufficient statistical power – as in case of individual randomisation)
- randomisation procedure raises some ethical/moral issues about the fairness of denying services to particular individuals (the control group), however
 - if because of budget restrictions it is not possible to deliver the intervention to everyone who needs/deserves it then it should not be a problem to use randomisation techniques
 - some authors point out that ethical issues only arise when an intervention has already been proven effective but it is still not delivered to everyone (when the effectiveness has yet not been proven it is OK to do scientific experiments, otherwise not)

Which one of the experimental designs to use?

- The two-group before-after randomised experiment (both groups are measured before and after interventions' running period): use it when assessing both the absolute (observe changes/progress between the two subsequent measurements) and the relative effectiveness of each intervention (campaign alone vs. integrated campaign).
- Multiple time series randomised experiment (in both groups several measurements are taken both in the baseline and the follow-up period). Use it when:
 - the budget/time restrictions allow taking multiple measurements
 - observational data or accident statistics are used (self-reported data is rather inconvenient, because asking the same subjects the same questions on different occasions will probably result in high drop-out rates)
 - the internal validity has to be strengthened (and being more confident in the conclusions about the cause-effect relationships)

In the above examples, where the allocation of communities to specific (additional) actions is dictated by other decision-making practices (social, political, etc.) and is not randomised, the only-campaign, and integrated-campaign groups are considered to comprise quasi-experimental groups, thus the risks associated with quasi-experimental designs as described above such as lack of internal validity, are introduced. However, in spite of some weaknesses compared to experimental designs, quasi-experimental designs are more easily applicable in practice and thus more useful for practitioners. Instead of using a randomisation procedure in the examples cited above, one can simply pick out some units (communities, regions, schools, etc.) to deliver additional actions (e.g. education, enforcement). However, in such cases one should be careful with causal inferences (conclusions) as it might be some specific characteristics of these units that could have affected the results (and not only additional actions integrated with a media campaign).

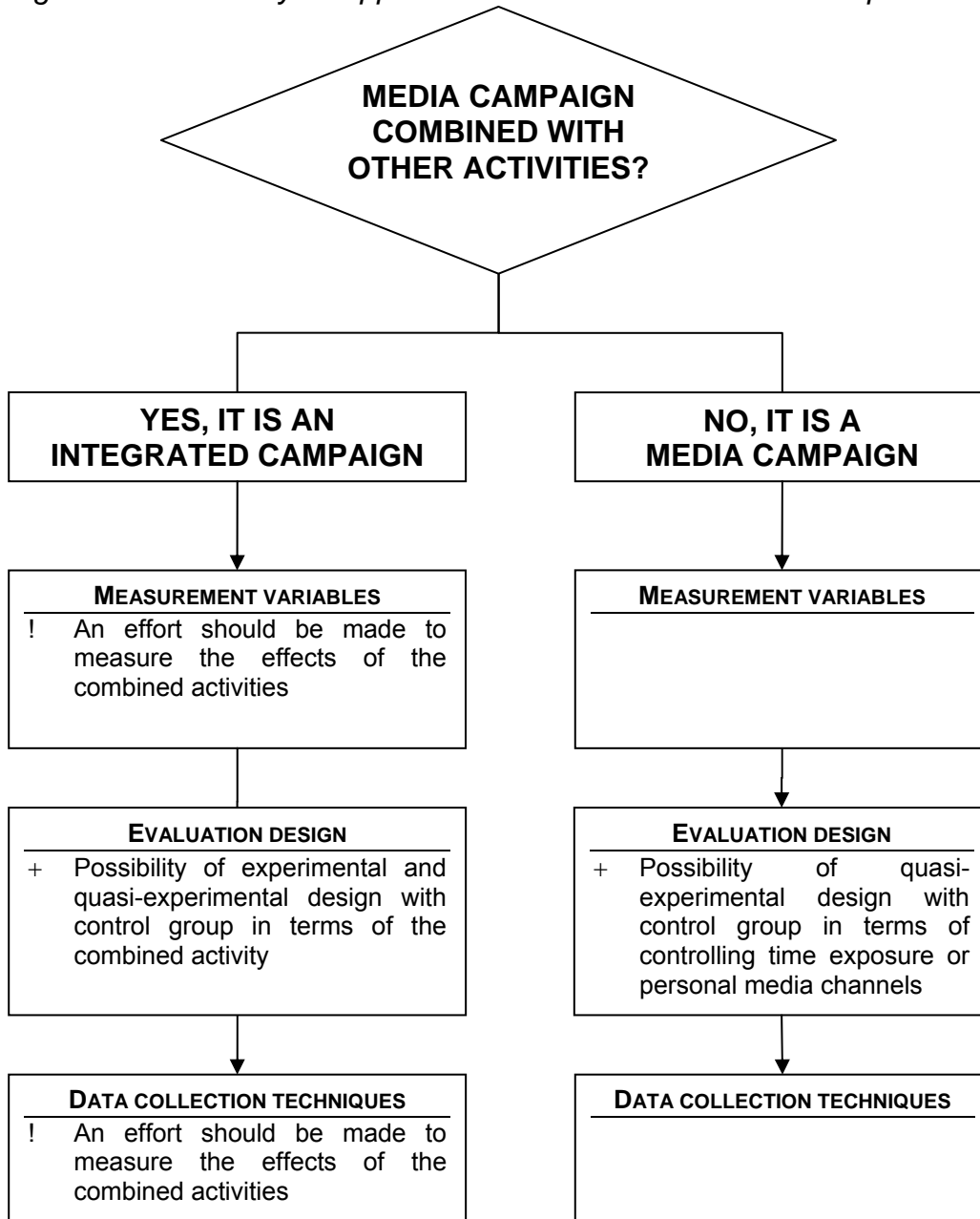
Somewhat more complicated quasi-experimental designs may also be used to evaluate the effects of integrated campaigns, especially when accompanying actions do not strictly overlap. One of them is the counterbalanced design which is useful for a comparison of the effectiveness of two or more interventions. In that case the required number of independent groups of subjects (or distinct geographical regions) equals the number of interventions to be compared. If for example two interventions are to be compared then group I first receives intervention I (I1) and afterwards intervention II (I2) while in group II the order of introducing interventions is reversed (first I2 and then I1).

Example: Counterbalanced design for an integrated campaign

Region 1: first a campaign with enforcement and then a campaign alone;
Region 2: first a campaign alone and then an integrated campaign.

Similarly, when different actions are introduced in a sequence (e.g. a media campaign alone is first implemented, then enforcement activities are added and somewhat later enforcement is again withdrawn) it is possible to use either the time series design or the equivalent time sample quasi experimental design. Both of them require several measurements – at least after each action is introduced or withdrawn. The advantage of these two designs is the possibility of using them with one single target group, also in case of national campaigns.

Figure 4.4: Summary – supportive activities and evaluation components



Note: (+): positive consequence for the evaluation component, (-): negative consequence for the evaluation component, (!) keep in mind.

5 Collecting data and data analysis

After the choice of the measurement variables, data collection method and techniques, and the evaluation design, it is time to plan the following steps of your research. First of all, you need to collect your data. This is related to the selected data collection technique that is largely dependent on the investigated subject(s), the existing knowledge and the data collection method chosen to conduct the study (see section 3.2). Set aside the concrete objectives of the study, the fact is that you need to describe the nature of a population – that is, a group or a class of subjects, variables, concepts, or phenomena. The usual procedure is to take a sample from the population. As sampling is a very determinative step in your evaluation study, we introduced sampling in a separate section 5.1. Raw data, in the form in which they have been collected, are rarely used directly to produce the results. A following step is that data need to be processed, edited, coded, or tabulated and then analysed. The results have to be understood and interpreted in accordance to the aim of the study. This phase will be discussed in section 5.3.

5.1 Introduction to sampling

To measure the effect on the target audience, it is necessary to conduct a study in order to describe the knowledge, attitude or behaviour of the targeted population. The process of examining every member of such a population is called a census. As restrictions on available time or money usually prohibit observation of an entire population, a sample will be taken. A sample is a subset of the population that is taken to be representative of the entire population (Wimmer, 1994). Assuming that a sample is chosen according to proper guidelines and is representative of the population, regardless its size, the results from a study using the sample can be generalised to the population.

A first step is to define the population to be studied (Kalton, 1983). Here the term 'population' is used in the technical sense of the totality of the elements under study, where the 'elements' are the units of analysis. The elements may be persons (e.g. holders of a driving license), but they could be alternatively be schools, companies, or any other unit. The population definition needs to be as precisely and specified according to the target group and the objectives of the campaign. Once the population has been defined, the question of taking a sample from it can be addressed. To collect data from only a part of the population is clearly less costly and, providing the estimates are sufficiently precise, sampling is thus more economic than organising a census. Unless the population is small, sampling is almost always used.

5.1.1 Sampling techniques

There are different ways to select a sample. A basic distinction to be made is whether the sample is selected by a probability mechanism or not. With a *probability sample*, each element has a known, nonzero chance of being included in the sample (e.g. Kalton, 1983; Wholey, 2004). Non-probability sampling covers a variety of

procedures, including the use of volunteers and the purposive choice of elements for the sample on the grounds that they are 'representative' of the population.

A variety of probability sampling techniques have been developed to provide efficient practical sample designs. An advantage of these techniques is that selection biases are avoided (see further), and statistical theory can be used to derive properties of the sample used to gain insight into characteristics of the population.

The most straightforward technique is called *simple random sampling* (e.g. Kalton, 1983; Peck, 2005; Scheaffer et al, 2006). *Simple random sampling* provides researchers with a sampling method that is objective and free of selection bias. This means that a simple random sample contains as much information on the population as the result of a census. A (simple) random sample is a sample chosen using a method that ensures that every individual member of the population has the same chance of being selected.

Example: Simple random sampling

Suppose that a study is to be conducted in a school to find out about the students' traffic safety habits. A list of the school's students is available, with the list being ordered by for example the students' identification numbers. One way to draw a sample would be by a lottery method. Each student's name or identification number is placed in an (imaginary) urn, they are thoroughly mixed, and then a certain number of them are selected at random.

When selecting a random sample, researchers can choose to do sampling with or without replacement. Sampling with replacement means that once an individual from the population is selected for inclusions in the sample and the corresponding data are recorded, the individual is placed back in the population and can be selected again in the sampling process. In practice, sampling with replacement is rarely used. Instead, the more common method is not to allow the same item to be included in the sample more than once. Sampling in this manner is called sampling without replacement. In some settings, however, alternative sampling methods may be less costly, easier to implement, or more accurate.

A commonly encountered feature is that a certain amount of information is known about the elements of the population to be studied. The essence of stratification is the classification of the population into subpopulation, or strata, based on some supplementary information (auxiliary variable). In *stratified random sampling*, a simple random sample is independently selected from each group (strata). The choice how the strata are formed is determined by the objectives for the stratification. The measure of success achieved in forming strata is their internal homogeneity with respect to the measurement variables of interest (Kalton, 1983).

Example: Stratified random sampling

Imagine that a study is organised to investigate truck drivers' safety behaviour. A possible way to collect data for the individual driver is to select truck drivers randomly selected at truck stops. These stops could be stratified on the size of truck stops approach by the number of parking spaces.

When the entire population can be divided into a set of non-overlapping subgroups, *stratified sampling* is often proved easier to implement and more cost-effective than random sampling. The real advantage of stratified sampling is that it often allows making more accurate inferences about a population than simple random sampling does. In general, it is much easier to estimate characteristics of a homogeneous group than of a heterogeneous group (Peck, 2005). Namely, when the strata are relatively homogeneous, a small sample from each stratum can provide reasonably accurate information about strata characteristics.

Sometimes it is easier (and more economical) to select groups of individuals from a population than it is to select individuals themselves. This is called *cluster sampling* and involves dividing the population of interest into non-overlapping subgroups, called clusters (for more information see D2.2 of the CAST project). Clusters are then selected at random, and all individuals in the selected clusters are included in the sample. Although we divide the population into groups for both cluster sampling and stratified random sampling, the techniques differ. In stratified sampling, we take a simple random sample within each group; in cluster sampling, we take a simple random sample of groups and then all individuals in the selected clusters are included in the sample (Peck, 2005). If only a sample of elements is taken from each selected cluster, the method is known as *two-stage* sampling (Kalton, 1983). Often a hierarchy of clusters is used: first some large clusters are selected (for example schools), next some smaller clusters are drawn within the selected larger clusters (for example classes within the selected schools), and so final elements are selected within the final-stage clusters (for example students within the selected classes). This is known as *multistage sampling*.

Example: Multistage random sampling

To minimise travel time and distance required conducting seat belt use observations, a multistage random sampling was chosen for this study. First, a random sample of geographical locations (such as provinces) was selected. Within each selected cluster, all roads were identified and several road sides were selected randomly. On these selected roads, all vehicles were observed in order to estimate the national safety belt usage rate.

As a consequence of the fact that only a sample of clusters being samples, the ones selected need to represent the ones unselected. To increase the chance of obtaining a sample that is representative of the population, it is best done when the clusters are as internally heterogeneous in the measurement variables of interest as possible (reflecting the variability in the population).

A lot of other sampling techniques exist. In case of a qualitative evaluation, Patton (2002) uses the term purposeful to describe the various sampling strategies possible. The choice depends in large part on what types of data are needed and who is best able to supply it. Among these choices are deviant case sampling, typical case sampling, maximum variation sampling, snowball sampling, convenience sampling, negative case sampling, and politically powerful sampling (Patton, 2002). To this lexicon we add sampling of experts or key informants. In describing each of these in BOX 5.1, we assume individuals are the sampling units.

BOX 5.1: Sampling techniques.

Deviant case sampling is driven by the need to learn about the outliers – persons (certain target group) who exemplify unusual successes or failures of the campaign. It can be used to identify a subgroup within a population for example drug users. By contrast, typical *case sampling* involves taking a sample of what one would call typical, normal or average for a particular phenomenon. It provides a cross section of a larger group.

Maximum variation sampling involves purposefully picking a wide range of variation on dimensions of interest. It is the choice when representativeness is aimed at cases that cut across wide variations in campaign processes and/or outcomes.

Snowball sampling is an essential when the population of interest is isolated, hard-to-reach, or suspicious of outsiders. Here a sample snowballs, or expands, by tapping into existing social networks.

Convenience sampling is undoubtedly the easiest approach, because it implies doing little more than taking advantage of cases at hand. Any number of constraints may steer the evaluation towards convenience sampling.

5.1.2 Sample size

One of the first questions that arises when sampling is: what sample size is needed? Determining an adequate sample size is one of the most controversial aspects of sampling. How large must a sample be to provide the desired level of confidence in the results? Unfortunately, there is no simple answer. There are suggested sample sizes for various statistical procedures, but no single sample-size formula or method is available for every research method or statistical procedure (Wimmer, 1994). The size of the sample required for a study depends on at least one or more of the following seven points:

- (1) project type;
- (2) project purpose;
- (3) project complexity;
- (4) amount of error willing to tolerate;
- (5) time constraints;
- (6) financial constraints; and
- (7) previous research in the area.

A few general principles guide researchers in determining an acceptable sample size. The random selection process allows being confident about selecting a representative sample. To determine an appropriate sample size, it is also necessary to specify the degree of precision required for the study (e.g. 95% probability). Another factor that needs to be included in the calculation of sample size is non-response. Usually, researchers can expect from 10%-25% of the sample to drop out of the study before it is completed. This adjustment serves only to produce the desired sample size; it does not address the problem of non-response bias (see further). Sample size is also controlled by cost and time.

An important consideration is the data collection technique used. For example: focus groups use samples of 6-12 people, but the results are intended to be generalised to the population from which the respondents were selected. A sample of 100 subjects per demographic group (such as adults 18-24 years old) is often used by researchers (Wimmer, 1994). Furthermore, unlike quantitative studies where there is a built-in stopping point (i.e. the instruments have been administered); qualitative data collection has no inherent endpoint (maximum sample size) beyond what is known saturation. Saturation occurs when the data analyses begin to reveal repetition and redundancy, when new data tend to confirm the existing findings rather than expand on them.

It is a common misconception that if the size of a sample is relatively small compared to the population size, the sample can't possibly reflect the population accurately. Sample data are usually selected with the intention of generalising results to the population from which the sample units were drawn. As a result, it is important that the sample is representative of the population (Peck, 2005). To be reasonably sure of this, the researchers must carefully consider the way in which the sample is selected. It is sometimes tempting to take the easy way out and to gather data in a haphazard way; but if a sample is chosen on the basis of convenience alone, it becomes impossible to interpret the resulting data with confidence. Furthermore, sample size may have a dramatic effect on an analysis (Wholey, J., 2004). When the sample size is very small, certain methods lose their statistical value.

As we know, both quantitative and qualitative methods can be used for collecting evaluation data. The conventional wisdom among evaluators is that qualitative and quantitative approaches have different strengths, weaknesses, and requirements that will affect evaluators' decisions about which techniques are best suited for their purposes. But the choice of an approach implies also the sample size. For all quantitative study types it is important to have sufficiently large numbers in the study to be sure that if an effect exists it is detectable (GRPS, 2007).

Information about the sample size is available in quite a few of published research. Consulting the work of other researchers provides a base from which to start. Generally speaking, the larger the sample used the better. However, a large unrepresentative sample is as meaningless as a small unrepresentative sample, so researchers should not consider numbers alone.

5.1.3 Bias in sampling

Bias in sampling arises when a sample tends to differ from the corresponding population in some systematic way (Peck, 2005). The amount that the sample mean differs from the population mean is called the *sampling error* (Wimmer, 2006). Bias can result from the way in which the sample is selected or from the way in which information is obtained once the sample has been chosen. The most common types of bias encountered in sampling situations are selection bias, measurement or response bias, and non-response bias (see BOX 5.2). A common misperception is that increasing the sample size is a way to reduce bias in observation studies, but this is not the case. For example, a larger sample size cannot compensate for response bias introduced by a poorly worded question.

BOX 5.2: Several types of bias

- *Selection bias* is introduced when the way the sample is selected systematically excludes some part of the population of interest if those who are excluded from the sampling process (e.g. those without telephone) differ in some systematic way from those who are included; the sample is guaranteed to be unrepresentative of the population.
- *Measurement or response bias* occurs when the method of observation tends to produce values that systematically differ from the true value in some way. This might happen if an improperly calibrated scale is used to weigh items or if questions on a survey are worded in a way that tends to influence the response. Other things that might result in response bias are the appearance or behaviour of the person asking the question, the group or organisation conducting the study, and the tendency for people not to be completely honest when asked about illegal behaviour or unpopular beliefs.
- *Non-response bias* occurs when responses are not actually obtained from all individuals selected for inclusion in the sample. If those who respond differ in important ways from those who do not respond, this type of bias can distort results. The biasing effect on the resulting sample is lowest when the response rate is high. To minimise non-response bias, it is critical that a serious effort be made to follow up with individuals who do not respond to an initial request for information. The non-response rate for surveys or opinion polls varies dramatically, depending on how the data are collected.

5.1.4 Non-response

Failure to collect the data from some sampled elements, or non-response, is a major survey problem that seems to have grown as the public has become less willing to participate in surveys (e.g. Kalton, 1983). The cause of concern about non-response is the risk that non-respondents with regard to the measurement variables, in which case the analyses on the respondents alone will be biased estimates of the overall population characteristics.

With interviews or face-to-face questionnaires, non-response can be classified into these categories: refusal to be interviewed; non-contact because the intended respondent is unavailable (not at home) or cannot be located; incapacity of the intended respondent to take part in the study for reasons such as illness, deafness, or inability to speak the language; and even completed questionnaires being lost in transit or processing. However, all that is known about non-response in case of most postal questionnaires is simply that the questionnaire has not been returned. The lack of response may be the result of one of several reasons, such as definite decision to refuse, a failure to get around to completing the questionnaire, or the questionnaire's failure to reach the respondent. The proportion of refusals varies greatly, depending on the subject of the study, the length of the questionnaire, and the skills of the research team. A variety of procedures is used in studies in an attempt to minimise the number of refusals, and even the choice of the data

collection technique is often influenced by the relative risks of refusals with different techniques.

Some examples:

- With interviews, interviewers are carefully trained in approaches to use to avoid refusals, and they are instructed to return to conduct an interview at a time more convenient to the respondent if necessary.
- Attempts to persuade the sample members of the value of the study; for example good sponsorship is likely to be particularly effective with a postal questionnaire.
- Assurances of anonymity and confidentiality are generally provided to eliminate any fears the respondents may have about the use of their responses.
- Questionnaires are usually organised to start with simple non-threatening questions to avoid the risk that the respondent will terminate to respond immediately.
- Sample members who are not at home in interview surveys should be treated by call backs. These call backs have to be made at different times of day, including some evening calls. Commonly interviewers are instructed to make at least four call backs if unable to contact a respondent (Kalton, 1983).
- A common strategy, in case of postal questionnaires, is to send a 'reminder letter' to those who have not replied after a given period and then to send a further reminder with a second copy of the questionnaire to those who still have not replied after a further period. This follow-up system has proved to be a valuable means of increasing the proportion of responses.

5.2 Collecting the data

The choice of a data collection method (e.g. method of asking, method of observing, and the method of document analysis) and the corresponding data collection technique (questionnaires, interviews, observations, accident statistics...) have been made. You need to consider what data already exists and what is necessary to be gathered in order to find an answer on your evaluation questions. If there was decided that new data of the targeted population must be collected, a sample size and type should be determined, this means who and how many persons, how, where and when. More information on sampling can be found in the former section. These are very important steps in an evaluation study as you will determine whether an implemented campaign has had the intended effect and reached the stated goals.

Collecting the data is a critical step in obtaining reliable information. When you collect data you should be attentive that the data are adequate in quantity and quality to produce a complete and credible analysis. It is important to keep in mind the questions that need to be answered on the basis of the resulting data (see section 3.1). Therefore you need to plan this step carefully. Who has the data or who can help you collect it? Consider also the extra work of data recoding. Sometimes your data is not readily collected or 'wrongly' aggregated in case of existing data. In the latter, you should check if the extra work is in proportion to the expected information to get from it. Planning needs to take into account:

- The expected quality especially regarding further data analysis

- Budget constraints
- Expected delivery dates - timing
- Personnel costs
- Sponsors and users of the data
- ...

The overall requirements should be clearly documented on beforehand. Think about the goal, cost, timing, scope, research design....

After the data are collected, the next step involves data analysis. Raw data (in the form in which they have been collected) are rarely used directly to produce the results. First, they have to be processed (evaluated for errors etc.), edited, coded, tabulated, and then analysed. Here rises the question if you will do the analysis yourself or if you need some assistance. Of course this will have consequences for the budget.

In practice data analysis begins before the data are collected (Haskins, 1993). For example, when designing questionnaires, researchers need to decide on the type of usable data the questionnaire responses will generate and thus pre-code the questions accordingly. The more organised one can be at the questionnaire construction stage, the easier the later tasks of data analysis, interpretation, and writing the research report will become.

5.3 Analysing the data

Once you have gathered all the information that you intend to collect, the next step is to examine it. Evaluators want to condense data into a meaningful outline that gives us insight into the effectiveness of the road safety campaign. The purpose of data analysis is to answer the evaluation questions (see section 3.1). It is very crucial to realise that data analysis is more than displaying the responses or data that were obtained (Logan, 2006):

- Are the preset hypotheses supported by the data;
- Do emerge meaningful trends or relationships out of the sometimes massive amounts of data;
- Are there statistically significant differences between before and after measurement or the control and the intervention group;
- ...

The time and effort required for data analysis and interpretation depends on the study's purpose and the methodology used (Wimmer, R.D., 1994). Analysis and interpretation may take several days to several months. Every analysis should be carefully planned and performed according to guidelines designed for that analysis. In practice, data analysis and interpretation usually are integrated activities carried out by the researcher/evaluator, or by an outside firm specialising in data tabulation and presentation (Haskins, 1993). The CAST project determined minimum standards for data analysis when evaluating road safety campaigns (see Box 5.3).

BOX 5.3: Minimum standards for data analysis.

- Describe the target group in terms of the relevant demographic variables (age, gender, education, etc.).

- Provide relevant descriptive statistics.
- Provide basic inferential statistics (significance testing and effect sizes)- explain if the difference between the two measurements is statistically significant (is it possible to generalise?) and what is the size of the effect.

The data analysis step requires the evaluator or researcher to select and apply the appropriate data analysis technique and statistical methods. The substantive questions identified to guide an evaluation, the data collection decisions made about how to measure the subject of interest (see also Box 5.4 about qualitative data analysis), and the type of audience the evaluator is addressing all affect selection of statistical techniques (Wholey, J., 2004). One consequence is that the way variables were measured limits the number of statistics available to evaluators. The most fundamental constraint is whether the variables were measured at the nominal, ordinal, or interval level of measurement⁵. For nominal measures, frequency distributions and contingency tables (that array frequency counts) are the most frequently used techniques for analysing data. For ordinal measures, contingency tables and frequency distributions are still the most likely choice for analysis, although some researchers prefer to treat ordinal measures as if they are equivalent to interval measures. However, unless an ordinal scale contains at least five values, it is probably best to treat the scale as a nominal measure. With interval measures evaluators have the widest range of alternatives.

BOX 5.4: Qualitative data analysis.

Qualitative data are in the form of text, written words, phrases, or symbols describing or representing people, actions, and events in social life. Qualitative data analysis can also be systematic and logically rigorous, although in a different way from quantitative or statistical analysis. It is less standardised, less abstract and often inductive (Neuman, 1997). Instead of testing a hypothesis, a qualitative study may illustrate in evidence showing that a theory, generalisation, or interpretation is plausible. Most often it comes down to two different tasks. The first of these involves generating codes, or conceptual themes, from the data and using these to guide additional analysis. The second approach does not fracture texts to extract meaning across cases but instead focuses on rich description of individual cases.

The following section gives only an introduction to data analysis. A wide variety of statistical tests, statistical techniques and examples can be found in other research reports, statistics textbooks or books of the marketing and social sciences field or academic journals. Nevertheless, this section gives a non-exhaustive overview of the main inferential statistics

⁵ Logan, T.K. et al. (2006): The nominal level of measurement is descriptive – meaning that names or terms are assigned to various categories that are sorted by a distinguishing characteristic (e.g. respondents might be grouped into those who are male or female, employed or unemployed etc.). Ordinal level of measurement uses rankings. The difference is that with ordinal data there is a clear hierarchy that represents gradations of amounts (e.g. strongly agree, agree, undecided, disagree and strongly disagree). Data measured at the interval level have known distances and equal intervals between the units (e.g. age questions will furnish continuous or interval data when no response categories are provided).

5.3.1 *Why are statistics necessary?*

Data collected are often in the form of numbers. These numbers represent values of variables, which measure characteristics of subjects, respondents, or other cases. These values have to be understood and interpreted in accordance to the objectives of the campaign, conclusions should be drawn which may suggest an answer to the effectiveness question, and/or new problems could be identified. Therefore it is the task of the evaluator to do several things to the raw data in order to see what they can say about the evaluation questions and the preset hypotheses.

To conduct an evaluation study with valid and reliable results you should base your research on a scientific method. This implies the necessity of statistics. Statistics are mathematical methods to collect, organise, summarise, and analyse data (Wimmer, 2006). As your data is coupled with the appropriate research design (as it should be) it can provide objective evidence that the campaign was or was not successful (information not depending on the evaluators' judgement) provided that statistical procedures are used in analysing the data. Moreover statistical procedures lend credibility and professionalism to your final report (Logan, 2006). Tukey (1986) identifies four purposes of statistics:

- to help summarising
- to help understand what is going on
- to extract information from the data
- to aid in communication

Researchers must determine also through analysis whether their work is valid internally and externally. To assess internal validity, one has to assess whether the study really investigates the proposed research question. Control over research conditions is necessary to enable researchers to rule out all plausible rival explanations of results. Any such variable that creates a rival explanation of results is called an extraneous or confounding variable. The presence of such variables indicates that the study has failed to investigate its hypothesis. These variables may arise from several sources such as history, maturation, instrumentation, sample selection, and so on. External validity refers to how well the results of a study can be generalised across the populations, settings, and time (Cook & Campbell, 1979). The external validity of a study can be severely affected by the interaction in an analysis of variables such as subject selection, instrumentation, and experimental conditions (Campbell, 1963). Most procedures to guard against external validity relate to sample selection. Cook and Campbell (1979) describe following considerations: use random, representative, and heterogeneous samples and replicate the study several times.

If statistics (such as averages, frequency distributions, measures of variations, multiple regression analysis etc.) are used to describe a group of items, the figures presented are called *descriptive statistics* (Wholey, 2004; Neuman, 1997). But researchers often want to do more than describing. *Inferential statistics* use the probability theory to test hypotheses formally, permit inferences from a sample to a population, and test whether descriptive results are likely to be due to random factors or to a real relationship. To illustrate this difference, following box (Box 5.5) gives examples of possible research questions. The next two paragraphs explain further what is meant by descriptive and inferential statistics.

BOX 5.5: Descriptive and inferential research questions.

Descriptive questions

- What is the average rate seat belt use?
- How many persons have seen the road safety campaign?
- Is the new traffic law known?
- What is the attitude towards speeding?

Inferential questions

- Does gender relate to speeding?
- Is the attitude regarding speeding changed after the campaign?

5.3.2 Descriptive statistics

The first step in making sense of a series of numbers resulting from the collection of data is to describe them (Haskins, 1993; Logan, 2006). Evaluators usually start their data analysis step by requesting ordered arrays of the data one variable at a time (also called univariate analysis). For each value you (or the computer) have to count how many times (or how frequently) it occurs. If the frequency distributions are produced by a computer program (such as SPSS) an advantage is that the evaluator can request easily measures of central tendency (see further) (Logan, 2006).

Reasons for looking at each variable individual or for descriptive analysis:

- to determine if any errors were made in preparing or recording the data, sometimes it will indicate that a certain group of individuals was inadvertently missed or that the range of responses was restricted;
- to condense data sets for easier interpretation (Wimmer, 2006).

Campaign evaluators can use two principal methods to make the collected data more manageable (see Box 5.6): summary statistics and data distributions (Wimmer, 2006). One category of summary statistics are the measures of central tendency such as mean, median, or mode. They determine the typical score of a distribution.

BOX 5.6: Explanation terminology summary and distribution statistics.

Summary statistics

- *Mean*: the actual numeric or arithmetic average of scores, calculated by the sum of the values divided by the number of answers, easily distorted when extreme scores emerge.
- *Median*: the middle value or the midpoint of the frequency distribution, this implies that half of the scores lie above it and half lie below the median position.
- *Mode*: the most common or most frequently score, value or category in the distribution.
- *Frequencies*: gives the occurrence of each answer thus by counting the number of times each value occurs.
- *Percentages*: to compare scores from different population samples (of varying sizes) it is better to determine percentages of each score.

- *Proportion*: you select one value of each variable and compute its rate of occurrence (or proportion), for example the ratio of the number of accidents among 18-25 year old persons and the total number of accidents.

Distribution or dispersion statistics

- *Sample distribution*: is the distribution of some characteristics measured on the individuals of analysis that were part of a sample.
- *Range*: the distance between the highest and the lowest value in a distribution, be careful as the range often increases with the sample size because larger samples tend to include more extreme values (or outliers) (Wimmer, 2006).
- *Standard deviation*: it represents information about how much scores deviate or vary from the mean of the distribution; it is defined as the square root of the variance (Scheaffer, 2006; Haskins, 1993; Logan, 2006).
- *Variance*: a mathematical index of the degree scores tend to differ from each other, a small variance indicates that most of the scores in the distribution lie fairly close to the mean thus the scores are very much the same; a larger variance represents widely scattered scores.

5.3.3 Inferential statistics

Once you have edited the data, corrected any mistakes, and learned what you can from the descriptive analysis, you are ready to begin looking at variables two at a time. Evaluators do this to test hypotheses or to examine the strength of association (Logan, 2006). Besides you need to make the best guess you can about the population based on the evidence provided by the sample. Inferential statistics are the appropriate statistical procedures for helping you reach that goal.

To apply inferential statistics first, a *statistical hypothesis* (is a statement about what might be true) identifying the relationship between any two variables of interest must be specified. There are two alternatives: either the variables are not related in the population, or they are related. To test that there is no relationship between two variables, a *null hypothesis* is stated. This hypothesis of exactly no relationship (the null hypothesis) is tested in the hope of rejecting it. General examples of the null hypothesis are:

- to test differences in intervention and control groups: the assumption is that no difference exists between the two groups for the variable being compared; or
- more general in campaign evaluation; the campaign has no effect in achieving the intended outcome(s).

When the null hypothesis is rejected, the sample data do permit a conclusion that a difference exists or that the campaign is effective.

To control if the results are not likely to be due to chance factors (are the differences between before and after measurements real?), the *statistical significance* is determined. As a consequence statistical significance is the minimum requirement for a meaningful result. When measuring the effect of a road safety campaign, the first assessment of the magnitude of the effect is thus testing statistical significance (Rossi, 2004). Furthermore it indicates the probability that the null hypothesis is falsely rejected or accepted (see BOX 5.7).

BOX 5.7: Two categories of statistical error.

- *Type I error*: a statistical conclusion error that rejects the null hypothesis in error when it is true, a decision known as *false positive*. Type I errors lead you to conclude that a relationship exists between two variables. The significant differences were found just by chance. A Type I error is very problematic because it claims effective campaigning when in reality the influence of the intervention could be negligible.
- *Type II error*: a statistical conclusion error that says that a relationship does not exist, when in fact it does (falsely accepting a null hypothesis). Type II errors lead you to conclude that there is no relationship between two variables. But in fact, there is a relationship. Type II error claims that no differences exist between two groups (e.g. a treated one and an untreated one) when in reality a reliable difference (i.e. a statistically significant one) does exist. It is possible to estimate the probability of committing a Type II error when using different statistical tests

Source: Logan, 2006

Both errors Type I and II are to be avoided. Therefore a specified decision rule states how confident the evaluator wishes to be that a false position will not occur. This decision rule provides the *confidence (or significance) level* for the test. The confidence level reflects the amount of evidence evaluators want to have to ensure that they are correct in concluding that the campaign does produce the observed effect. In the physical and social sciences a 95 percent confidence level is conventionally used as a decision rule for testing statistical hypothesis. For many public campaign purposes, 95 percent may be excessive (Wholey, 2004). Conclusions for which evaluators are 80 percent or 90 percent confident may be quite adequate and reduce the size of sample needed, thereby reducing costs.

As said before, hypothesis testing is a method of making statistical decisions by using research data. The data actually obtained from the research (actual data) will be compared to the data that should have been obtained if the null hypothesis was true (expected data). The size of the discrepancy of the actual data and the expected data is summarised in one figure, the *testing statistic*. Generally statistics are divided in two branches namely parametric and nonparametric. To select the appropriate test, you should take into account the type of data and the goal. More information regarding this distinction can be found in BOX 5.8. The *p-value* is the probability that, if the null hypothesis is true, random chance could give you the same results. It has a value ranging from 0 to 1. If the p-value is smaller than or equal to the significance level, the null hypothesis will be rejected.

BOX 5.8: Distinction between parametric and nonparametric statistics

<i>Nonparametric statistics</i>	<i>Parametric statistics</i>
- are appropriate with only nominal and ordinal data	- are appropriate for interval and ratio data
- the results cannot be generalised to the population	- possible to generalise to the population
- no assumption about normality (distribution free)	- assumption about normally distributed data

If your sample is very small (i.e. fewer than 10 subjects), then it cannot be assumed that the population is normally distributed and thus you must in theory use nonparametric statistical procedures. These techniques are also used when the data consist of ranks (instead of interval data). Nowadays most researchers agree that both categories can be used with all types of data and that both are appropriate for generalising results to the population.

Source: Logan, 2006.

Commonly used parametric tests are t-test, analysis of variance, Pearson correlation test, and regression analysis. An example of nonparametric statistics is the chi-square test. In Box 5.9 you will find a short description of a couple statistics. The most basic parametric statistic and widely used in all mass media research is the t-test (Wimmer 2006).

BOX 5.9: Examples of nonparametric and parametric statistics.

- *T-test:*

The most elementary method for comparing the mean scores of two groups to test if a statistically difference exist between the two groups. In other words, the t-test allows researchers to investigate the effects of one independent variable on two samples of people. In case you compare two paired groups (for example recall scores of males and females, before and after measurement of the target group...of the same campaign) the appropriate statistical procedure is the *paired samples t-test*. In case you assessed your campaign with the aid of a control group the appropriate statistical procedure is the *t-test for independent samples or the unpaired t-test*.

In practice: it is necessary to consult a t-distribution table to estimate the significance. In case the calculated t equals or exceeds the table value, you may conclude that the differences between the groups are significant.

- *ANOVA*

In case data on three or more groups had been collected, and you need to make comparisons between two or more means, the analysis of variance is the appropriate statistical tool. With the aid of this tool the observed variance will be partitioned into components due to different variables. *One-way ANOVA* investigates one independent variable; *two-way ANOVA* investigated two independent variables. The two-way ANOVA can save time and resources because studies for each independent variable are conducted at the same time (Wimmer, 2006). Both tools calculate the ratio of variance namely the *F-ratio*.

In practice: if this F value equals or exceeds the F-distribution table value (which you can find on the Internet) then the ANOVA is considered to be statistically significant.

- *Pearson correlation*

The most common statistic to measure the degree of relationship between two variables is the Pearson correlation. In fact, a correlation expresses whether two variables tend to increase or decrease together. Although correlations are limited in what they can reveal, they can still be useful in understanding data. Example: to convince the readers of an evaluation report that a campaign was worthwhile, you can verify if the campaign correlates with successful outcomes.

In practice: the Pearson correlation value varies between -1.00 (perfect negative relationship) and +1.00 (perfect positive correlation). If you square the Pearson r , this will give the percentage known of information necessary to make a perfect prediction from one variable to another. Other kinds of correlation coefficients are Spearman's rho and Kendall's W.

- *Linear Regression Analysis:*

This is a collective name for techniques to analyse the relationship two or more independent variables (or predictors) and a dependent (measured) variable. In other words, regression will predict the dependent variable using information derived from an analysis of the independent variables (namely by weighted linear combinations of them). Thus a formula will be developed that explains as much variance in the dependent variable as possible. The correlation between the predicted dependent variable and the weighted linear combination of the independent variables is represented by the coefficient of correlation R . However, R^2 calculated from a sample tends to overestimate the population value of R^2 and this bias increases as the ratio of independent variables to sample size increases. Adjusted R^2 takes this bias into account (Hankins, French & Horne, 2000).

In practice: the R^2 indicates the proportion of variance in the independent variable(s). The higher R^2 (the closer to 1.00), the more accurate the prediction is considered to be. However, in instances when several independent variables are included in the regression analysis, for example when theoretical models such as the theory of planned behaviour is applied, it is recommended to use Adjusted R^2 (see Hankins et al., 2000, for further statistical guidelines regarding the use of linear regression analysis).

- *Logistic Regression Analysis*

When the dependent variable is qualitative rather than quantitative, for example when participants were either wearing a seatbelt or not, the proper way to predict their behaviour is by using logistic regression analysis (e.g., Skog, 2004). A central concept in logistic regression analysis is *odds ratio*. The odds-ratio is a parameter which indicates how many times larger (or smaller) the odds are when the independent variable increase with one unit. An odds-ratio equal to 1 indicates that the odds do not change as a result of an increase in the independent variable (i.e., no relationship). When the odds-ratio is greater than 1, the odds increase as a function of an increase in the independent variable (i.e., positive relationship). Finally, when the odds-ratio is smaller than 1, the odds decrease as a result of an increase in the independent variable. The most common and intuitive statistical tests used in logistic analysis is the Wald-test (cf. Skog, 2004), a test which is similar to the t -test used in linear regression.

- *Chi-square*

If you want to determine whether the change in frequency is actually significant, you can perform a chi-square test. It is used in situations where you can cut the data in two groupings (e.g. persons who have improved and persons who haven't or a more successful effect with women than with male). A chi-square is a simple value that shows the relationship between the expected and the observed frequencies. Example: to determine the relationship between gender and use of media to obtain information about the campaign, crosstab analysis with a chi-square test is frequently used in mass media research.

In practice: to determine whether the chi-square value represents a significant difference in frequencies, you need to consult a chi-square significance table. If the calculated chi-square value equals or exceeds the value found in the tables, the differences are considered to be statistically significant. In literature you can find the limitations when using this test which you need to take into account.

5.3.4 Magnitude of a campaign effect

Next to showing that the campaign effect is statistically significant, a separate judgement should be made regarding the magnitude of the effect that is being measured (Wholey, 2004). Judgements about the size of the effect (slight, moderate, strong) reflect the evaluators view on the practical importance of the measured effect. There are no standards available for evaluators to use when interpreting the magnitude of the size of the observed effect (or observed relationship between two or more measures). The best way to evaluate such numbers is to compare them to appropriate referents such as comparable figures for previous years, for other administrative units, or for comparable campaigns. Appropriate and meaningful comparisons are absolutely essential to lend credibility to measures of magnitude.

5.3.5 Reporting statistics

Next to failing in using statistics when they are needed, it is also feasible to inundate the readers of your evaluation report with too much data, tables and results of loads of statistical tests. To avoid this, it is essential to consider the audience who read your evaluation report (Logan, 2006). Evaluators should keep in mind the pragmatic use of the findings. On the Internet you can find examples illustrating how to report statistics in APA style in case of a research report. Sometimes it can also be important to report non significant statistically relationships especially those that result in recommendations or suggestions to the management (Logan, 2006). In box 5.10 you will find a non exhaustive list of do's and don'ts. But, the CAST project recommends to report the precise p-values to accommodate the future meta-analyses (see reference Vaa, T., et al, 2009).

BOX 5.10: Do's and don'ts when using statistics in reports

- don't present statistics that are not likely to understand

- write that statistically significant differences were found but don't report the actual t or chi-square values
- don't show any formulae used in the calculation of statistics
- present information that is important to understand the major findings
- don't add everything contained in the computer printout

Source: Logan, 2006

6 Reporting and dissemination of results

The final stage of the research process is reporting the results and conclusions, suggesting solutions, formulating new theories, hypotheses.... A campaign evaluation report should provide all the information needed to understand the outcome in a language that is comprehensive for people outside of the research team who performed the evaluation study. More details are described in this chapter.

There are many conventions regarding the organisation of research papers (such as evaluation reports) and writing styles (Herzog, 1996). This knowledge allows communicating clearly, concisely, and with professionalism to both peers in the sciences and the general public. Besides, given the wide variety of approaches to research, it stands to reason that the approaches to writing a research report are equally varied (Wimmer, 1994). Although there is no standardised outline to follow in writing an evaluation report, it is always a good idea to address the evaluation's goals at the very beginning (Logan, 2006). The first two paragraphs (6.1 & 6.2) cover two topics: the content and the writing style. The last paragraph (6.3) is about dissemination.

6.1 Major sections of an evaluation report

As already mentioned, it is important to write a detailed and structured report. The information as well as the order in which this information is presented should be standardised (see Box 6.1). This way it is easy for the reader to find the exact information he or she is looking for and it is also ensured that all essential information is included in the report. This again allows campaign makers and evaluators to learn from the reports and facilitates future meta-analyses (Vaa, T. et al, 2009). Therefore the reporting tool (Boulanger, A. et al, 2009) was developed which provides a step by step guideline on how to write an evaluation report. An introduction to this tool is presented below. It is recommended to obtain a copy of the tool when writing an evaluation report as it provides boxes with examples of the type of information required and space to write notes to aid the report writing.

Usually a report starts with an abstract even if it contains an executive summary. The abstract is very brief (max. 200 words) and only states the absolute essential information including the conclusions and recommendations. An English version of the abstract should be provided to ensure that the international community has access to the information.

The sections *introduction and description of the campaign design* of the evaluation report set the stage for the evaluation and provide enough information to allow the reader to understand what the study was about. The first job of the report writer is to provide some information about the background and the nature of the investigated problem (e.g. the danger connected to particular traffic behaviour).

In this first part of the full report, the description of the road safety campaign concerned (its history, goals, target group, media plan...) and the goal of the evaluation will be described. The background information of the campaign (former

campaigns, literature review, statistics...), can be followed by a discussion of the ways the problem has been studied in the past. These methods could have had a direct bearing (or not) on the approaches used in the campaign and the way the data are analysed. The information should be accurate and relevant. Special emphasis should be placed on the objectives of the research study as it indicates the limits of the study's intended investigation (Haskins, 1993). Just like the objectives for the road safety campaign, the objectives for the evaluation study should be articulated clearly and agreed to by those parties undertaking the research.

This leads the reader to the *evaluation study and methodology* section. The report should provide enough information about the evaluation methodology components, including the research design, the data collection techniques, sampling techniques, and data analysis including the strategies for enhancing the rigor of the study. In other words, the methodology section of the paper describes in detail how the evaluation was conducted. The section should furnish the reader with enough information to allow another evaluator to replicate the study. Commonly, subsections and subheadings are used to differentiate the various components of the methodology section.

One subsection describes how the study has been carried out. It should include: the research design, the selected measurement variables including the evaluation questions and description of control techniques used. Each measurement variable requires some justification for its use; variables may not be added without reason. Every evaluation report also requires a description of the chosen data collection techniques.

Another subsection describes the sample of participants and should contain the following information: sample size, response rate, relevant sample characteristics (such as the male/female compositions, average age, and other demographic information but also geographic information) of the participants, and details of the sampling procedure (related to the used research design), including qualifying criteria and controls on sampling.

Some authors discuss in an optional subsection how they plan to or did analyse their data as data analysis should have been planned at the time that the evaluation design was considered.

Once the methodology section has described all the important evaluation components, it is time to address the findings in the results section.

The *results* section of the report is the heart of the report (Haskins, 1993) and describes what was found in the study and usually summarises the statistical analysis of the data (Herzog, 1996). The techniques used to analyse the data should be mentioned. But just the facts are presented in this section – statistically significant differences, results of before and after measurements, and so on. The challenge for the evaluator is to provide a user-friendly explanation of all decisions made and a critical assessment of the statistical accuracy that the tests can be reasonably expected to provide. The section also provides the answers to the questions that were previously posed (has the campaign reached its objectives?).

Here follows a list of some guidelines:

- Write the results in a clear and concise manner appropriate to both the research question and the individuals who will read the report.

- Decide which findings are the most important or have the greatest implications. One should be aware of the danger of trying to report every statistically significant correlation or t-test.
- Adapt the level of detailed information of statistical results to the target audiences' expectations and preferences. In case the campaign being evaluated is too complex for the audience and if there are numerous important findings, prepare a technical report and give it to the group decision makers and other interested members of the public.
- Report clearly both statistical significance and the size of the campaign effects. Both findings should be reported and interpreted for the audience (Wholey, 2004). For example, a difference between treatment and control groups may be minuscule, yet be statistically significant at a specified confidence level.
- Communicate next to positive findings, also unfavourable or ambiguous findings. It is important to identify areas of strength and areas for potential revision. However, establishing such relationships may not be feasible within some evaluation approaches or in large national evaluation with legislative or other policy making audiences. Although in some instances stakeholders might be able to help clarify some of the ambiguity found in the evaluation study.
- Never conceal information that might influence the interpretation of their findings (Wimmer 1994). For example, if 2 weeks elapsed between the testing of the experimental group and the testing of the control group, this delay should be reported so that other researchers can discount the effects of history and maturation on the results. Every research report should contain a full and complete description of the evaluation methodology and of any departure from standard procedures.
- Provide user friendly tables and graphs (providing statistical results). The writer must make sure that they are well-crafted and that their purpose is clear. They should be presented simply and be well-labelled. If tables and figures are used, there must also be some text describing them.
- Avoid unfamiliar abbreviations, acronyms, and other jargon as it confuses the readers (see also section 6.2).

Before concluding on the results they should be discussed. It is important to remember that it is not unusual that only the discussion, conclusion and recommendation part of the report is read. It is therefore important to write these very carefully.

When writing the *discussion section* it is important to try and explain the results both desired and undesired. When doing this all factors that could have influenced the results should be reported and their possible influence on the results explained. The most important factors are elements of the campaign itself – what worked and why, what did not work and why. But also elements from the evaluation itself can influence the results. Thus, limitations (and strengths) of the study should be discussed. The evaluator should provide an objective assessment of any inherent flaws or problems in the research design, any problems in the data collecting or analysis phase, and any other factors that might have affected the outcome of the study or limited its generalisability. Finally, external factors may influence the results of the campaign or the behaviour of the target group for example if a major road accident which has received a lot of media coverage occurred, while a speeding campaign was running, it may have influenced people to drive slower and thus inflated the effect of the campaign itself.

Example: Possible content of the discussion section

- If your findings run counter what was predicted or expected, provide an explanation for why results turned out the way they did.
- If you discover there was some source of bias what was not obvious at the time the study was planned, identify it.
- If the study was not implemented in a way consistent with the 'manual' or as originally had been proposed, point out the major limitations of the study.
- If time and space permit, you may link your findings to other evaluations of similar campaigns to contextualise them for your audience.

After having discussed possible explanations for the results you have to write some clear *conclusions*. As mentioned above it is all too often the only section thoroughly read by the audience. It is the place to put the most concise and insightful summary. When concluding it is important to relate to the objectives of the campaign and conclude whether these were met. The conclusion section clearly explains which hypotheses were supported and which were not. If a cost benefit or cost effectiveness analysis was performed conclude if the effect or benefit justified the costs. In the conclusion it is not unusual to have some ambiguous findings; definitive answers to all evaluation questions may not have been obtained.

Recommendations are one of the most critical products of an evaluation. The *recommendations* section handles the implications of the findings (what the findings may mean practically). In writing the recommendations, the evaluator must review the objectives of the study, the situation analysis with respect to the object of investigation, including the findings and conclusions. Because they are suggestions for decision making, the evaluator must be very careful to present well-considered recommendations in as objectives a manner as possible and including all of the known caveats and qualifications (Haskins, 1993). Possibilities range from highly specific recommendations that dictate a particular action or choice to more general recommendations framed as alternatives to consider (Shaw et al, 2006). Highly specific recommendations should be handled carefully as it concerns the extent to which the evaluator goes beyond areas of evaluation expertise and ventures into a consulting role requiring specific campaign expertise. In any case, they should be bounded by the constraints of the results of the evaluative data and the specific implications derived from them. An advice is to draw a clear, physical boundary between recommendations and conclusions (Shaw et al, 2006). If relevant it is also an option to give recommendations regarding campaign evaluations.

Finally, there is also the issue of summaries. One strongly advocated summary is popularly referred to as the *executive summary* (mostly at the beginning of the report). The time and patience of potential recipients of evaluation reports is frequently limited. When confronted with a massive and sometimes imposing report, there may be a tendency to avoid it. Evaluators can help ease the load by guiding potential readers through the report via the executive summary (Shaw et al, 2006). It is suggested in the Evaluation Standards (Joint committee, 1994) that the executive summary should contain condensed versions of all of the sections of the full report. Although it is also advisable to think of the executive summary as a one-page or two-page guidebook for learning about the most important things that stakeholders and

other readers might want to know. To give some advices, an executive summary should:

- focus primarily on the evaluation results and recommendation (if any);
- be ordered in a way that corresponds with the priority of potential readers' interests;
- go right to the point and state conclusions quickly;
- employ descriptive phrases only long enough to convey the essential idea;
- use bullets, stars, or other highlighting devices, along with indentations and differences in typeface and size; and
- serve also as a reference or a guide for the further report (a referencing summary).

Moreover, it is highly recommended to write an English version of the executive summary to ensure that the international community also has an opportunity to learn from the results of the evaluation.

BOX 6.1: Summary of the structure and content of an campaign report		
Main structure report	Sections of the report	Content of sections
Basic elements to start a report	Front page	
	Acknowledgements	
	Abstract and keywords	- Very brief overview of the absolute essential information (including the conclusions and recommendations)
	Table of content	
	Executive summary	- Structured overview of the content of the report following the chapters of the report - (Including general results, conclusions and recommendations)
Background	Introduction	- Background information (former campaigns, literature review, statistics, situational analysis...) - Problem definition (target audience, theoretical model...) - Goals and objectives of the campaign - Purpose of the evaluation
Campaign strategy	Description of the campaign design	- Description of the road safety campaign (slogan, timing...) - How does it address the targeted problem (message, media plan, supportive actions...)
Evaluation	The evaluation study and methodology	- Types of evaluations: process, outcome, and/or economic - Research design of the evaluation - Measurement variables and

		evaluation questions - Data collection techniques - Sampling techniques - Validity and reliability
	Data analyses and results	- Data analyses performed - Results of the data analyses - Answers to evaluation questions - Charts and graphs of findings - Cost benefit / cost effectiveness analysis
	Discussion	- Which factors could have affected the results and in which way
Conclusions & main recommendations	Conclusion	- Conclusions about campaign effectiveness - Campaign recommendations
Basic elements to end a report	References	
	Appendices	

6.2 Writing style

In all evaluation report writing it is important that the writer has a sense of the reader, and an evaluator's most important readers are potential users and decision makers. This section offers a few suggestions and guidelines (e.g., Shaw et al, 2006).

- *Write for the audience;*
In all writing it is important to write for those who will be the readers of the evaluation report. While close knowledge of the stakeholders enhances communication, such knowledge is not always possible. Sometimes the report is written for policy decision makers, or for a nationwide audience.
- *Use an accessible style and avoid jargon;*
Evaluation literature addresses the importance of writing in a clear, accessible style, a style that is direct, relatively free of jargon, and not convoluted with an abundance of complex syntactic and passive voice constructions. Research reports are almost always written in third person. It is best to avoid the specialised, professional language of evaluation and its related fields when writing the report. Of course, there may be times, when some jargon is necessary. The writer should then define it in context. Remember that the goal of scientific writing is to explain findings clearly, simply, and accurately (Wimmer, 1994).
- *Use illustrations and metaphors;*
Communication can be enhanced through the strategic use of illustration and metaphor. The value of metaphor is the vivid comparison that can help a reader to understand a complex process or vexing situation.

6.3 Dissemination of results

One of the most disappointing things that can happen to an evaluator is that the evaluation report, after weeks of hard work, is placed on a bookshelf or filed away to be seldom noticed or referred to again. When planning an evaluation, think about issues of dissemination and utilisation of the data by others (see also Box 6.2).

Once an evaluation is complete it is important to provide feedback to the stakeholders involved in the campaign (Drinking, 2007). There are a variety of audiences for evaluative reports – practitioners, policy makers, researchers, and the general public (Logan, 2006). Dissemination of the results will help garner further support for the campaign if it is successful, and help others gain support for the introduction of similar campaigns. Publicity from dissemination activities may also increase the impact of the campaign. If the campaign has not been successful it is important to share this with others so that weaknesses or relevant issues are considered in other similar interventions, including whether or not to introduce such interventions. Dissemination may involve presenting the results at public meetings, using the media to publicise the outcomes of the campaign, or publishing reports and papers in the scientific literature (Drinking, 2007). A variety of written, verbal and visual presentations might be needed for various audiences (Robson, 2001).

BOX 6.2: Four issues that increase the chances that an evaluation report will be used.

- Present the report in a manner that gets and holds the attention of readers and stakeholders.
- Incorporate the needs and concerns of the campaign designers and policy makers.
- Deal with negative or unexpected findings (present of course some positive information).
- Use the evaluation as a building process (e.g. changes have often budget implications thus it is impossible to change a campaign overnight).

Source: Logan, 2006

6.4 Use for future campaigns

In addition to providing information on the effectiveness of the road safety communication campaign, the evaluation is also of significant benefit to future campaigns. Before starting designing a road safety campaign, practitioners should constantly consider previous campaigns which have been thoroughly evaluated.

Here follows an overview of the possible use of the evaluation study for future road safety campaigns. In general it will improve the implementation of future campaigns by:

- minimizing the risk of implementing inappropriate future campaigns;
- giving information on how to reach future similar target groups;
- determining pertinent theoretical frameworks;

- demonstrate accountability to the funding sources, the stakeholders, policy makers and the public;
- collecting good ideas and avoiding wrong ones;
- facilitating future fundraising;
- anticipating both the means and budget to overcome problems common to road safety campaigns (Delhomme, P. et al, 2009).

Practitioners should note that all elements of a campaign evaluation can be of extreme utility as for the effectiveness of future road safety campaigns. Even campaigns without any major effect offer valuable information as their evaluation signposts pitfalls to avoid. Furthermore, you should always be careful by copying the exact campaign and evaluation design in other countries, regions....

Conclusion

A small investigation of the current situation of the evaluation reports revealed the strengths and the weaknesses of these reports. The most striking weaknesses refer to the following issues: not clearly defined campaign objectives and thus unclear success criteria, inconsistencies between the campaign objectives and the used measurement variables, no profound data analysis, no information on the evaluation costs, and missing information on the respective supportive activities. To deal with these shortcomings, CAST provided this evaluation tool for campaign practitioners. The aim of this tool is to provide practical guidelines and recommendations on how to carry out a proper effectiveness evaluation study.

The evaluation of road safety campaigns is essential to establish whether these campaigns are effective at improving road user safety. Conducting evaluation studies cost money but the importance of it may not be undervalued. Therefore, it is necessary to allocate a certain amount of time and budget resources specifically for the evaluation study. This budget should be clearly distinguished from the total campaign budget.

When evaluating, the first step is to choose the appropriate outcome evaluation methodology. This is dependent on the specific characteristics of the concerned campaign. Speaking about the proper evaluation methodology you need to take decisions about the measurement variables, the research design, and the data collection method (and techniques). The choices and decisions are strongly influenced by the scope of the campaign, the target group, the campaign objectives and the supportive activities. Each of these campaign characteristics has its particular implications for the evaluation study. Besides, all the components of the evaluation are interrelated. During the decision process of the evaluation components, you need to take into account the following general CAST recommendations:

- Assess the right measurement variables in relation to the pre-set objectives of a campaign in order to determine the success of the campaign.
- Choose an appropriate design to evaluate the impact of an intervention, a certain degree of trading between rigor and applicability (internal vs. external validity) must usually be considered.
- Keep in mind that the advantages and disadvantages of the different data collection techniques (can) complement each other.

In practice, the best or ideal evaluation methodology is not always practical feasible and therefore CAST stated minimum standards. The purpose is to ensure some satisfactory level of evaluation procedures throughout Europe. Thus this is the least one should do when evaluating a campaign. Of course, it is always highly recommended to do more than what is defined by minimum standards.

Here you find a summary of the CAST minimum standards:

What to measure?

- Always measure the campaign exposure.
- Always measure the campaign objectives plus behaviour (self reported or observed).

- In case the campaign is based on a theoretical model: measure the respective concepts.
- In case the campaign is not based on a model: measure at least attitude, subjective and descriptive norms, perceived behavioural control and behavioural intentions.

How to measure?

- Always measure before and after the campaign.
- In case the campaign is regional or local wide: add a control/comparison group.

How to analyse your evaluation data?

- Describe the target group in terms of the relevant demographic variables (age, gender, education, etc.) and/or other background information (e.g., car driving frequency).
- Provide relevant descriptive statistics.
- Provide basic inferential statistics (significance testing and effect sizes) - explain if the difference between the two measurements is statistically significant (is it possible to generalise?) and what is the size of the effect.

When these steps of the evaluation study are finalised, the findings of your campaign evaluation should be reported and disseminated. To avoid missing information in the evaluation report in the future, CAST provides guidelines for reporting the campaign and its effects in a standardised way (see Boulanger et al, 2009).

Bibliography

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.

Ajzen, I. (2002). *Constructing a TPB questionnaire: conceptual and methodological considerations*. Retrieved 9 November 2006 from <http://www.people.umass.edu/aizen/pdf/tpb.measurement.pdf>.

Bauman, A., B. J. Smith, et al. (2006). Evaluation of mass media campaigns for physical activity. *Evaluation and Program Planning*, 29, 312-322.

Boulanger, A., Daniels, S., Delhomme, P., Deugnier, M., Divjak, M., Eyssartier, C., Hels, T., Moan, I., Nathanail, E., Orozova-Bekkevold, I., Ranucci, M., Schepers, P., Van den Bossche, F., Zabukovec, V. (2007a). *Comparison of research designs*. CAST project, Brussels: Belgian Road Safety Institute

Boulanger, A., Divjak, M., Orozova-Bekkevold, I. & Zabukovec, V. (2007b). *Typology of evaluation methods: current practices and campaign evaluation* (A. Boulanger & M. Divjak, Eds.). CAST project, Brussels: Belgian Road Safety Institute.

Boulanger, A., Hels, T., Larsen, L., Meng, A., Orozova-Bekkevold, I. (2009). *Reporting tool for effects of a road safety campaign*. CAST project. Brussels: Belgian Road Safety Institute.

Brault, M. & Letendre, P. (2003). *Evolution des comportements et attitudes face à la vitesse au Québec entre 1991 et 2002*, Québec: SAAQ.

Cameron, M. and W. A. Harrison (1998). *Developing an effective public education strategy for road safety based on successful experience in Australasia*. Paper presented at the 9th International Conference Road Safety in Europe, Germany.

Campbell, D. & Stanley, J. (1963). *Experimental and Quasi-experimental Designs for Research*. Boston: Houghton Mifflin Company.

Cauzard (2004). European car drivers and road risk. In *SARTRE 3 reports Part I: Report on principal results*, Arcueil: INRETS.

CDC (2001). *Program Evaluation. Program Operations Guidelines for STD Prevention*, Atlanta: Centers for Disease Control and Prevention.

Christchurch City Council (2004). *Evaluation Guide*. Christchurch: Research and Policy Unit.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory on normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58 (6), 1015-1026.

Conner, M., & Armitage, C. J. (1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of Applied Social Psychology, 28*, 1429-1464.

Conner, M., Smith, N., & McMillan, B. (2003). Examining Normative Pressure in the Theory of Planned Behaviour: Impact of Gender and Passengers on Intentions to Break the Speed Limit. *Current Psychology: Developmental, Learning, Personality, Social, 22* (3), 252-263.

Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Boston: Houghton Mifflin.

Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

De Dobbeleer, W. (ed.), Nathanail, T., Adamos, G. (2009). *Pan-European Road Safety Campaign: final report*. Brussels: Belgian Road Safety Institute. (to be published).

Delaney, A., Lough, B., Whelan, M., & Cameron, M. (2004). *A review of mass media campaigns in road safety*. Accident Research Centre Report No. 220, Victoria, Australia: Monash University Accident Research Centre.

Delhomme, P., Vaa, T., Meyer, T., Harland, G., Goldenbeld, C., Järmärk, S., Christie, N., Rehnova, V. (1999). *Evaluated road safety media campaigns: An overview of 265 evaluated campaigns and some meta-analysis on accidents*. EU GADGET project. Arcueil: INRETS/RR-00-006-FR.

Delhomme, P., De Dobbeleer, W., Forward, S., Simões, A., Adamos, G., Areal, A., Chappé, J., Eyssartier, C., Loukopoulos, P., Nathanail, T., Nordbakke, S., Peters, H., Phillips, R., Pinto, M., Ranucci, M. -F., Sardi, G. M., Trigosso, J., Vaa, T., Veisten, K., Walter, E. (2009). *Manual for designing, implementing and evaluating road safety communication campaigns*. Brussels: Belgian Road Safety Institute.

Drinking and driving: a road safety manual for decisions-makers and practitioners. (2007). Geneva: Global Road Safety Partnership.

Elder, R. W., R. A. Shults, et al. (2004). Effectiveness of Mass Media Campaigns for Reducing Drinking and Driving and Alcohol-Involved Crashes - A Systematic Review. *American Journal of Preventive Medicine 27*(1).

Elliott, B. (1989). *Effective road safety campaigns: A practical handbook*. Report prepared for Federal Office of Road Safety. Sydney: Elliott & Shanahan Research.

Elliott, M. A., Armitage, C. J., & Baughan, C. J. (2003). Drivers' compliance with speed limits: An application of the theory of planned behavior. *Journal of Applied Psychology, 88* (5), 964-972.

Evaluation budgeting (n.d.) Retrieved 1 August 2007 from http://www.opportunitystudies.org/repository/File/evaluation/evaluation_budgeting.pdf.

Frechtling, J., & Sharp, L. (1997). *User-friendly handbook for mixed method evaluation, Directorate for education and human resources*. Retrieved September 29, 2006, from <http://www.ehr.nsf.gov/EHR/REC/pubs/NSF97-153/START.HTM>

Hankins, M., French, D., & Horne, R. (2000). Statistical guidelines for studies on the theory of reasoned action and the theory of planned behaviour. *Psychology and Health, 15*, 151-161.

Harré, N. & Wrapson, W. (2004). The evaluation of a central-city pedestrian safety campaign. *Transport Research Part F*, 167-179.

Haskins, J. & Kendrick, A. (1993). *Successful advertising research methods*. Lincolnwood: NTC business books.

Herzog, T. (1996). *Research methods in the social sciences*. New York: Harper-Collins College Publishers.

Joint Committee on Standards for Educational Evaluation (1994). *The program evaluation standards, 2nd edition: How to assess Evaluations of Educational Programs*. Thousand Oaks, CA: Sage.

Kalton, G. (1983). *Introduction to survey sampling*. University of Michigan, California: SAGE Publications.

Logan, T.K., Padgett, D.K., Thyer, B.A. & Royse, D. (2006). *Program evaluation: an introduction*. Belmont, California: Thomson Brooks/Cole.

MacDonald, G., G. Starr, et al. (2001). *Introduction to Program Evaluation for Comprehensive Tobacco Control Programs*. Atlanta (GA): Centers for Disease Control and Prevention.

Manstead, A. S. R. (2000). The Role of Moral Norm in the Attitude-Behavior Relation. In D. J. Terry & M. A. Hogg (Eds.), *Attitudes, Behavior and Social Context*. New Jersey: Lawrence Erlbaum.

Mercer, G.W., Cooper, P.J. & Kristiansen, L. (1996). A cost/benefit analysis of a 5-month intensive alcohol-impaired driving road check campaign. In *40th Annual Proceedings, 7-9 October 1996*. Vancouver: Association for the advancement of automobile medicine.

Millett, R. A. (1998). *Evaluation Handbook*, Michigan: W.K.Kellogg Foundation.

Mruk, Ltd. (2004). *Evaluation of the 2003-2004 Festive drink drive campaign*. London: Scottish Executive Social Research.

Neuman, W. L. (1997). *Social research methods: qualitative and quantitative approaches*, 3rd edition. Massachusetts: Allyn and Bacon.

OECD. (2004). *Stakeholder involvement techniques*. Paris: Organisation for Economic Co-operation and Development.

Parker, D., Manstead, A. S. R., & Stradling, S. G. (1995). Extending the theory of planned behaviour: The role of personal norm. *British Journal of Social Psychology*, 34, 127-137.

Patton, M.Q. (2002). *Qualitative evaluation and research methods*, 3rd edition. Newbury Park, CA: Sage.

Peck, R., Olsen, C. & Devore, J. (2005). *Introduction to statistics and Data Analysis*. 2nd edition Belmont, California: Brooks / Cole – Thomson Learning.

Prigogine, Y., L'Hoste, J., Roche, M., Sheppard, D. & Wind, G.H. (1975). *Manual on road safety campaigns: guidelines for preparation, design and effectiveness control*. Paris: OECD.

Prigogine, J. (2003). *Zorg dat BOB rijdt. Altijd. - Evaluatie*. Brussels: Belgian Road Safety Institute.

Robson, L. S., Shannon, H. S., Goldenhar, L. M., & Hale, A. R. (2001). *Guide to Evaluating the Effectiveness of Strategies for Preventing Work Injuries: How to Show Whether a Safety Intervention Really Works*. Washington: Department of Health and Human Services, Public Health Service, Centre for Disease Control and Prevention, National Institute for Occupational Safety and Health.

Rojas, L. and F. Serpa (2006). *Introduction to Program Evaluation*, Professional Research and Evaluation Group.

Rudmin, F. W. (1999). Norwegian short-form of the Marlowe-Crowne Social Desirability Scale. *Scandinavian Journal of Psychology*, 40, 229-233.

Scheaffer, R.L., Mendenhall, W. & Ott, R.L (2006). *Elementary survey sampling*. 6th edition. Belmont, CA: Thomson Brook/Cole.

Schopper, D., J.-D. Lormand, et al., Eds. (2006). *Developing policies to prevent injuries and violence: guidelines for policy-makers and planners*. Geneva: World Health Organization.

Shaw, I.F., Greene, J.C., & Mark, M.M. (2006). *Handbook of evaluation: policies, programs and practices*. London: SAGE Publications.

Simmonds, A.G. (1981). The effects on vehicle speeds of roadside safety posters. *Traffic Engineering & Control*, 22(8/9), 480-485.

Solomon, M., Ulmer, R. & Preusser, D. (2002). *Evaluation of the Click it or Ticket Model Programs*. Trumbull, CT: Preusser Research Group Inc.

Sparks, P. (1994) Attitudes towards food: applying, assessing and extending the theory of planned behaviour. In Rutter, D. R. and Quine, L. (Eds.), *Social Psychology and Health: European Perspectives* (pp. 25–46). Avebury: Aldershot.

Studsholt, P. (1990). Campaigns against Drunken Driving among Young Drivers. In *Proceedings (23-33) of Road Safety and Traffic Environment in Europe, Gothenburg, Sweden, 26-28 September 1990*. Linköping: Statens väg- och trafikinstitut (VTI).

Sutton, S. (2002). Using Social Cognition Models to Develop Health Behavior Interventions: Problems and Assumptions. In D. Rutter & L. Quine (Eds.), *Changing Health Behaviour: Intervention and Research with Social Cognition Models*. (pp. 193-208). Buckingham: Open University Press.

SWOV (2007), Stichting Wetenschappelijk Onderzoek Verkeersveiligheid. Retrieved 14 March, 2007, from <http://www.swov.nl>.

U.S. Department of Health and Human Services (2005). *Introduction to program evaluation for public health programs: a self-study guide*. Atlanta, GA: Centers for Disease Control and Prevention. Office of the Director, Office of Strategy and Innovation.

Vaa, T. & Phillips, R. (eds), Adamos, G., Areal, A., Ausserer, K., Delhomme, P., Divjak, M., De Dobbeleer, W., Forward, S., Krol, B., Meng, A., Moan, I., Nathanail, E., Pohlmeier, E., Sardi, G.M., Schepers, P., Sedá, E., Ulleberg, P., Walter, E. (2009). *Effects of Road Safety Campaigns*. Brussels: Belgian Road Safety Institute.

Verplanken, B., & Orbell, S. (2003). Reflections on past behaviour: A self-report index of habit strength. *Journal of Applied Social Psychology*, 33, 1313-1330.

W.K.Kellogg Foundation (2004). *Logic Model Development Guide*, Michigan: W.K.Kellogg Foundation.

WHO (2006). *Helmets: a road safety manual for decision-makers and practitioners*. Geneva: World Health Organization.

Wholey, J.S. (2004). *Handbook of practical program evaluation*. San Francisco: Jossey-Bass.

Wimmer, R.D. & Dominick, J.R. (1994). *Mass media research: an introduction*, 4th edition. Belmont, CA: Wadsworth Publishing Company.

Annex 1: Elements of a detailed description of a road safety campaign: summary

Before every campaign evaluation, different stakeholders of the campaign should be engaged. Why, how and who stakeholders are will be discussed in the first section. The next step of the campaign evaluation process is to describe the road safety campaign in terms of its various components. Sections 2 until 8 discuss the different parts of a campaign description (U.S. Department of Health and Human Services 2005).

1. Engaged stakeholders

Stakeholders in a road safety campaign evaluation are people or organisations that invested in the campaign, are interested in the results of the evaluation and/or have a stake in what will be done with the results (U.S. Department of Health and Human Services 2005). Because of their presence, in one way or another, in the road safety domain, it is good practice to include their needs and interests in the campaign evaluation. Stakeholders can be involved in the campaign operations (e.g. government, police officers), they can be affected by the campaign (e.g. road users), or they can be the users of the evaluation findings (e.g. research institutes, government).

Who are the stakeholders?

Stakeholders are, in one way or another, related to the road safety campaign. For topics of common interest, like road safety, the list of possible stakeholders can be quite large, as will become clear from the examples given in Table 1. Even the community as a whole can be considered as an interested party, and their involvement is important. In particular, special effort should be made to ensure the involvement of relevant community groups (Schopper, Lormand et al. 2006), which are the groups in the community that in some way or another are related to the campaign. The choice of which stakeholders to involve is a function of the purpose of the evaluation and who will use the results (CDC 2001).

In several documents (for example Christchurch City Council 2004; U.S. Department of Health and Human Services 2005), lists of criteria are provided that can be used to classify stakeholders according to their specific importance in the evaluation process. The classification of stakeholders is important, as it may determine the way in which stakeholders should be involved in the road safety campaign. The following classes are considered:

1. Stakeholders that are involved in setting up and conducting the campaign.
2. Stakeholders that invest in the initiative.
3. Stakeholders that are served or affected by the initiative (the target groups).
4. Stakeholders who are willing to carry out similar campaigns in the future.
5. Stakeholders who are intended users of the evaluation findings.
6. Opponents of the initiative.

For example, when the government launches a campaign for helmet use among motorcyclists, then this government is a stakeholder in the groups 1 and 2. The motorcyclists themselves are expected to be affected by the campaign (group 3), while pressure groups and research institutions might be interested in the results of the evaluation (group 5). Regional road safety organisations might be interested in the campaign because of plans for organising similar campaigns in the future (group 4). Helmet use might also have opponents, who do not believe in the value of this risk-decreasing device, and therefore will presumably not support the campaign (group 6). If they are not willing to be involved in the campaign evaluation at the start, it is good practice to keep them informed about the progress of the evaluation, in the hope they may ultimately be persuaded to join at a later stage.

Because of the potentially large list of stakeholders that might be generated, it is crucial to identify those that matter most (U.S. Department of Health and Human Services 2005).

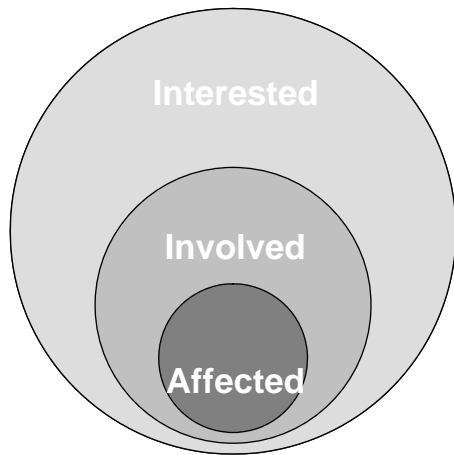
How to select stakeholders?

In this section, some examples of possible stakeholders in the context of road safety campaigns are given. Identifying and selecting stakeholders is an ongoing and iterative process, based on the growing understanding of stakeholders' identity, importance and objectives. However, the list of stakeholders should be derived *tailor-made*, because campaigns can differ considerably. The following steps can be used as a guideline to obtain a useful selection of stakeholders:

1. List all possible stakeholders connected to or involved in the project.
2. Categorise the stakeholders as affected, involved, and interested.
3. Explore the characteristics of each group (social background, formal or informal status, structure, current situation related to the campaign...).
4. Identify interests, motives and attitudes in each group.
5. Illustrate the relationships (in terms of the evaluated campaign) among the affected, involved and interested stakeholders. The diagram in Figure 1 can be helpful in this respect. Relationships can be friendly, neutral or hostile. Stakeholders can have comparable or conflicting interests concerning the campaign.
6. Evaluate the potential of each group. What are the strengths and the weaknesses?
7. What would be the implications for the campaign process if certain stakeholders are left out? If a group is left out, would they be informed about the progress, or ignored completely?
8. Based on this information, choose the final stakeholders that will be included in the list.

Note that the selection of stakeholders is ideally done at the start of the campaign, although some stakeholders can have a different role in the evaluation compared to the design phase. Reconsidering the stakeholders list and their specific roles at the start of the campaign is therefore recommended.

Figure 1: Interested, involved and affected stakeholders



Based on the gathered information about stakeholders, Table 1 can be constructed:

Table 1: Stakeholder identification table

Potential stakeholder	Role	Characteristics	Problems	Interests	Potential	Linkages	Priority
Whose interest do you want to prioritize?	Is the stakeholder affected by / involved in / interested in the campaign and its evaluation?	What are the specific characteristics of the stakeholder?	What problems (related to the campaign) does the stakeholder face?	What does the stakeholder want from the campaign evaluation?	What might the stakeholder contribute to the campaign evaluation?	Are there any links of conflict / cooperation / dependency?	What is the importance of the stakeholder compared to others?
Campaign organisers							
Funding agencies (local or state governments)							
Schools and educational organisations							
Road safety research institutions							
Private citizens (or road users)							
Campaign critics							
Pressure groups							
Law enforcement representatives (police forces)							
...							

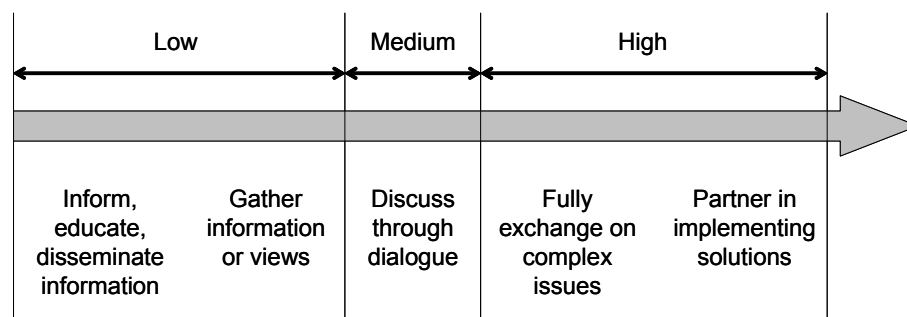
How should stakeholders be involved?

Stakeholder involvement will vary with the type of evaluation and can be realised at various levels. For example, sometimes it might be helpful to involve stakeholders in the question development, data collection and analysis, while in other cases it might be sufficient to ask their advice when developing questions and/or to communicate the main evaluation findings. A complete involvement of stakeholders in the evaluation is called a participatory approach to evaluation. In any case, it is

advisable to ensure that stakeholders have the opportunity to provide input to designing evaluation questions and are kept informed of the progress and the results of the evaluation.

Roughly speaking, the way stakeholders are involved is related to the nature of the participation (advisor, collaborator, etc., mainly based on whether a stakeholder is affected, involved or interested), the form of the participation (active member, observer,...) and the participation mode (as a group, a company, an individual,...) (WHO 2006). It should be clear by now that some stakeholders might beforehand expect a particular level of involvement. Therefore, knowing the stakeholders and preliminary contacts are very important in order to position them at the appropriate level of involvement. The intensity of involvement of a stakeholder can be represented as a continuum, as the one presented in Figure 2. Every stakeholder on the list should be positioned on this line, such that the involvement and the mutual expectations are clear right from the start.

Figure 2: Stakeholder involvement continuum (based on OECD)



For some stakeholders, a low level of involvement is sufficient. Informing, educating and disseminating are needed when the campaign, its evaluation and the corresponding results should be described for the stakeholder. A slightly higher level of involvement consists of listening to and gathering information from the stakeholders. When two-way information exchange is needed, the involvement is labelled as medium. These stakeholders have an interest in the issue and will likely be affected by the campaign and its evaluation. This type of involvement is especially needed when there is an opportunity of influencing the final outcome, for example the number of road victims in the targeted group of road users. When stakeholders are expected to talk to each other about the subject under study, hoping that this interaction will create an opportunity to affect related policies, then the level of full exchange on complex issues is required. Once stakeholders are involved in managing the campaign and develop solutions for the problems at hand, they are at the highest level of involvement.

As an example, consider a road safety campaign for impaired drivers. As this is an issue of safety and health, often related to the implementation of regulations and legislations, it is necessary to involve the public in one way or another. The public can be seen as a stakeholder who requires the necessary information in order to create support for the topic. Especially the target groups will be educated and informed about the campaigns and its effects. Governments and scientific institutions may be the right partners to provide information about the topic that is evaluated. Pressure groups should have a higher level of involvement, as they may influence the final outcome if attitudes are changed through the campaign. Also producers of breathalysers are at the level of medium involvement, because knowledge about the

measurement devices is crucial when collecting data. Policy makers who are in a position to formalise specific evaluation outcomes in regulations, and are able to decide on future campaigns, should be at a level of full exchange. Also those who are responsible for data analysis are at a high level of involvement. Finally, every institute or organisation that is involved in the evaluation process or can implement solutions to the problem should be considered as partner in the campaign process.

Checklist

In a nutshell, the following checklist summarises the main objectives of a stakeholder analysis (adapted from Rojas and Serpa 2006; WHO 2006):

1. Identify key stakeholders, define their characteristics and examine how they will be affected by the campaign. Review the list of stakeholders to ensure all appropriate stakeholders are included.
2. Assess their potential influence on the development, approval and implementation of the campaign. This will help identifying stakeholders' roles in the evaluation.
3. Understand the relationship between stakeholders and possible conflicts of interest that may arise.
4. Assess the capacity of different stakeholders to participate in the campaign and the likelihood of their contribution to the process. Knowing their areas of expertise and interest will optimise their input in the campaign evaluation. For example, it is not unwise to ask stakeholders to suggest evaluation questions that are in line with their own interests and experiences.
5. Decide how they should be optimally involved in the campaign.
6. Understand and respect stakeholders' values.

2. Need for the campaign

The description of the need for the campaign should explain the (road safety) problem addressed by the campaign. For example, there might be increased attention for the level of impaired driving among young road users during the weekend. It should include an analysis of the magnitude of the problem, and present related road safety figures (number of victims, accidents, etc.), for various stratifications if needed (according to gender, rural versus urban areas, etc.). The identification of the need requires more than just accident statistics showing a low level of road safety for particular road users. Also the rationale for why a campaign may be beneficial in relation to changing behaviour and raising awareness should be made clear, as well as the availability of the support for the campaign over a defined time frame (Bauman, Smith et al. 2006). Important questions that should be answered at this stage are:

- What is the road safety problem and what are the consequences for the state or community?
- What is the size of the problem, overall and in various segments (roads, road users...)?

- What are the determinants of the road safety problem? To what extent is it related to the infrastructure, the road user or the transport mode? Is it determined by road users with specific characteristics (age, social group...) or by the geographical properties (urban versus rural), etcetera.
- What changes or trends are occurring?

3. Target groups

The target groups of the campaign are the groups or audiences that need to move into action or need to change behaviour in order to make progress on the road safety problem. Road safety campaigns typically aim to change the behaviour in one or more target groups, and to create an environment that reinforces sustained adoption of these changes, with the intention to prevent accidents and injuries in the specific target group (U.S. Department of Health and Human Services 2005). Road users in the first place, but also governmental services, police forces or other parties involved, like parents of young road users, may be considered as campaign targets. Therefore, it can be interesting to distinguish between direct and indirect targets. The identification of the target group is not only important for the development of the campaign itself, but also for evaluation purposes, as the campaign should be evaluated in the light of its effects realised on the target group.

It is important to determine the target groups based on the nature and the definition of the behaviour that will be focused on by the campaign. Identification of the target behaviour leads to identification of the characteristics and circumstances most frequently associated with the behaviour, and hence, identification of the target group (Cameron and Harrison 1998; cited in Delaney, Lough et al. 2004). Depending on the characteristics of the target group, another campaign setting might be optimal. For example, a campaign to stimulate the use of protection systems for children in cars might need different communication channels compared to a campaign to prevent impaired driving. Also, the message content needs to be realistic and credible, portraying members of the target group (Delaney, Lough et al. 2004). This asks for careful identification of the target groups.

The main question when determining the target population is: who are you going to serve (Millett 1998). This might seem a very general question, but it is important here to be as specific as possible. Also, it might be necessary to define several subgroups within a target population. For example, when setting up a campaign to raise helmet use among young bicyclists, also the parents of these children might be a target. Therefore, it makes sense to distinguish and name them clearly.

4. Description of the campaign results (objectives)

The result of a campaign can be seen as its contribution to the reduction of the level of road accidents and victims. However, also less ambitious results can be related to the fact that a campaign has been implemented. Changes in attitude or awareness-raising are favourable results of a campaign, and should be evaluated as well. As will

be seen in Figure 3, a hierarchy of effects can be considered, each of which will finally lead to a change in the level of road safety.

The results of a campaign can be considered from two points of view. An evaluator considers the *outcomes* of a campaign as the main study topic, while a planner focuses on *objectives and goals*. The latter view is from an evaluation designer, who wants to set objectives for the long, intermediate and short term. The former is from an evaluator, who can measure the outcomes of a campaign that has been implemented. In the end, these two views match perfectly with each other. They will both be introduced shortly.

An evaluator's view: outcomes

The outcomes are the changes in attitude or behaviour of the targets that will hopefully result from the campaign. They specify how and in what way the targets need to change. Typically, the ultimate outcome of a road safety campaign is quite ambitious and usually has a long-term character (i.e. reducing the number of accidents and victims that are related to the problem highlighted by the campaign). Therefore, not only long-term outcomes but also short-term and intermediate outcomes should be formulated in a campaign description. Outcomes can be very diverse: intensity of contact with the campaign, interest for, acceptance of and reaction to the campaign, learning about different road user behaviours, adopting actions, changes in the environmental conditions or changes in the level of road safety. It is clear that outcomes can be defined for the whole process of the campaign, and evaluations will be stronger when they deal with both short-term and long-term outcomes.

In various sources, the outcome variables of a campaign are therefore organised in a theoretical hierarchy of effects (see e.g. U.S. Department of Health and Human Services 2005; Bauman, Smith et al. 2006). The concept of a hierarchy of effects is also clearly applicable to road safety campaigns.

- *Participation*: A campaign initiates in the first place some participation among the target group. This is expressed in terms of the number of people reached, their characteristics, and so on. In terms of campaigns, one may also add awareness of the campaign, dose of exposure and source of exposure.
- *Reactions*: In the next step, the target group may show some degree of interest for and feelings towards the message communicated by the campaign. They accept the activities and recognise a perceived personal relevance of the message.
- *Learning*: In the learning step, the target's knowledge and opinions are affected, as they know the message of the campaign and understand the theme and the recommendation.
- *Actions*: Following the learning process, the target audiences will take appropriate actions, that is, they adopt their behavioural patterns towards the conceived message. This is the stage where the target group reflects the message to its own situation, and is considering (or is at least showing the intention of) changing behaviour and attitudes.
- *System and environment change*: The set of actions undertaken will subsequently initiate a social support for the campaign and its message. This is

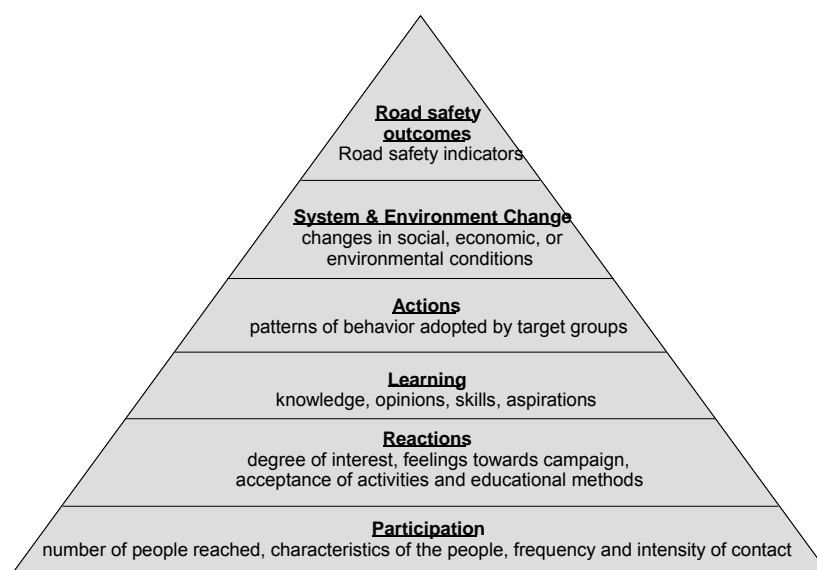
the result of the complete set of recommendations, actions, policies and practices that have been implemented.

- *Road safety outcomes*: Once the social support is present and the shift in norms concerning the targeted road safety problem has been made, the campaign can finally result in a long-term effect on the road safety outcomes, in terms of number of road users killed or the number of accidents.

It is clear that the higher levels in the hierarchy are more difficult to achieve and less straightforward to measure, although they are the genuine reasons for setting up the campaign. Therefore, an evaluation should show evidence at various levels of the hierarchy.

Although the effects lower in the hierarchy are needed to obtain effects on a higher level, sometimes a distinction in importance is made along the lines of the hierarchy. For example, Elder et al. (2004) investigate the effectiveness of mass media campaigns for reducing drinking and driving and alcohol-involved crashes. The selected outcome measures are crashes and the measured blood alcohol concentration. These are clearly related to the ultimate outcome of interest, namely crash related injuries. The intermediate outcomes, such as message recall or knowledge and attitude change may, according to the authors, be only weakly related to actual behaviour and will not measure the effectiveness of the campaign. However, it is reasonable to assume that the final effect of a campaign will be a combination of several factors that can have an impact on human behaviour. Also, intermediate outcome measures are important to find improvements in case that the goals are not reached. These outcomes will allow distinguishing between a campaign that was not set up properly, and a well designed campaign that was indeed not effective.

Figure 3: A hierarchy of effects (based on U.S. Department of Health and Human Services 2005)



A planner's view: objectives and goals

While the terminology of outcomes is common in evaluation studies, there might be confusion with the terms *objectives and goals*, which are typically used in the planning of a campaign (U.S. Department of Health and Human Services 2005). A goal in planning will correspond to the long-term outcomes in the evaluation, while objectives in planning are comparable to short-term and mid-term outcomes in evaluation. Objectives correspond to the amount of change expected from the campaign for a given road safety problem, for a specified target group within a given time frame. These objectives are divided into short-term, intermediate and long-term objectives. In the short term, changes in attitudes, behaviour, knowledge and skills are expected to result from the campaign activities, usually at the level of individuals in the target group (W.K.Kellogg Foundation 2004). Intermediate objectives build on the short-term objectives. For example, reading information messages on the BOB campaign (in the short-term) leads to increased knowledge about the campaign (as an intermediate outcome). The long-term objectives involve organizational, community or system level change, expected to result from the campaign activities. This might include an improved level of road safety, with a decreasing amount of accidents and persons killed. BOX 1 gives an example on how an objective is usually stated.

BOX 1: Elements of an objective

Who will achieve how much of which outcome by when?

Who?	The target group
How much?	A percentage or target amount
Which outcome?	The measured concept
By when?	Period after the campaign begins

Source: MacDonald, Starr et al, 2001

Although the need for the campaign may be formulated on a rather abstract and ambitious level, the goals and objectives of the campaign are restricted by the available resources and the maturity of the campaign. For example, a first edition of a bicycle helmet campaign might have objectives that are different from those of a campaign that has been recurring for 10 years in the same region. Also the resources (in terms of funding, human capital, time...) will determine the outcomes that can be expected from the campaign.

A clear definition of the goals and objectives will help in identifying the targets by which the campaign's effectiveness will be measured (MacDonald, Starr et al. 2001). The final goal of a campaign is its overall purpose. Every road safety campaign wants to contribute to the reduction in the number of persons killed and injured in traffic. Objectives describe the results to be achieved. They should be specific and measurable, as they will help in monitoring progress towards achieving goals. In many sources, objectives are considered well-defined if they are SMART (MacDonald, Starr et al. 2001): Specific, Measurable, Achievable and Ambitious, Relevant and Time-bound (see BOX 2).

BOX 2: SMART objectives.

S	Specific	It identifies a specific event or action that will take place.
----------	-----------------	--

M	Measurable	It quantifies the amount of change to be achieved.
A	Achievable Ambitious	It is realistic given available resources, yet challenging enough to accelerate campaign efforts.
R	Relevant	It is logical and relates to the campaign's goals.
T	Time-bound	It specifies a time by which the objective will be achieved.

For example, in a seatbelt wearing campaign, an objective might be to increase the percentage of young adults (18 – 25 years old) who report they wear a seat belt all the time from 80% in 2005 to 95% in 2010. This is a *specific* objective, because it identifies the event of increasing the reporting percentage of seat belt wearing in the target group. It is *measurable*, since it starts from a given percentage in 2005 (80%, assuming an existing data source in this example) and indicates the quantity that the intervention should achieve (95%). It is *achievable*, for example based on historical growth rates of the percentage, yet *ambitious*, because achieving this objective would be a significant improvement. It is *relevant*, as it directly relates to the social support for a higher seat belt wearing rate. Finally, it is *time-bound*, given the specified time by which the objective should be achieved.

5. Activities

The activities are the actions that will be undertaken to move the target groups to a certain level of change or to specific actions, in order to achieve the desired outcomes in the target groups. Activities may include products (like promotional material, posters, gadgets, and television campaigns), services (training, education in schools) and infrastructure (buildings, promotional cars). In order to get a clear description of the campaign and what should be evaluated, it is important to describe the different activities, determine how they relate to each other and to the goal of the campaign, and identify the different actions expected to occur (MacDonald, Starr et al. 2001).

Activities are clearly related to the outcomes that have to be achieved. Therefore, it is good practice to list all activities and the corresponding outcomes in a table, like the example shown below (U.S. Department of Health and Human Services 2005).

Example: Inventarisation of activities and related outcomes

Activities

What will the campaign and its staff actually do?

...

Outcomes

What are the changes that hopefully will result from the campaign activities?

...

6. Outputs

Outputs are the tangible deliverables that will be produced by the campaign activities (the number of posters in the streets, the number of cars controlled, etc.). These typically reflect the activities that have been redefined in countable terms, like the

number of posters visible in the streets, the number of cars controlled by police forces, and so forth.

Defining outputs as measurable representatives of the campaign activities is important in the light of an evaluation. If one is interested in the magnitude of the campaign, related to the outcomes obtained, then the outputs will quantify all activities that were performed. If the expected results of a campaign fail to appear, this might be related to the level of activities performed and should therefore be documented in the evaluation.

7. Inputs

The inputs are the funding sources and other resources (people, information, etc.) that are needed to ensure a successful and effective implementation of the activities. These are usually the factors on which the success or failure of the campaign depends, and if the intended outcomes have not been achieved, one possible issue in the evaluation might be the lack of implemented resources. For road safety campaigns, legal issues should often be dealt with in order to enable enforcement actions, police officers should have access to training programmes, resources (both funding and personnel) should be available to develop the campaign, etcetera.

It is important for various reasons to take inputs into account when describing the campaign for evaluation purposes. First, the evaluation often includes the accountability for resources to funding partners (that is: is the money well spent?). If intended outcomes are not achieved, looking back at the funding resources is a first step. Either the resources were not managed well, or they were insufficient to perform the required activities. Second, inputs are often fixed before a campaign is actually started. Therefore, activities and corresponding resources (budget, people...) should be considered together in order to end up with a campaign design that is realistic. The costs of an outcome evaluation should be considered very early in the process, preferably together with the budgeting of the campaign itself. Looking at the outcomes in relation to the inputs also forces the campaign planners to come up with effective activities as much as possible, such that the available resources are invested optimally.

8. Stage of development and context of the campaign

A campaign can roughly be in three stages of development: planning, implementation and outcomes achievement. The stage of development is important to set a realistic evaluation focus. Indeed, a campaign that has just been planned will be evaluated differently from a campaign that has been repeated for several years, mainly because the evaluation data that is available and the experience that has been built up will be completely different.

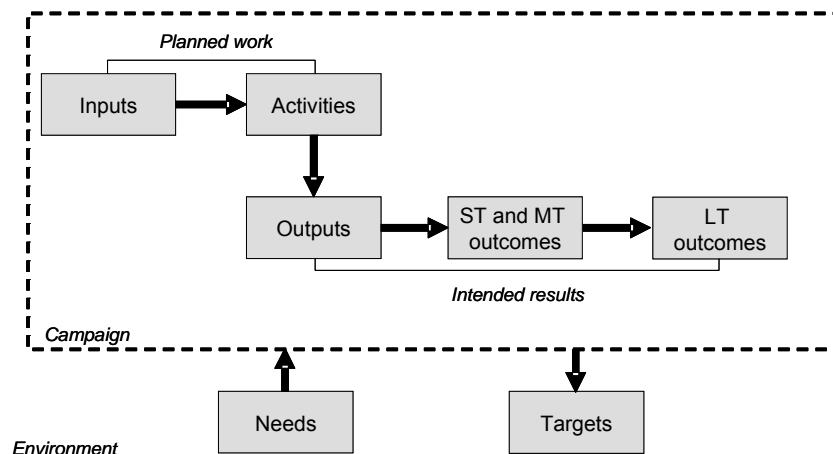
A campaign description should also include the possible opportunities and hindering factors in the larger environment. The environment can contain both threats and opportunities for the campaign. Environmental issues include politics (e.g. increased attention for the road safety problem on the political agenda), competing interests (for example when campaigns try to influence the activities of trucks in the transportation sector), funding (a partner that might be forced to stop funding the

campaign), social and economic conditions, special concurring events (for example when a campaign for young road users is launched during a youth music festival, the campaign should take advantage of the increased exposure to the campaign activities), and so on. Knowing the context will also contribute to a more realistic interpretation of evaluation findings, and hence to a better assessment of their generalisability (U.S. Department of Health and Human Services 2005).

Representing relationships: a logic model

Once these components are identified, the relationships among them should be clarified. A logic (conceptual) model is a graphic representation of the relationships between the activities and the intended outcomes of a campaign. Logic models are a common tool employed by evaluators (U.S. Department of Health and Human Services 2005).

Figure 4: A logic model for road safety campaigns (adapted and extended from U.S. Department of Health and Human Services 2005)



A logic model is a systematic and visual way to represent the relationships among the resources that are available to set up the campaign, the activities that will be performed with these resources and the changes or results that are hoped to be achieved (W.K.Kellogg Foundation 2004). A basic logic model is represented in Figure 4. This scheme shows that a campaign is initiated by a certain need and oriented towards a certain target. It converts the planned work (inputs and activities) into intended results (outputs and outcomes) and is further embedded in the larger environment in which the campaign will be developed. It is clear that all aspects discussed above are depicted in this logic model. Also, the logic model depicts, in a way, a timeline from planning to results and thereby introduces conditional relations between the building blocks. For example, the activities will only be performed given the resources; outputs can only be delivered if activities are accomplished, and so on.

There are several reasons why presenting a campaign in a logic model is fruitful (W.K.Kellogg Foundation 2004). First, logic models can better position campaigns for success. This is not only true for campaign design, planning and implementation, but also for evaluation and reporting purposes. A logic model presents campaign information and progress towards formulated outcomes in an informative and clearly

structured way, which makes it useful to present to campaign stakeholders. Second, a logic model may increase the interest of investors in the campaign. It is a clear representation of the ideas behind the campaign and of the inputs it needs to be operated, and it allows capturing, documenting and disseminating the results. If investors realize that the campaign is not an impulsive project but a well-structured and outcome-oriented plan, then they will be more inclined to invest in the campaign. Third, when a problem is investigated in a structured way, it will be more flexible to adapt when the campaign and/or the environment change over time.

In Table 2, the elements of a logic model implementation are given. As an example, activities and outcomes from the BOB campaign are shown (Prigogine 2003). Of course, depending on the stage of the campaign and the objectives of the evaluation, a logic model can be more or less elaborate.

Table 2: Logic model development: implementation

Resources	Activities	Outputs	ST and MT outcomes	LT outcomes
Resources needed to perform activities	Activities to conduct to solve the problem	What is expected to result from the activities	When outputs are delivered, this is expected in 1-3 or 4-6 years	When outputs are delivered, this is expected in 7-10 years
<ul style="list-style-type: none"> ▪ Funding ▪ Experts ▪ Material resources ▪ ... 	<ul style="list-style-type: none"> ▪ Posters ▪ Television spots ▪ Television programs ▪ Radio spots ▪ Website ▪ Article in road safety magazine ▪ Messages along the road ▪ Gadgets for restaurants and pubs ▪ Stickers ▪ Bob-van ▪ Public transport activities ▪ Police controls ▪ ... 	Quantify the activities: <ul style="list-style-type: none"> ▪ Number of posters ▪ Minutes of television time ▪ Website hits ▪ Number of persons who used public transport on New Year's Eve ▪ Number of stickers distributed ▪ ... 	<ul style="list-style-type: none"> ▪ Familiarity with the BOB campaign ▪ Knowledge of the messages ▪ Recall of the campaign ▪ ... 	<ul style="list-style-type: none"> ▪ Change in attitude (in %) ▪ Change in behaviour (in %) ▪ Change in impaired driving (in %) ▪ Change in impaired driving accidents (%) ▪ Change in the number of accidents (%) ▪ ...

Setting up a logic model

Setting up a logic model involves several steps (W.K.Kellogg Foundation, 2004; U.S. Department of Health and Human Services, 2005). Following these steps will finally result in a complete filled-in Table 2 which will subsequently be used to design the different steps according to Figure 4.

1. Develop a list of activities and intended outcomes.

This is at the core of the logic model, as it brings clarity in the relationship between activities and outcomes. Many existing documents on the campaign (strategic plans, key informants...) might contain information that meets the definition of an activity or an outcome. Another approach to develop this list is to work backward from the outcome (how to obtain a given outcome?), and trying to discover the factors, variables and actors that will be involved in an expected change (U.S. Department of Health and Human Services, 2005). Working forward from the activities (what happens when an activity is performed?) to discover their added value for the campaign can also be helpful to find out relationships between outcomes and activities. The main questions in this step are (W.K.Kellogg Foundation, 2004):

- For each of the activities you plan to do, what short-term and intermediate outcomes do you expect to achieve?
- For each of the activities you plan to do, what long-term outcomes do you expect to achieve for the target group?
- Given your knowledge about the problem at hand and possibly available theoretical insights, what specific activities have you planned?

2. *Subdivide the lists to show the logical sequencing among activities and among outcomes.*

Some activities will work only when they are preceded by other activities. Also, certain outcomes will result from the campaign only after other outcomes. A logic model can bring clarity in the order in which activities and outcomes are expected to occur. Typically, changes in attitude and knowledge will precede changes in behaviour. A straightforward way to obtain an idea of this sequencing among activities and outcomes is to split up the respective columns in Table 4 into: early and later activities, and early and later outcomes.

3. *Add any inputs and outputs.*

The flow of activities and outcomes can be enriched with information on the inputs (before the activities) and the outputs (before the outcomes). Note that the outputs do not represent any change yet. They are only a quantitative representation of the activities performed in the campaign. Outputs should be indicated for every activity that is planned to be done. The inputs are the resources that can be used to support the planned activities.

4. *Draw arrows to depict intended causal relationships.*

Based on the three previous steps, the campaign description already benefits from the structural approach of a logic model. However, when the logic model will be used for evaluation purposes, it is very instructive to add arrows that show the (expected) causal relationships among activities and outcomes. Note that these relationships can go from (early) activity to (later) activity, from activity to outcome, or from (early) outcome to (later) outcome.

These steps will finally result in a model that is comparable to the one in Figure 4 but with a much higher level of detail. Of course, these models can be further elaborated, and the detail that is needed will depend on the campaign and on the purpose of the logic model. Further details on building logic models can be found for example in (W.K.Kellogg Foundation, 2004) and in (U.S. Department of Health and Human Services, 2005).

Source sections 2 – 8: U.S. Department of Health and Human Services 2005

Annex 2: Advantages and disadvantages of data collection techniques: overview

Method of asking

Data collection technique	Advantages	Disadvantages
Mail and self-administered questionnaires	<ul style="list-style-type: none"> - Relatively cheap to administer. - When properly done, a survey produces reliable and valid information. - You can ask many things in one time. - Closed questions are easy to collect and it is easy to compare and analyse the results afterwards. - Open questions can be used to collect more detailed information. - Anonymity of questionnaires may improve response rate for sensitive questions or response from shy individuals who may not respond to other collection methods. - Respondents can complete the questionnaire when it is convenient and interviewer bias is avoided. - Minority groups can be represented fairly. - Analysis of coded responses is relatively straightforward and can provide powerful statistical evidence for the effectiveness of the programme. 	<ul style="list-style-type: none"> - Development of effective questionnaires requires specialist knowledge. Surveys are easily misused, resulting in invalid data and information. - Questionnaires can limit response rates by being perceived as 'boring'. - A low response rate can result in a biased sample (although this can be minimised). - Risk of incomplete questionnaires or someone other than the sampled person completed the survey. - Closed questions can restrict response categories and limit the depth of the evaluation. - Categorising and coding responses to open ended questions is time consuming. - No signals available as to the honesty of respondents' answers. - Literacy skills required for responding can exclude some individuals. - They often ask people questions upon which the respondent has never been asked to express an opinion before. The scaled questions could encourage a response. An attitude may reveal which does not exist outside the context of a questionnaire survey. - People are notoriously forgetful when reporting their own behaviour. - The researcher cannot control the conditions under which a

		mail questionnaire is completed.
Interviews	<ul style="list-style-type: none"> - Usually yield richest data, details, new insights, provide opportunity to explore topics in depth - Afford ability to experience the affective as well as cognitive aspects of responses - Allow interviewer to explain or clarify questions, increasing the likelihood of useful responses - Allow interviewer to be flexible in administering interview to particular individuals or circumstances - The most appropriate form of testing for sensitive subjects - Generates a greater range and depth of response than other methods, especially if a rapport exists between interviewer and interviewee - Can raise issues of which the interviewer was previously unaware, as the topic guide is often very flexible - Small samples, if interviewed in-depth, can provide a large range of views - Higher response rates than questionnaires - Can be face-to-face or via telephone or online - Valuable for developing more effective survey materials for use in an evaluation - Useful for evaluating respondents with low levels of literacy 	<ul style="list-style-type: none"> - Expensive and time-consuming - The less structured the interview, the more difficult and time consuming it is to analyse - The less structured the interview, the more opportunity for bias to creep in to the questioning or interpretation of the answers - Need well-qualified, highly trained interviewers - Interviewee may distort information through recall error, selective perceptions, desire to please interviewer - Researcher's presence may bias responses - Questions must be skilfully phrased so as to avoid leading the interviewee towards a particular response - Large gap between the respondent's knowledge and that of the interviewer - Provides information in a designated 'place' rather than the natural field setting (as in an experiment or a field observation) - Large volume of information may be difficult to transcribe and reduce data - Special equipment to record and transcribe interviews is usually required.
Telephone interviews	<ul style="list-style-type: none"> - You can quickly reach many people across long distances. - More expensive than a mail questionnaire but less expensive than face-to-face interviews. - Less time consuming than face-to-face interviews 	<ul style="list-style-type: none"> - Limited interview length. - Respondents without telephone are impossible to reach. - The call may come at an inconvenient time and thus result in poor answers or no answers at all. - Open-ended questions are difficult to use, and questions requiring visual aids are impossible.

Face-to-face interviews	<p>Very high response rate and permits the longest questionnaires.</p> <p>Possibility to observe the surroundings, nonverbal communication and make use of visual aids.</p> <p>Good approach to gather in-depth attitudes, beliefs, and anecdotal data from individual targets.</p> <p>Personal contact with participants might elicit richer and more detailed responses.</p>	<p>The appearance, tone of voice, question wording, and so forth of the interviewer may affect the respondent</p> <p>It requires time and a quiet area to conduct interviews</p> <p>Requires special equipment to record and transcribe interviews</p>
Key informant – expert opinion	<ul style="list-style-type: none"> -Expert opinion can be used to carry out measurements in areas where objective data are deficient. It is relatively inexpensive and quick data collection technique. -Information concerning causes, reasons, best approaches from an ‘insider’ point of view. -Advice/feedback increases credibility of study. 	<ul style="list-style-type: none"> - Time required to select and get commitment may be substantial. - Informants may interject own biases and impressions, expert opinion presents a credibility problem. - The validity of the measurement can be questioned; therefore expert opinion should not be used as the sole source of data for an evaluation.
Focus group	<ul style="list-style-type: none"> -Excellent approach and a highly efficient way to gather in-depth attitudes, beliefs, behaviours and anecdotal data from a large group at one time -Produces a lot of useful ideas from the target groups themselves -Group dynamics might generate more ideas than individual interviews, can explore the attitudes and opinions that groups have about road safety, rather than just those of individuals, participants’ comments often stimulate a wide variety of ideas amongst the group -Participants are not required to read or write. Technique relies on oral communication -Can be effectively used to focus on details regarding issues found through surveys or other data collection methods 	<ul style="list-style-type: none"> - The environment and situation is not natural to the respondents and may discourage people to give freely comments, peer pressure can introduce conformity to the opinions of the group, making some participants reluctant to offer their real views - Discussion can be hindered if some participants are seen to be ‘experts’ - The data collected could tend to have low validity due to the unstructured nature of the discussion - Requires a skilled and strong facilitator to guide the discussions, keep the group on topic and ensure participation by all members - It requires time to set up and facilitate focus groups, to identify and schedule participants for focus group

	<ul style="list-style-type: none"> - Is relatively inexpensive to run and can generate information fairly quickly (compared to an experimental study) 	<ul style="list-style-type: none"> - Recruitment of those who willingly volunteer can bias the discussion group by excluding those people who do not usually like to participate in groups - Requires special equipment to record and transcribe focus group discussion - The analysis of the discussions can be time-consuming and relatively expensive
On-line focus group	<ul style="list-style-type: none"> - Anonymity of method can be useful when discussing a sensitive topic. - Using the Internet can overcome the expense and logistical complications of participants from different areas travelling to a single location. - On-screen displays lessen the influence of the interviewer's personal characteristics. - Transcripts of the discussion can be automatically printed to vastly reduce the time needed for analysis. - Novelty of the experience will, for some, stimulate both participation and further IT skills. - Relative ease of administration allows several groups to be run in quick succession. 	<ul style="list-style-type: none"> - Anonymity of method can allow participants to create a false impression of themselves and their views. - Restricted participation for those who cannot type quickly (or at all) or those who are not familiar with the Internet as a way of communicating. - Content can be difficult to follow due to the variation in the speed of participants' responses. - Requires a skilled facilitator to keep discussions on topic. - Difficult to explore individual comments in more detail. - Technology may overshadow the purpose of the discussion or even interrupt the discussion.

Method of observing

Data collection technique	Advantages	Disadvantages
On-site observations	<ul style="list-style-type: none"> - Yield data sensitive to changes caused by interventions and can be analysed for time trends - Excellent techniques to discover, observe, and evaluate actual behaviour of participants in a natural setting. - Provide direct information about behaviour of individuals and groups. - Provide good opportunities for 	<ul style="list-style-type: none"> - It can be expensive and time consuming to plan and carry out field trips to collect representative data. It requires time to observe and record observations - Need well-qualified, highly trained observers; the reliability and validity of observations depend on the skills of the observer and on the observer's

	<p>identifying unanticipated outcomes and unusual aspects.</p> <ul style="list-style-type: none"> - Permit evaluator to enter into and understand the situation/context - Researcher can record information as it is revealed - Useful in exploring topics that may uncomfortable for participants to discuss or who have difficulty directly communicating their views 	<p>awareness of any bias he or she brings to the task. Another person carrying out a similar set of on-site observations may observe the same phenomena differently.</p> <ul style="list-style-type: none"> - Observers may encounter difficulties determining gender, age and behaviour of participants. The data recorded is subjective - Restricted viewing angles may obscure important details - Cannot ask questions of participants during observation for example it is not possible to identify those who have/have not seen the road safety campaign - Being observed may affect a participant's behaviour, if they are aware they are being observed
--	--	---

Method of document analysis

Data collection technique	Advantages	Disadvantages
Content analysis	<ul style="list-style-type: none"> - As written evidence, it saves a researcher the time and expense of transcribing. - Useful for determining value, interest, positions, political climate, public attitudes, historical trends or sequences. - An unobtrusive source of information – can be accessed at a time convenient to the researcher. - The study of the minutes of meetings and correspondence can provide an insight into the intervention and the decisions that have been made. - A file review may produce a useful framework and basis for further data gathering. - Audio and visual material provides an opportunity for 	<ul style="list-style-type: none"> - Analysis may be time-consuming. - Requires transcribing or optically scanning for computer entry. - The documents may not be authentic or accurate. - A file review rarely yields information on control groups, except in special cases, such as when files on rejected applicants to a study exist. The lack of such data may make it necessary to collect new data, but these data may not be comparable with the original file data.

	<p>participants to directly share their 'reality', creative in that it captures attention visually.</p>	
Existing statistics	<ul style="list-style-type: none"> - An unobtrusive source of information – can be accessed at a time convenient to the researcher. - Can be used to collect baseline data. - Analysis of statistics is relatively straightforward and can provide powerful statistical evidence for the effectiveness of an intervention. 	<ul style="list-style-type: none"> - Missing data is especially a problem when researchers cover a long time periods. - In order to obtain reliable and good quality results, a large amount of data is needed. - It is often difficult to determine the accuracy of secondary data, the evaluator obviously has no control over the methodology used to collect the data, but still must assess their validity and reliability. - Equivalence reliability can also be a problem. For example, a measure of police controls across a nation depends on each police department's providing accurate information.

Source: Boulanger, A. et al (2007a)

Annex 3: Implications of the evaluation designs on the budget

To summarise, the most expensive design is the counterbalanced design (in yellow in Table 1). This design is very complex, it has several measurements with several intervention groups but it is also very reliable and strong in terms of internal validity.

The one group single measurement design is the least expensive one, with only one group and one measurement (in blue in Table 1). Although the one-group designs pose some serious restrictions to internal validity they are in some case, as presented in this text, the only possible designs to use, as in the case of national campaigns for example (see section 4.2.1).

Table1: Designs according to the number of groups and the number of measures.

		Number of groups		
		1	2	3 and more
Number of measures per group	1	- One group single measurement	- Separate pre-post sample - 2 group after only randomised design	One group multiple measurements
	2		- 2 groups before-after randomised experiment - Non equivalent control group design - Proxy pre-test design (the 2 measures can be made at the same time)	- Switching replications randomised experiment - Switching replication design
	3 and more	- Equivalent time sample - Time series		- Counterbalanced design

- *Non-experimental designs*

One group single measurement: this design is the cheapest one. It is very easy to design and implement but the main problem is its internal validity, it is not reliable.

- *Experimental designs*

One-group multiple measurement: this design is not expensive, as far as it has only one group but, even if there are several measures, the main problem is its internal validity.

The two-group after-only randomised experiment: the two-group after-only randomised experiment is relatively inexpensive and easy to execute, although its shortcomings should be kept in mind. Besides, having only one wave of

measurement can be particularly convenient when, data collection (e.g. administering a questionnaire) is expensive or when the researcher is only concerned with testing cause-effect relationships.

The two group before-after randomised experiment: the costs associated with this type of experimental design are normally somewhat higher than the costs of the simple two-group after-only randomised experiment, simply because additional measurement is taken before introducing the intervention. But at the same time more valid conclusions about the true impact of an intervention can be drawn upon with this type of evaluation design. Besides, it should be kept in mind that the implementation costs vary substantially not only across different evaluation designs but also with respect to the measurement variables chosen for the evaluation – executing observational studies certainly is a more costly alternative compared with surveys.

The switching replications randomised experiment: the implementation costs are believed to be somewhat higher than the costs associated with the two-group before-after randomised experiment, simply because the same intervention is delivered twice and three waves (instead of commonly two) of measurement need to be carried out.

- *Quasi Experimental designs*

Time series: Depending on the mode of implementation, time series design may be inexpensive when statistical data, already collected from other sources (e.g. Statistical Bureau of Accidents) are being used. It becomes proportionally expensive as the number of questionnaire surveys increases, where subjective measurements are required (e.g. attitude or intentional surveys).

Non equivalent control group: The cost of this type of design is higher than the two-group after-only randomised experiment, which rely only on after measurements (because two waves of measurements need to be obtained). Also, the type collected data affect the cost. As it applies in all types of design, observational data collected by other sources, such as a statistical bureau, are less expensive than data which are collected, through a designated questionnaire survey.

Proxy pre-test: The cost for undertaking this test is significantly less than other tests because the before intervention measurements can be performed at the same time as the after intervention measurements, therefore time may be saved. Costs are also reduced, in case the before measurements can be retrieved from an archive (Archived Proxy Pre-test design) and tailored to suit the needs of the particular design. However, the low cost of the design is counterbalanced by the risk of having less-reliable before measurements, due to the proxy variable.

Separate pre-post samples: The cost of this design can be controlled by the designer. The results of previous surveys can be used as a pre-test. Exploitation of results of previous measurements, as in the example above, may lower the overall cost.

Complicated: Normally, this design should be expensive according to the number of groups and the length of the evaluation.

Equivalent time sample: Costs are affected by the type of output data required for the assessment of the interventions. Observation data, already being collected by other procedures, result in a less expensive design, than behavioural data, which require specifically designed measurements.

Switching replication: The costs of this design are higher than in the case where only one before and one after measurement is foreseen. Also, the control group (switching role of the second group), increases if not doubles the anticipated costs of one group only.

Annex 4: Formulas to measure behavioural, normative and control beliefs

- *Measuring behavioural beliefs*

The belief strengths and outcome evaluations for the different accessible beliefs provide substantive information about the attitudinal considerations that guide people's decisions to engage or not to engage in the behaviour under consideration. However, beliefs strength and outcome evaluation can also serve to compute a belief composite that is assumed to determine the attitude toward the behaviour in accordance with an expectancy-value model:

$$A \propto \sum_{i=1}^n b_i e_i,$$

where b_i is the perceived probability that the behaviour will lead to the consequence i , e_i represents the individuals evaluation of the consequence i , and n is the total number of behavioural beliefs. These two elements are multiplied for every consequence and the sum of the products determines the indirect attitude measure (behavioural beliefs)

- *Measuring normative beliefs*

Measures of normative belief strength and motivation to comply with respect to each accessible referent offer a *snapshot* of perceived normative pressures in a given population. An overall normative belief composite is obtained by applying the expectancy-value formula to these measures:

$$SN \propto \sum_{j=1}^n n_j m_j,$$

where n_j is the belief that behavioural performance will be approved of by a specific referent j , m_j reflects the motivation to comply with that referent, and n is the total number of important others. The two elements b and m are multiplied for every referent, and the sum of the products determines the indirect measure of subjective norm (normative beliefs).

- *Measuring control beliefs*

Examination of the average strength and power of different control beliefs provides a picture of the factors that are viewed as facilitating or impeding on performance of the behaviour. Using an expectancy-value formulation, it is possible to compute a control belief composite:

$$PBC \propto \sum_{j=1}^n c_j s_j,$$

where c_j is the perception of how often different facilitating or inhibiting factors appear, s_j is the perceived strength of the same factors, and n is the total number of facilitating or inhibiting factors. Control beliefs are determined by the sum of the products of c s.

Source: Ajzen, 2002

Annex 5: Assumptions in regression analysis

One of the basic assumptions in multiple regression analysis is that the dependent variable should be normally distributed at all points along the regression line, i.e., the residuals should be normally distributed. A second basic assumption is that the variance in the residuals should not be associated with the predicted value of the dependent variable, i.e., the residuals should be homoscedastic (Hankins, French & Horne, 2000). Royston's (1982) extension of the Shapiro and Wilk's *W* statistic can be used to test whether residuals are normally distributed. A Shapiro-Wilk score which is not significantly different from 1 indicate normality. Whether or not residuals are heteroscedastic, i.e., whether the variance in the residuals are associated with the predicted value, can be tested by making a scatter plot of the standardized predicted value of the dependent variable and the standardised residuals (Hankins et al., 2000).

Like in linear regression, logistic regression rests on some basic assumptions. One assumption in logistic regression is that the logistic curve gives a correct description of the empirical relationship (cf. Skog, 2004). When testing a model which includes several independent variables, the Hosmer-Lemeshow test can provide a good estimate of the models fit to the data (Skog, 2004). In short, the result from the Hosmer-Lemeshow test tells us whether the discrepancy between model and data is significant or not.

Annex 6: An indicative case study on statistical data analysis

In the following paragraphs, an example is given, constituting of the overall data analysis process for the assessment of a campaign, based on the listed methods and techniques in the CAST project. Depending on the specific characteristics of the campaign, the data analysis process may be adjusted accordingly.

The example assumes that the evaluation of a road safety campaign has been conducted following on two groups, the experimental and control, and measurements have been made before and after the campaign implementation on both groups.

Two main categories of analysis methods may be implemented: *Descriptive* and *inferential* analyses.

- *Descriptive analyses*

In the framework of the first category, a number of characteristics of the sample (like total number of sample size and population, age, gender, etc.) have to be described.

This is addressed by estimating the frequency distribution per gender, per age etcetera, as well as the mean values and the standard deviations. The estimation of the above statistical values is conducted for the overall sample or the individual groups that have been used, according to the experimental design. For example, in case we have separated our population into *control* and *experiment* groups and we have made measures before and after the implementation of the campaign, the following tables show a possible presentation of the results:

Table 1: Participants gender across the four groups

Group	Gender	Frequency (absolute number)	Percentage
Experiment pre-test	Male		
	Female		
Control pre-test	Male		
	Female		
Experiment post-test	Male		
	Female		
Control post-test	Male		
	Female		

Table 2: Age statistics across the four groups

Group	N	M	Max	Min	SD
Experiment pre-test					
Control pre-test					
Experiment post-test					
Control post-test					

N: number of subjects in group, M: mean value of age, Max: maximum age, Min: minimum age, SD: standard deviation

If we want to present any other important information about our sample (percentage of participants being involved in accidents, etc) we can follow exactly the same procedure by presenting this information through tables or diagrams.

- *Inferential analyses*

After we have described our sample and its characteristics, we now have to test the effectiveness of our campaign. Inferential statistics are used for this purpose. This type of statistics use hypothesis testing to assess the evidence on whether the values of the parameters between two groups are statistically different, and develop regression models, to predict a dependent variable, based on various parameters.

Hypothesis testing

In the hypothesis testing, we have to test two hypotheses, the null and the alternative one. As an example, the null hypothesis could be that there is no difference between our experiment group before and after the campaign. By rejecting the null hypothesis, we are able to conclude that the campaign was finally effective.

In addition to the above hypothesis testing, *p-value* is also estimated, in order to indicate the level of significance, above which the hypothesis rejection stands. In other words, p-value is used to control for type I and II errors. This level is usually set at 0.05, meaning that there is 95% probability that our results are due to the campaign and have not occurred accidentally.

In our example, where we have taken measures before and after the campaign implementation, the most common hypothesis testing in order to compare the results between the before and after phase is *t-test analysis*, which allows us either to reject or accept the null hypothesis we have made and decide whether the means of the two groups are statistically different from each other. When the hypothesis testing procedure is following the situation in which each person of the group is measured twice (before and after) is called a t-test for dependent means. In the case where the before and after measurements are conducted randomly, the test is called a t-test for independent means.

Suppose now that we have to compare values concerning more than two groups. The statistical procedure that is suggested for this purpose is called the analysis of variance (ANOVA), which tests the variation among the means of more than two groups. Of course, we could use ANOVA for a study with two groups only, but t-test is a simpler procedure with the same result.

In order to reduce the error variance and increase the statistical power of the research design by controlling for factors relevant to the (outcome) measurement variable, the analysis of covariance (ANCOVA) is used. In our example, suppose that the pre-test measures are not equivalent for the two groups, the mean pre-test of the experiment group is higher than the mean pre-test score of the control group and also the post-test mean for the experiment group is higher than the post-test for

the control group. The problem here is that we do not know to what extent that differences are due to the difference in the pre-test measures. The analysis of covariance gives the answer to this by checking the significance between the post-test measures for the two groups by holding constant the pre-test measures.

In case that the values of the comparable measures do not follow a normal distribution, data transformation is implemented. Usually methods such as dividing the values with the standard deviation or data ranking may be used.

If the variables are nominal, then chi-square test may be implemented.

Prediction models

In order to build a prediction model that correlates a dependent variable to various parameters, the suggested procedure we could follow is *regression analysis*. Simple regression analysis correlates the dependent variable with one parameter, while in multiple regression analysis; we could use any number of variables as predictors. Predictors have to be selected amongst the variables, which significantly influence the dependent variable.

Suppose that the campaign of our example follows the Theory of Planned Behaviour (TPB). According to this model, we would like to correlate behavioural intentions (dependent variable) with the behavioural (BB), normative (NB) and control (CB) beliefs (independent variables). In case we have more than one variables measuring the same construct (e.g. for BB we have parameters BB1, BB2, BB3), we have to run an *alpha test* Cronbach's alpha measures how well a set of variables can compose a single unidimensional latent construct. Usually, an accepted alpha value for composing a number of parameters to one construct is >0.6 . If in our example the alpha value is greater than 0.6, then the construct to be used in the model is estimated as $BB = (BB1 + BB2 + BB3) / 3$.

All conditions required for implementing a regression model have to stand, i.e. linearity, normal distribution of errors, homoscedasticity of data etcetera.

The regression model is then evaluated against its statistical accuracy and power, using a number of statistical indicators, such as R^2 , standard error of estimate, t-tests, F-tests etc. Attention has to be made to the logical relation indicated in the model, amongst the variables, as well as to the possibility of identifying the constructs for the implementation of the model.