

Practical Guidelines for Assessing the Effects of Drink Driving Interventions

**Advice for road safety professionals in
low- and middle-income countries**

Practical guidelines for assessing the effects of drink driving interventions

Advice for road safety professionals in low- and middle-income countries

Comissioned by IARD

Editors:

Ms. Ingrid Van Schagen, SWOV Institute for Road Safety Research

Dr. Sjoerd Houwing, SWOV Institute for Road Safety Research

With contributions from:

Mr. Niels Bos, SWOV Institute for Road Safety Research

Prof. Dr. Jacques Commandeur, SWOV Institute for Road Safety Research

Dr. Charles Goldenbeld, SWOV Institute for Road Safety Research

Dr. José Pulido Manzanero, National School of Health, Carlos III Health Institute, Madrid, Spain and Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain.

Project Leader: Dr. Sjoerd Houwing

January 2016

SWOV Institute for Road Safety Research

P.O. Box 93113

2509 AC Den Haag

The Netherlands

Telephone: +31 70 317 33 33

Telefax: +31 70 320 12 61

E-mail: info@swov.nl

Internet: www.swov.nl

PREFACE

Road traffic crashes are one of the main causes of injury and death worldwide. It is widely known that impairment by alcohol is an important factor in causing crashes and in worsening the consequences of crashes. Alcohol impairment increases the likelihood of a crash since it increases reaction time and decreases visual acuity; it reduces vigilance and concentration. It results in poorer judgement and slower reflexes on the one hand, and increased confidence on the other. When alcohol is consumed in high quantities, users may experience confusion, blurred vision, clumsiness, memory loss, nausea, passing out, and even coma and death. Alcohol impairment also leads to more serious consequences of a crash since it affects other aspects of traffic safety such as seatbelt wearing, helmet use, and speed choice.

Countries all over the world are taking action to reduce drink driving. IARD, the International Alliance for Responsible Drinking, is working to enhance these efforts and optimize the use of resources. One key aspect of maximizing the impact of available resources is the careful and continuous evaluation of interventions. This helps to ensure that effective interventions will be maintained and replicated, and that ineffective interventions can be improved or eliminated. This publication, developed and written by SWOV, provides general guidelines, dos and don'ts, and advice and considerations for performing evaluation studies. The guidelines target road safety professionals in low- and middle-income countries who do not have a scientific or statistical background. The main aim is to give the audience a taste of evaluation research, allowing them, for example, to make decisions about the set-up of a study, and to understand the scope and limitations of study results.

About SWOV

SWOV Institute for Road Safety Research (www.swov.nl) was founded in 1962 by the Dutch Ministry of Transport and by organizations representing the private sector. SWOV's mission is to improve road safety by knowledge from scientific research. Research results are disseminated to policy makers and those whose work involves road traffic and road safety in the Netherlands and abroad. SWOV is a non-profit, inter-disciplinary research institute. It gives impartial recommendations to road safety professionals based on research results. About 40 researchers, plus supporting staff, work on the areas of the road user, vehicles, the road infrastructure, analysis of road safety, and support of decision-making processes in this field. Other areas, which have a relationship with road safety, are also involved in the research. Both nationally and internationally, SWOV has an excellent reputation, thanks to its high standard of research and its scientifically founded recommendations. More information about SWOV and its research is available at <http://www.swov.nl>.

About IARD

IARD is a not-for-profit organization dedicated to addressing the global public health issue of harmful drinking. Our mission is to contribute to the reduction of harmful drinking and promote responsible drinking worldwide. This is a problem that requires new insights, urgent action, and open dialogue. Central to IARD's work is our role as Secretariat of the Beer, Wine and Spirits Producers' Commitments to Reduce Harmful Drinking.



ACKNOWLEDGEMENTS

We thankfully acknowledge the valuable input and feedback from a large number of people in different stages of the work. We would especially like to thank the three members of IARD's Drink Driving Senior Advisory Group, Mr. David Silcock, Ms. Kathy Stewart, and Prof. Fred Wegman, for their very useful comments on the draft version of these guidelines. We also specifically thank IARD staff and local drink driving program managers for their input and feedback for making the guidelines as practically useful as possible, notably:

Mr. James Yu (China)
Mrs. Mariana Guerra Rendon (Mexico)
Dr. Olanrewaju Onigbogi (Nigeria)
Ms. Margarita Plotnikova (Russia)
Mrs. Lan Huong Nguyen (Vietnam)

CONTENTS

7 **Introduction**

- 7 Drink driving, a threat for road safety
- 8 IARD's Program of Action
- 8 Evaluations and the need for guidelines
- 8 These guidelines

10 **Part I: Basic Theory**

- 10 1. **What is an evaluation study?**
- 11 2. **How to design an evaluation study**
 - 11 2.1. The ideal experimental design: randomized controlled trials
 - 13 2.3. Single data point studies or time series studies
 - 14 2.4. Hierarchical order of the study designs
- 14 3. **Which indicators to use to measure the effects**
 - 14 3.1. Direct crash-related indicators
 - 15 3.2. Indirect crash-related indicators
 - 15 3.3. Safety performance indicators: observed and self-reported
 - 15 3.4. Surrogate indicators
 - 15 3.5. Hierarchical order of indicators
- 16 4. **Hierarchical matrix of study design and indicators**
- 17 5. **Which research method to apply**
- 18 6. **Measurement errors, bias, and confounding**
 - 18 6.1. Random errors
 - 19 6.2. Systematic errors: bias
 - 19 6.3. Confounding variables
- 20 7. **Some remaining requirements**
 - 20 7.1. Costs of the study
 - 20 7.2. Legal and ethical issues
 - 20 7.3. Data quality
- 21 8. **Concluding remarks**

22 **Part II: Evaluation in Practice**

- 22 Related to the legal BAC limit
- 22 Related to enforcement activities
- 22 Change in the penalty system
- 22 **Step 1. What type of intervention do you want to evaluate?**
- 23 Publicity campaigns
- 23 Education
- 23 Introduction of alcohol interlock devices in vehicles
- 23 Direct crash-based indicators
- 23 **Step 2. Which indicators can you use?**
- 24 Observed safety performance indicators
- 24 Self-reported safety performance indicators
- 25 Indirect crash-based indicators

CONTENTS

25	Surrogate indicators
26	Roadside surveys
26	What and how in brief
26	Legal/ethical issues
26	Cooperation of police and/or road authorities
26	Step 3. Which research method do you consider?
27	Size of the sample
36	Hospital studies
38	Offender database and other database studies
39	Contingency tables (Chi-square tests)
39	Analysis of Variance (ANOVA)
39	Step 4. How to analyze the data
40	Logistic regression analysis
47	<i>ANNEX I: Study design and time series analyses</i>
47	1. The ideal study design – a randomized controlled trial
48	1.1. No pre-test
48	1.2. No reference group
49	1.3. No pre-test nor a reference group
49	2. Practice – observational studies
50	3. Time series analysis
52	3.1. Forecasting developments had the intervention not been introduced
54	3.2. Intervention analysis including confounding variables
56	3.3. Intervention analysis with a reference group
59	4. The Empirical Bayes analysis method
59	5. Conclusions
61	<i>ANNEX II. Example of roadside survey questionnaire</i>
63	<i>ANNEX III. A work plan for a mail survey</i>
64	<i>ANNEX IV. Formulating questions about drink driving</i>
64	1. Setting up a questionnaire and defining its scope
65	2. Example questions for a drinking and driving questionnaire
70	3. Final remarks

INTRODUCTION

Drink driving, a threat for road safety

Road traffic crashes are one of the main causes of injury and death worldwide. It is widely known that impairment by alcohol is an important factor in causing road crashes and in aggravating their consequences. According to the World Health Organization, drink driving is one of the five key risk factors influencing traffic crashes and injuries, together with speed and not using motorcycle helmets, seatbelts, and child restraints (WHO, 2014).

First of all, alcohol impairment increases the likelihood of a crash since it increases reaction time and decreases visual acuity, produces lower vigilance, poor judgement, slower reflexes, and increased confidence. In high quantities, users may experience confusion, blurred vision, clumsiness, memory loss, nausea, passing out, and even coma and death. Secondly, alcohol impairment leads to more serious consequences of a crash since it affects other aspects of traffic safety: alcohol results in less seatbelt wearing and helmet use, and is often associated with higher speeds (Valencia-Martin et al., 2008). Blomberg et al. (2005) estimate the risk for drivers with a blood alcohol concentration (BAC)¹ of 0.5 mg/ml to be approximately 40% higher than without alcohol. At 1.0 mg/ml, the risk is almost four times higher, and at a BAC of 1.5 mg/ml it is around 20 times higher. For young drivers, being inexperienced drivers and inexperienced drinkers, the risk starts increasing at even lower BAC levels (Keall et al., 2004).

In high-income countries, around 25% to 30% of fatally injured drivers on average exceed the legal BAC limit: around 25% in Europe (European Commission, 2014), between 25% and 30% in Australia (Australian Transport Council, 2011), and approximately 30% in the United States (NHTSA, 2012). Data from low- and middle-income countries are limited. A literature study by Odero and Zwi (1995) identified 16 relevant studies covering the period between 1966 and 1994. Eight studies looked at alcohol levels in fatalities and found that 33% to 63% of the fatally injured drivers had consumed alcohol; in eight non-fatality studies, the percentage of drivers who had consumed alcohol (determined by blood analysis, breath tests, and interviews) varied between 8% to 28%. The authors correctly conclude that, due to substantial methodological differences, direct comparison of results is inappropriate and that the true prevalence of alcohol-related traffic injuries in these countries remains unknown.

The frequency of drinking and driving also varies around the world. In Europe, based on roadside measurements of a random sample of car drivers in 13 countries, some 1.5% had levels equal to or greater than 0.5 mg/mL (Houwing et al., 2011). In Ghana, in a similar type of study, it was found that over 7% of drivers had a BAC level above 0.8 mg/ml (Mock et al., 2001). When looking at self-reported drink driving frequency, it appears that in Europe 31% of car drivers reported that they had driven after drinking alcohol, and 15% that they had driven over the legal limit during last month (Cestac & Delhomme, 2012). In Vietnam, 77.9% of male drivers reported that they had driven after drinking alcohol in the previous month (Ngoc, Thieng & Huong, 2012). In Nigeria, 67.2% of commercial drivers reported that they had consumed alcohol during the working day and 47% reported that they had been drinking "heavily" (Abiona, Aloha & Fatoye, 2006). Several studies have shown other personal factors associated with drink driving, such as low socioeconomic status, low educational level, alcoholism, and use of illegal drugs (Valencia-Martin et al., 2008; Cestac & Delhomme, 2012; Moskowitz & Fiorentino, 2000). Drink driving is more frequent at night and during weekends (EMCDDA, 2012).

In high-income countries, different measures came into effect over the last few years that reduced the number of crashes caused by drink driving. The success generally rests on the implementation of a combination of:

- A well-communicated legal BAC limit;
- Effective enforcement of this BAC limit, preferably through
 - random breath testing, and
 - a proper, consistent, and swift punishment;
- Further restrictions for young or inexperienced drivers; and
- Treatment for recidivist alcohol-impaired drivers.

1 Blood alcohol concentration (BAC) is a measure to express the level of alcohol intoxication. A BAC of 0.1 means that there is 0.10 mg of alcohol for every ml of blood. The legal drink driving threshold is also expressed in BAC. For example, the most common limit in Europe is 0.5 mg/ml.

INTRODUCTION

This all supported by education and publicity campaigns to increase knowledge and awareness of the problem, as well as public support for interventions. These measures are most effective when implemented as part of a comprehensive, multi-faceted strategy (Shults et al., 2009).

IARD's Program of Action

In late 2012, 13 CEOs from the world's largest beverage alcohol producers announced that they would build on long-standing efforts to reduce harmful drinking by implementing the Beer, Wine and Spirits Producers' Commitments to Reduce Harmful Drinking over five years (2013–2017). The Commitments represent a pledge to implement targeted efforts focusing on five broad areas, one of which is to reduce drink driving by continuing to support the six Global Actions programs in China, Vietnam, Colombia, Mexico, Russia, and Nigeria. In 2015, the program was expanded to begin project implementation in Cambodia, Dominican Republic, Namibia, and South Africa. The projects in each country focus on capacity building and training, implementation of projects at the local level, and monitoring, evaluation, and dissemination of global best practices.

Evaluations and the need for guidelines

An evaluation should be a normal and natural part of interventions and countermeasures. An evaluation study gives an indication to what extent an intervention is effective, which target groups are affected most, and whether the effects are sustainable. This information in turn helps inform whether resources are spent effectively, whether the intervention is to be continued, and whether the focus can be maintained or needs to be redirected. This way it may convince national and local policy makers to make budget for such interventions available.

In order to provide this type of information, an evaluation study needs to be designed correctly, use the appropriate indicators, and collect data in a reliable and sound way. Conducting a good evaluation study involves many challenges, especially when resources and good quality data are limited.

These guidelines

What is included in these guidelines?

The current guidelines give general advice, dos and don'ts, and considerations for performing evaluation studies of drink driving interventions. Among the topics discussed is the study design, answering the question why it is better to have a before and after measurement and explaining the added value of a reference group. The guidelines also discuss the pros and cons of specific indicators: what are the advantages and disadvantages of looking at crash data or at self-reported behavior or at opinions and attitudes about drink driving? What do you need to consider when setting up a roadside survey or a questionnaire study or when analyzing existing crash databases? All examples focus on evaluating drink driving interventions; the general principles however have wider validity.

What is not included in these guidelines?

This document does not discuss possible drink driving interventions and their effectiveness. For this, we refer to the IARD drink driving resources and to the 2007 Drinking and driving – an international good practice manual; a joint production of WHO, FIA Foundation, Global Road Safety Partnership, and World Bank, available in Chinese, English, Portuguese, and Spanish.

For whom are these guidelines?

The guidelines target road safety professionals and their consultants in low- and middle-income countries. It may be especially useful for those who do not have a scientific or statistical background. Some interest in research in general and in evaluating the effect of particular interventions is, however, an advantage.

Why these guidelines?

The main aim of the guidelines is to give the readers a taste of evaluation research. It allows the reader to take the initiative for setting up an evaluation study, to discuss and make decisions about the set-up by balancing quality requirements and practical feasibility of different options, and to understand the scope and limitations of study results. Similarly, it also allows the reader to define the technical specifications for calls for tenders and to judge the scientific quality of the subsequent proposals. The guidelines, however, are not intended to be a "how-to" manual to carry out evaluation research. For example, the guidelines do not present full data collection and data analysis protocols; several aspects of evaluation research require in-depth expertise and experience, and far more detailed information and extensive

INTRODUCTION

study than can be provided in these guidelines. Involving a professional company or a research institute to perform the evaluation is an option to be considered when planning the project.

How are they structured?

The guidelines consist of two parts:

- The first part contains an overview of the theory of evaluating drink driving projects, describing possible research methods and safety indicators, and their value.
- The second part provides the more practical details of performing an evaluation study. The reader can choose the best suitable indicators and research method that fits his/her circumstances and needs by following different steps in the decision tree.

PART I: BASIC THEORY

About Part I

Part I of these guidelines briefly introduces the general theoretical concepts related to evaluation studies. *Chapter 1* describes what, in general, an evaluation study is and what its aims are. *Chapter 2* presents different study designs for evaluating drink driving interventions, pointing out their strong and weak characteristics. *Chapter 3* lists the possible indicators (i.e., the information or data that can be used to measure the effects of a project), again pointing out possibilities and limitations. *Chapter 4* synthesizes the designs and indicators and ranks them in a hierarchical matrix according to their scientific value and robustness. *Chapter 5* introduces the main research methods and *Chapter 6* presents the main sources of research errors and biases. *Chapter 7* gives an overview of some remaining requirements with respect to costs, legal and ethical aspects, and data quality. *Chapter 8* gives some concluding remarks.

As indicated these guidelines focus on the evaluation of drink driving interventions. However, it is very important to consider this evaluation phase in a wider context. This means that the selection of a drink driving intervention should always be guided by existing knowledge of its effectiveness in similar situations or countries. Hence, any decision about an intervention should be preceded by a review of the relevant literature and the experiences elsewhere in order to get an indication of the effects you can expect. In turn, the results of your evaluation study should be published and disseminated so that they can be added to the overall knowledge base of drink driving interventions, helping others when looking for promising interventions. This evidence-based approach ensures optimal use of existing knowledge, preventing the implementation of interventions through trial and error or by reinventing the wheel, and as such preventing inefficient use of scarce resources.

1. What is an evaluation study?

Evaluation is not only meant to provide feedback on the effectiveness of an intervention or program; it can also help to determine whether the intervention is appropriate for the target population, whether there are any problems with its implementation and support, and whether there are any ongoing concerns that need to be resolved as the program is implemented (WHO, 2007). Today, evaluation is more often used to guide continuous quality improvement: “what needs to be changed to improve the effectiveness of a program?” (THCU, 2007). Today, evaluation is more often used to guide continuous quality improvement: “what needs to be changed to improve the effectiveness of a program?” (THCU, 2007).

An evaluation can be defined as “an objective process of understanding how an intervention was implemented, what effects it had, for whom, how, and why” (HM Treasury, 2011).

An evaluation is an objective process of understanding how an intervention was implemented, what effects it had, for whom, how, and why.

Source: HM Treasury, 2011.

This definition distinguishes two main components: how the intervention was implemented and what its effects were. The evaluation of the implementation is called process evaluation. Relevant questions for a process evaluation are whether the implementation was efficient, whether staff was sufficiently qualified, whether program management was efficient, whether the budget was appropriate, etc. The evaluation of the effects is called outcome evaluation: did the intervention have the intended effect on the participants (e.g., in terms of crash involvement, behavior, knowledge, or attitudes). A third type of evaluation is the impact evaluation that looks at either longer term consequences of an intervention or at a much broader set of consequences than those for which the intervention was specifically designed.

The outcome evaluation is probably the most common form of evaluation as it provides information on whether an intervention actually had an effect. The current guidelines focus on this type of evaluation. Outcome-based evaluations assess the effects of an intervention in such a way that any effects found are most likely the result of the intervention itself and not of other co-incidental factors. The central question always is: “what would have happened to those receiving the intervention if they had not in fact received the program?” (World Bank, 2008).

Central to an outcome evaluation is the question: what would have happened to those receiving the intervention if they had not in fact received the program?

Source: World Bank, 2008

PART I: BASIC THEORY

A program or intervention often consists of different elements that are implemented at the same time, for example, the introduction of a new legal limit, accompanied by a publicity campaign and targeted police enforcement. Please note that in that case the results of an evaluation study tell you something about the effect of the combined elements. No statements can be made about the relative contribution of the individual elements.²

For more information about evaluation studies we refer to the IARD Guide to Evaluating Prevention Programs (IARD, 2015).

2. How to design an evaluation study

This chapter is an extract of the more elaborate discussion illustrated with several practical examples in [Annex I](#).

2.1. The ideal experimental design: randomized controlled trials

The ideal evaluation study is a randomized controlled trial. This is a full experimental design that applies:

- A random draw of a sample from the target population (i.e., the population on which the intervention is supposed to have an effect);
- A random assignment of the people in the sample to the treatment group (the people who are subjected to the intervention) and the control group (the people who are not subjected to the intervention);
- Measurements of the indicator (the variable that the intervention is supposed to affect) both before and after the intervention (pre-test and post-test) and in both the treatment and the control group.

Because of the random sampling, this study design allows you to generalize the effect of the intervention observed in the sample to the total target population. In addition, because of the random assignment of the people in the sample to the treatment and the control group, this design can show with close to 100% certainty that the effects, if any, are actually caused by the intervention and not by any other co-incidental factors. Having both before and after measurements allows you to obtain an estimate of the size of the effect.

Unfortunately, when evaluating the effects of road safety interventions such as drink driving program, it is hardly ever possible for practical or ethical reasons to randomly draw a sample from the population and to randomly assign people to a treatment and a control group. For example, when a law or policy is changed in a country, all drivers are exposed to the change, making it impossible to randomly assign some to the new law condition and others to the previous law. In addition, it is probably unacceptable when part of the drivers receive heavier penalties for drink driving than other drivers. Hence, in those cases it is impossible to draw a random sample of drivers who are and are not exposed to the intervention.

2.2. The alternative: quasi-experimental designs

As an alternative, there are several “quasi-experimental study designs.” In these quasi-experimental designs, the challenge is to try to come as close to the ideal study design as possible by:

- Trying to establish what the situation in the post-test period would have been had the intervention not been introduced;
- Controlling statistically instead of experimentally for confounding variables;
- Comparing a (sample from the) treatment population with a (sample from a) reference population that is not expected to be influenced by the intervention, but is as similar as possible to the treatment population in all other respects (e.g., age, gender, education level, region, etc.).

Some of these quasi-experimental designs are more successful in meeting these requirements and hence are stronger than others. The ideal quasi-experimental design consists of a measurement before the intervention (the pre-test) and

2 Scientifically it is of great interest to identify the relative contribution of the different elements of a package of measures. Indications of the individual effect of one element and the added effect of the other elements can be found if the elements are implemented one after another and the additional effects are measured after each phase, assessing what is called the “dose-response” relationship. Since in practice this will be difficult to realize, the guidelines do not further elaborate this approach.

PART I: BASIC THEORY

a measurement after the intervention (the post-test), where part of the population is exposed to the intervention (the treatment group) and part of the population is not (the reference group):

	PRE-TEST	INTERVENTION	POST-TEST
	+	+	+
	+	-	+

When either a pre-test or a reference group is missing, this severely limits the implications of the results.

No pre-test

If there is no pre-test, measurements of the relevant indicator (e.g., the number of alcohol-related crashes or the prevalence of drink driving) in the treatment and reference group is only available for the period after the introduction of the intervention. It is now still possible to test whether there is a significant difference between the treatment and the reference group. However, without a pre-test we are not allowed to conclude that this was the result of the intervention. An observed difference could be caused by other uncontrolled factors, certainly if treatment and reference group are not randomly composed. For example, the region where the treatment group lives may have much larger amounts of traffic than the region where the reference group lives, affecting the number of crashes. Or, maybe the drivers in the treatment region happen to have higher incomes than those in the reference group, thus allowing them to buy better and more protective cars.

If there is no pre-test and no random assignment, observed differences between treatment and reference groups may be explained by other factors than the intervention.

No reference group

With a design without a reference group, we would typically measure the relevant indicator before the introduction of the intervention, and again after the introduction of the intervention. In this situation, it is again possible to statistically test whether there is a difference between the pre-test and the post-test. Should we find a difference, however, then we simply cannot conclude that this was caused by the intervention, since there can be several alternative explanations for the observed change. Other external events may have occurred at the same time as the intervention that are also responsible for the effect. For example, a serious alcohol-related crash took place that generated a lot of publicity in the local media and which may have affected the opinions about drink driving more than the intervention itself. Or, when looking at an effect in terms of the number of crashes, there can be a general decreasing trend in the number of fatal crashes or there may have been a change in the registration of the number of crashes. Last, but not least, there is also the possibility of statistical regression-to-the-mean: interventions are often applied to the most dangerous locations or regions. In that case, the number of crashes in the post-test will automatically tend to shift in the direction of the average number of crashes in all regions due to chance alone.

In other words, without a reference group, we have no idea what would have happened had the intervention not been introduced. The assumption that nothing would have changed between the pre-test and the post-test had the intervention not been introduced is rather hard to defend.

No pre-test nor reference group

The worst case scenario is a situation where we only have data for the treatment group after the introduction of the intervention. In this case, nothing can be said about the effect of an intervention.

If there is no reference group, we cannot unequivocally conclude that a difference between the pre-test and the post-test was caused by the intervention.

PART I: BASIC THEORY

2.3. *Single data point studies or time series studies*

Quasi-experimental designs can refer to studies where there is in principle only one observation or measurement of the relevant indicator in the pre-test and one in the post-test, but also to studies where several repeated observations of the indicator are available in the pre-test and the post-test. The former type of study is called a single data point study, the latter a time series study.

Single data point studies

Single data point studies generally only collect data on the indicator once in the pre-test and once in the post-test, specifically for the evaluation of the intervention at hand. Sometimes, mainly when it concerns a targeted intervention with a relatively short time span, the same people are involved twice (repeated measurements). An example: participants of a driver improvement course have to complete a questionnaire to assess their attitudes about drink driving and their knowledge on the effect of alcohol, once before the course and once after. More often, however, the people tested in the post-test are not the same as those tested in the pre-test. This is called a cross-sectional study. An example: a roadside survey is conducted to assess the alcohol levels among drivers, once before the introduction of higher fines for drink driving and once after. For evaluation, single data point studies are more common than time series studies.

With respect to the study design, the same principles apply as described in *Section 2.2*. Ideally, these studies collect data before and after an intervention and distinguish between people or regions randomly assigned to the treatment group and the reference group.

For some before-after single data point studies the Empirical Bayes method can be applied. The basis of this method is the understanding that interventions (e.g., reconstructing an intersection into a roundabout) are not implemented to a random sample from the target population (intersections), but have been selected for the intervention because they are relatively very unsafe. This selection bias may well result in the statistical regression-to-the-mean effect (i.e., the effect that the number of crashes in the post-test will automatically tend to shift in the direction of the average number of crashes due to chance alone). There are research and statistical techniques that can help correct for potential biases in the data collected with this kind of design (e.g., the Empirical Bayes method) (See [Annex I](#)).

Time series studies

Time series are repeated and sequential measurements over a longer period of time of one and the same phenomenon (e.g., the annual or monthly number of alcohol-related crashes, or opinions about drink driving). Evaluations based on time series typically use existing databases. Again, such repeated measurements are preferably available both in the period before and after the intervention and both for a treatment group and a reference group. An example of such an intervention-based time series with a reference group is an annual nationwide inventory of opinions and attitudes about alcohol consumption, including drink driving. The results for drink driving could be used to assess the effect of theoretical lessons about drink driving during the driver licensing phase, which were introduced in some provinces or states but not in others. For this type of time series study, it is important that the questions in the inventory have remained exactly the same; otherwise, comparisons are impossible. Adding questions over time is not a problem.

For many interventions, especially nationwide interventions, it is difficult to find a suitable reference group. If there is no reference group, it is even more important that the data collection methods remain the same over the study period. For example, if the definition of an alcohol-related crash changes, or if the method of the roadside survey changes (other type of locations, other data collection periods), data are no longer comparable. In that case, a change over time may not reflect an actual change (e.g., due to specific interventions), but can be the result of these changes in data collection methods or definitions. Suppose that before the intervention the drink driving prevalence was assessed during weekend days and after the intervention during weekdays, and you found a decrease in drink driving. It is very likely that this is caused by the fact that drink driving is less common on weekdays, rather than by the intervention.

An advantage of time series studies over single data point studies is that they allow you to evaluate the effect of an intervention while correcting for general trends (e.g., the trend in road crashes, trends in public opinions, trends in alcohol consumption), for seasonal patterns (if we are dealing with quarterly or monthly data for example), and for confounding variables such as changes in mobility (if this data has been measured and is available).

PART I: BASIC THEORY

2.4. Hierarchical order of the study designs

Based on the main characteristics of quasi-experimental designs as described in Sections 2.2 and 2.3, we distinguish six types of study design. As explained, some designs are stronger than others. *Table 1* lists the six designs in a hierarchical order with the strongest design on top. This overview excludes the best design in theory: the randomized controlled trials, which are hardly feasible when evaluating drink driving interventions. Furthermore, study designs with no reference group and no pre-test are excluded as well since these designs cannot tell anything about the effect of an intervention. In other words, *Table 1* says that time series studies are to be preferred above single data point studies; that studies with a reference group are to be preferred above studies without a reference group, and that studies with a pre-test and post-test are to be preferred above studies with a post-test only. Obviously evaluation studies without a treatment group or a post-test do not exist.

Table 1: The hierarchical order of study designs.

	TREATMENT GROUP	REFERENCE GROUP	PRE-TEST	POST-TEST
Time Series Study	Yes	Yes	Yes	Yes
Single Data Point Study	Yes	Yes	Yes	Yes
Time Series Study	Yes	Yes	No	Yes
Single Data Point Study	Yes	Yes	No	Yes
Time Series Study	Yes	No	Yes	Yes
Single Data Point Study	Yes	No	Yes	Yes

The hierarchy is based on the ability of the design to exclude alternative explanations for the effect of an intervention. The lower in the ranking the more difficult it becomes to exclude alternative explanations for an effect.

The hierarchical order provides a relatively simple and theoretical overview of the designs. In the end, the usefulness of a design is largely determined by practical factors as well. Even if a design is theoretically ranked higher, it will be better to choose a lower ranked design if the “best” study is not feasible.

3. Which indicators to use to measure the effects

Another aspect of evaluating drink driving interventions is the choice of indicators: what are we going to measure to assess whether the interventions have the intended effect? This chapter provides some theoretical background on the possible types of indicators for drink driving interventions and ranks them in a hierarchical order with the theoretically best indicators at the top of the list.

3.1. Direct crash-related indicators

Direct crash-based indicators (i.e., road crashes or casualties) are best for evaluating road safety initiatives, since they are most directly related to the safety level of the road system. In that case, the question is “does my project result in fewer crashes or casualties?” They are directly related to the subject of interest, in our case, consumption of alcohol by drivers. An example of a direct crash-related indicator is the proportion of alcohol-related fatalities. An alcohol-related road fatality can be defined as “any death occurring as a result of road accident in which any active participant was found with blood alcohol level above the legal limit” (Assum & Sørensen, 2010). Passengers are not active participants; if a sober driver with two drunk passengers drives off the road resulting in the death of one of them, this is not considered as an alcohol-related fatality. Direct crash-related indicators can only be used if, as a standard, blood alcohol concentrations (BACs) are assessed of all active participants involved in a road crash. Another specific problem with the use of crash data is

PART I: BASIC THEORY

that the information is not complete. Not all crashes are recorded and stored in the database. Particularly the less severe crashes are generally underreported. It is important to keep this in mind when interpreting the absolute crash and casualty numbers (see also [Part II, Step 3](#) on crash database studies).

3.2. *Indirect crash-related indicators*

There are also indirect crash-based indicators. Indirect crash-based indicators do not require information about alcohol levels of crash-involved road users, but instead look at types of crashes that have been proven to relate to the consumption of alcohol by drivers. By far the most valid indicator in this category is the number of night time single vehicle crashes. Alcohol is known to be overrepresented in these types of crashes. One could, for example, look at the number of these crashes as a proportion of the total number of crashes or focus on weekend nights only.

3.3. *Safety performance indicators: observed and self-reported*

A safety performance indicator is a behavior or attitude that has a proven causal relationship with the number or severity of road crashes. A useful safety performance indicator related to drink driving is the consumption of alcohol by drivers. Examples of a safety performance indicator for drink driving are:

- Proportion of drivers over the legal limit in general;
- Proportion of drivers over the legal limit on weekend nights;
- Number of kilometres driving under the influence of alcohol.

There are two approaches to get this type of information: either by observations (road side measurements) or by self-reports from drivers (questionnaires or interviews). Attitudes and opinions about drink driving can also be used as safety performance indicators. These are per definition self-reported. Generally, information from observations is more reliable than from self-reports. The latter may be flawed by, for example, memory limitations and the fact that people may have difficulty confessing that they drink and drive (social desirability).

3.4. *Surrogate indicators*

Finally, there are what we call surrogate indicators. Surrogate indicators are indicators that are, at most, indirectly related to alcohol consumption by drivers and alcohol-related crashes. There is no causal relationship with the prevalence of drink driving. Examples of such surrogate indicators are national alcohol sales or alcohol consumption per capita. If alcohol sales or alcohol consumption increase, this does not necessarily mean that drink driving becomes more common. Maybe drivers have become more aware that drinking and driving should be separated and have chosen alternative transport modes or, more often, drink at home. Similarly, a decrease in alcohol sales does not necessarily mean that drink driving becomes less common; people may have started to brew or distil their own alcohol beverages.

3.5. *Hierarchical order of indicators*

Table 2 provides an overview of the hierarchical order of the indicators. On top are the direct crash-based indicators. This is the most direct safety measure. Second in rank are the observed safety performance indicators, followed by the self-reported safety performance indicators and the indirect crash-based indicators. The surrogate indicators are last in this hierarchical order, only to be used if no better data is available; given the very indirect relationship with road safety, their use is not recommended.

The hierarchical order provides a relatively simple and theoretical overview of the indicators. In the end, the usefulness of an indicator is largely determined by practical factors as well, such as the availability, quality, and reliability of the information needed for an indicator. Even if an indicator is theoretically ranked higher, if the required information is not available or of low quality, it would be better to choose a lower ranked indicator. For example, if an intervention affects a relatively small area, the number of crashes will be a less suitable indicator, simply because their number will be too small to analyze in a statistically meaningful way. Similarly, if crash data or the information about the alcohol consumption of the crash-involved road users is known to be incomplete or unreliable, it is better to choose a relevant safety performance indicator rather than a theoretically higher ranked crash-based indicator.

Anyway, it is recommended to apply two or even more indicators. This may help with interpreting the results, especially if the quality or reliability of an indicator is questionable.

PART I: BASIC THEORY

Table 2: An overview of the hierarchical order of the indicators.

INDICATOR TYPE	EXAMPLE
Direct Crash-Based Indicators	Proportion of alcohol-related fatalities
	Proportion of alcohol-related crashes
	Proportion of Proportion of injured drivers with alcohol levels above the legal limit
Observed Safety Performance Indicators	Observed proportion of drink drivers on weekend nights
	Observed proportion of offenders among breath tested drivers
Self-Reported Safety Performance Indicators	Self-reported prevalence of drink driving
	Attitudes towards drink driving
Indirect Crash-Based Indicators	Proportion of night time single vehicle crashes
	Proportion of night time single vehicle crashes on weekend nights
Surrogate Indicators	Alcohol sales per 100,000 inhabitants
	Liters of alcohol consumed per capita

4. Hierarchical matrix of study design and indicators

Figure 1 is a matrix that combines the hierarchical order of the study designs ([Chapter 2](#)) and the indicators ([Chapter 3](#)). This matrix can be used to decide which study design with which types of indicator will provide the best insight in the effect of a drink driving intervention on road safety. On the left hand side, you see the four types of indicators presented in the previous chapter. On top of the matrix are the study designs, which distinguish between designs with and without a reference group. In the case of a reference group, there is the choice between time series studies and single data point time studies, in both cases with a before measurement (before-after) or without a before measurement (after-only). In the absence of a reference group, only before-after studies are possible, whether as a times series or as a single data point study. The ideal, but in practice hardly feasible, study design with randomized control trials (see [Chapter 2](#)), is not included in our matrix.

The design-indicator combination in the left upper corner of the matrix is the theoretically best combination of study design and indicator. This is a time series study with before and after data of both a treatment group and a reference group, using direct crash-based indicators. However, if reliable direct crash-based indicators are not available and cannot be collected, the use of observed safety performance indicators would be the next best indicator. Using more than one indicator would be helpful when interpreting results. If a before-after time series study is not possible, a single data point before-after study would be the next best alternative. The weakest design-indicator combination is situated in the right lower corner and is a simple before-after single data point study without a reference group, using surrogate indicators.

PART I: BASIC THEORY

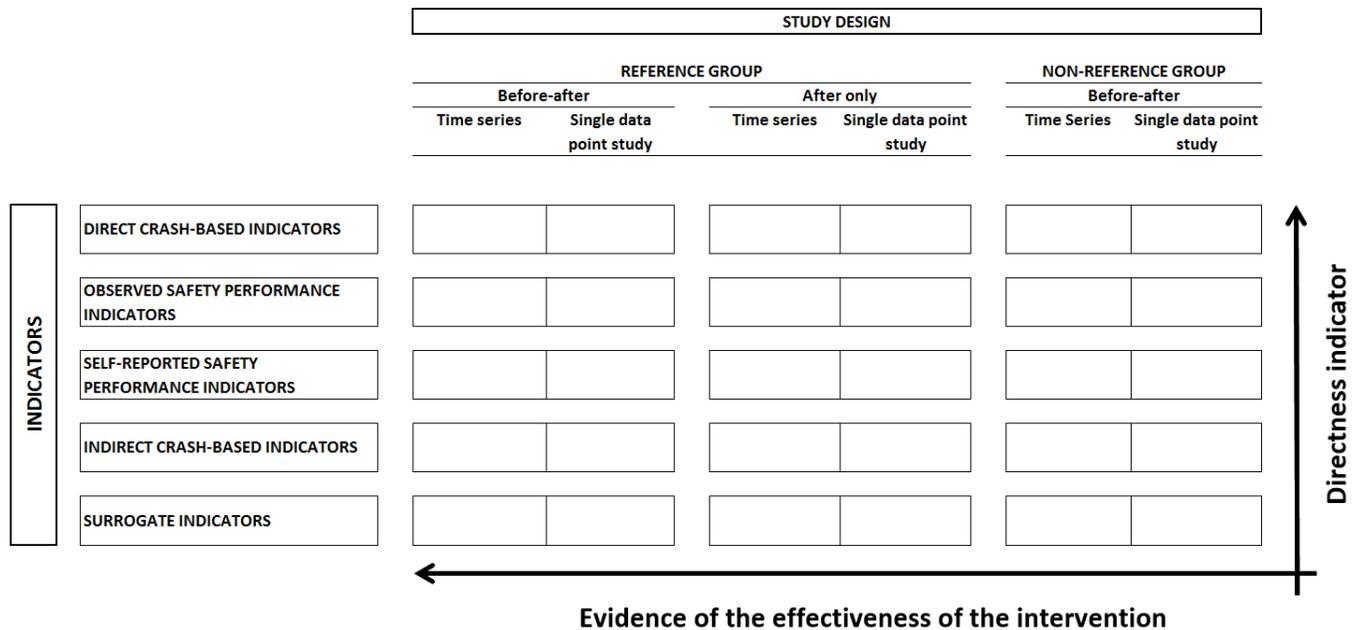


Figure 1. Hierarchical matrix of indicators and methods to evaluate drink driving studies.

5. Which research method to apply

The data from the selected indicators can be collected by different research methods. In this study we have selected four general types:

- Roadside surveys;
- Questionnaire and interview surveys;
- Crash database studies; and
- Hospital studies.

More elaborate information about the practicalities of these research methods can be found in [Part II](#) of this document.

Roadside surveys

A roadside survey provides information about the alcohol consumption by drivers in actual traffic. It is a behavioral safety performance indicator. During a roadside survey, a random sample of drivers is stopped. The alcohol level of each is assessed by means of alcohol breath testing. Some basic information about the driver (e.g., age, gender) and the trip (e.g., length, motive) is observed or asked.

Questionnaire and interview surveys

Questionnaire and interview surveys are means to get information about road users' self-reported behavior, behavior intentions, attitudes, personal and social norms, opinions, etc. Generally, a representative sample of the population of interest is invited to participate.

Crash database studies

A crash database study uses reported crash data to assess the number or proportion and type of relevant crashes and related casualties. For the evaluation of drink driving interventions or for monitoring development in drink driving over time, relevant crashes are alcohol-related crashes (as a direct indicator) or "typical" alcohol-related crash types (e.g., single vehicle night time crashes, which is an indirect indicator).

Hospital studies

In a hospital study, road traffic casualties who are admitted to a hospital are tested for the presence of alcohol in the blood as an indication of the proportion of alcohol-related crashes. It gives information about the percentage of the crash-involved road users who were under the influence of alcohol at the time of the crash.

PART I: BASIC THEORY

Which of the research methods to apply largely depends on the chosen indicator. *Table 3* provides an overview of the appropriate research methods per indicator type. The indicator to choose, in turn, depends on the type of intervention that is going to be evaluated (see [Part II, Step 1](#) for examples).

Table 3: *Appropriate research methods per indicator type.*

INDICATOR TYPE	APPROPRIATE RESEARCH METHOD
Direct Crash-Based Indicators	Crash Database Study
	Hospital Study
Observed Safety Performance Indicators	Roadside Survey
	Offense Database Study
Self-Reported Safety Performance Indicators	Questionnaire Study
	Interview Study
Indirect Crash-Based Indicators	Crash Database Study
	Hospital Study
Surrogate Indicators	Other (database) information

6. Measurement errors, bias, and confounding

Collected data should be a good representation of reality. Two types of errors endanger the representativeness of the data: random errors and systematic errors. Systematic errors are generally referred to as bias. Confounding variables are variables that unintentionally affect the outcome of the study and may form an alternative explanation for a change observed after an intervention. The next three sections briefly discuss the main types of errors, bias, and confounding. It is important to be aware of potential sources of bias in evaluation studies and to try to avoid them. Not all types of bias can always be avoided. Recognizing them and taking them into account when analyzing the data and interpreting the results is then the way to deal with it.

6.1. Random errors

Whenever we measure something – whether we are using a ruler to measure length, a scale to measure weight, or a breathalyzer to measure alcohol consumption – our measurements are subject to fluctuations. For example, when we repeatedly measure the weight of the same object, no matter how careful and precise these measurements are at each weighing, there will always be small fluctuations. Such chance fluctuations are called random error. Random errors tend to push measurements up and down around an exact value, so that the average of all measurements over many trials is very close to the exact value. Random errors thus are likely to cancel out, on average, over repeated measurements. One special type of random error is errors arising from the fact that you base your study on a random sample of the population rather than on the complete population. This sample can, by chance, deviate from the full population on the indicator of interest. This type of random error is called sampling error. That is why study results are often presented by means of a confidence interval. For example, the outcome of a study can be described as: “if I repeat this study 100 times, then the observed reduction will be between 8% and 12% 95 times out of 100; five times out of 100 the observed reduction will be either smaller than 8% or larger than 12%”. The larger the sample, the smaller the sampling error and, consequently, the smaller the confidence interval, the more precise the results, and the more strength in making conclusions.

PART I: BASIC THEORY

6.2. *Systematic errors: bias*

In contrast with random error, systematic error tends to push measurements in the same direction and causes the average or mean value to be too big or too small. Unlike random errors, systematic errors do not cancel out over repeated measurements, but affect all measurements in roughly the same way. A systematic error is often referred to as bias. Bias can be described as a systematic difference between a measurement and the actual value of an object, or as a systematic difference between the value of an indicator obtained in a sample and its actual value in the population. Unlike random errors, systematic errors are not affected by sample size.

Selection bias

There are different types or causes of bias. An important one is selection bias. A selection bias originates from an incorrect selection of people or regions to participate in a study. The main reason this happens is because the selection is not random. For example, in a roadside survey aiming to determine the proportion of drink driving in traffic, it is important that each driver has an equal chance of being stopped and tested. If only male drivers are selected, or only drivers who behave strangely according to the police are selected, this is likely to give an overestimate of the prevalence of alcohol in traffic. Another example of a selection bias is the self-selection bias. A self-selection bias occurs if a treatment, for example, a training course or an alcohol interlock program, is offered on a voluntary basis rather than by random assignment. Those who choose to participate may be a different type of people (e.g., they may be more safety-minded than those who do not choose to participate). Subsequent differences may be partly caused by this initial difference rather than by the treatment. Related to the self-selection bias is the non-response bias. If in a questionnaire study about drink driving attitudes, there is an overrepresentation of elderly people completing the questionnaire and an overrepresentation of younger people not responding, this again will distort the outcomes of the study. The larger the non-response, the more likely that it results in a bias.

Information or measurement bias

Another type of bias is the information (or measurement) bias. This refers to a systematic error in the way data is collected or in the classification of study participants. An example: if in the intervention area drink driving is measured by breath testing and in the reference area by blood samples, there may be a difference that cannot be attributed to the intervention, but just to the different measurement method. Another example of an information bias is the situation where a subject is assigned to a wrong group (e.g., the group without reported crashes or a group who report to never drink and drive). Reasons for such a misclassification could be the use of questions that are multi-interpretable or selective recall by the respondent. In particular, in questionnaire studies and even more so in interview surveys, social desirability is a frequent cause of an information bias: respondents may not want to report behavior or attitudes that they consider socially undesirable. As a result they may, for example, be inclined to report less drinking, less drink driving, and stronger attitudes against drink driving than is actually true.

6.3. *Confounding variables*

Confounding is the phenomenon that there seems to be an effect of a particular intervention, but it cannot be ruled out that there is an alternative explanation for the observed change. For example, a country launches a large-scale national publicity campaign with billboards, radio and television spots, newspaper advertisements, etc. Before the campaign, 62% of the drivers say never to drink and drive; after the campaign, the percentage increased to 78%. Statistical tests show that the increase is statistically significant; it cannot be attributed to just chance. However, one week before the post-test, a severe alcohol-related crash took place with several casualties. The crash was extensively presented and discussed in all national media. In this case, it cannot at all be excluded that the measured effect was at least partly the result of the crash and the media attention.

This example clearly shows the importance of having a research design with not just a pre-test and a post-test, but also a treatment group and a reference group. The effect of the media attention for the alcohol crash would have been similar for the treatment group and the reference group. Hence, any extra effect on the treatment group could with more certainty be attributed to the publicity campaign.

Other examples of confounding variables are changes in crash reporting rates, general trends over time, and seasonal effects.

PART I: BASIC THEORY

7. *Some remaining requirements*

Conducting a study according to the best design and using the best indicator and research method for assessing the effects of an intervention is not always possible due to financial, legal or ethical issues, or data requirements. This chapter provides a general overview of some of the requirements that need to be considered when deciding on the design and indicator(s) for your study: costs, legal and ethical issues, and data quality.

7.1. *Costs of the study*

The costs of the study are based on several aspects related to the design of your study and the chosen research method. For example, more data are needed in a pre/post-test study with a reference group, than in a pre/post-test study without a reference group or in an after-only study with a reference group. The more data to be collected, the higher the costs. The amount of data not only depends on the study design, but also on the appropriate size of the sample. [Step 3](#) in *Part II* provides more information about the required sample size.

Secondly, it depends on the type of data that are needed (i.e., the indicator(s)). If a study can be based on existing data, for example, from national crash databases, police registers, or the results of an ongoing annual questionnaire, the costs for data collection are much lower than if data still have to be collected. Also, the research method (see [Chapter 5](#)) determines the costs of a study: for example, collecting data by means of a roadside survey requires much more staff capacity than a questionnaire survey.

Additional study costs that are easily forgotten but should also be included in the project budget are costs for using existing databases, financial compensation of participants for the collection of body fluid samples (e.g., blood sampling to determine the presence of alcohol), rewards for participants in questionnaire studies, and last but not least, costs for data analysis expertise and statistical advice.

A pilot study can help to get an indication of the exact costs of a study, once a study design and sample size has been determined.

7.2. *Legal and ethical issues*

Some study designs and indicators require legal or ethical considerations. An example of a legal issue is the ownership of a database. If the database is owned by a third party, its use may be restricted or require a fee. The use of databases may also be restricted or even prohibited for privacy reasons. This means that the necessary data may not be fully available.

Another issue is the collection of personal information for questionnaires or interviews. This data and the ways it can be used are often determined by strict privacy legislation. In some countries, this type of study might need formal approval of an ethics board. Requirements and procedures for application and approval widely differ.

Roadside surveys of drink driving among drivers also require several legal considerations, for example, related to the legal capacity to stop drivers in traffic and to take breath samples. Collection of breath or blood samples of injured or killed drivers may be restricted for ethical reasons or require an informed consent by the person in question or a relative. In some countries, postmortem testing is not allowed at all for ethical reasons.

7.3. *Data quality*

For some evaluation studies, data are collected specifically for that study. Other evaluation studies use existing data that have been collected for other purposes than your study. This may affect their usefulness for your study. For example, sometimes the police only test for alcohol among injured drivers if it is considered relevant for legal prosecution purposes. Another example: in hospitals, injured drivers are more likely to be tested for alcohol, if there is a medical need. If there is no medical need, the first goal of the hospital staff will be to take care of the patient. In both cases, not all drivers will be tested and, hence, the data do not tell the complete story. In addition, data collected for other purposes may use variables and values that are less useful for your study. For example, if an intervention is aimed at drivers with a BAC level of 0.8 to 1.3 mg/ml, and the available data only distinguish between drivers with a BAC lower or higher than 0.5 mg/ml, it is impossible to get a good picture of the effects of the intervention.

PART I: BASIC THEORY

8. *Concluding remarks*

Part I has shown that obtaining hard evidence for the effect of an intervention in observational studies is clearly not a trivial matter, especially when it comes to establishing whether the intervention actually caused an effect. Still, we hope to have shown that the evaluation of the effect of an intervention can be made considerably more convincing by, whenever possible:

1. Using a study design that includes a pre-test and a post-test of both a treatment group that was exposed to the intervention and a comparable reference group;
2. Using longer time series of observations in the pre-test period in order to be able to correct for general trends, seasonal patterns, and regression-to-the-mean effects;
3. Correcting explicitly for confounding variables that are known and have been measured in the pre-test and post-test periods by including them as covariates in the intervention analysis;
4. Correcting implicitly for confounding variables that are unknown and/or have not been measured by using information on a reference group or population in the analysis;
5. Applying different types of analysis in order to cross-validate the estimated effects of the intervention.

PART II: EVALUATION IN PRACTICE

About Part II

This second part of the guidelines provides practical information about relevant aspects of an evaluation study. This part is structured as a decision tree. Based on what type of measure you want to evaluate, we provide options for choosing a suitable indicator and some specific recommendations. Subsequently, given the indicator, we present options for the research method and the data collection methods, as well as potential pitfalls for interpretation. Finally, some basic information is given about suitable data analysis methods. These four steps are displayed in the flowchart below.



It should be noted that conducting a scientifically sound evaluation study is a complex and time consuming exercise. These guidelines provide sufficient information to make overall decisions about the type of study and to understand any limitations. For the actual data collection and analysis, specialist help is considered essential.

Step 1. What type of intervention do you want to evaluate?



Explanation/illustration:

You want to evaluate a specific intervention. The decision about which indicator you can use and, subsequently, the research method to use depends on the type of intervention you want to evaluate. Please choose the type of intervention you want to evaluate from the following list.

Related to the legal BAC limit

For example, introduction of a general legal BAC limit, lowering the existing general BAC limit, introduction of a specific BAC limit for special road user groups (e.g., young drivers or professional drivers).

Suggested indicators:

- Direct crash-related: Proportion of alcohol-related crashes
- Self-reported safety performance indicator: Self-reported drink driving behavior

Related to enforcement activities

For example, introduction of random breath testing operations, introduction of breath testing as legal proof in court, an increase in enforcement efforts/capacity, etc.

Suggested indicators:

- Direct crash-related: Proportion of alcohol-related crashes
- Observed safety performance indicator: Observed drink driving behavior
- Self-reported safety performance indicator: Self-reported drink driving behavior

Change in the penalty system

For example, the introduction of heavier penalties, accelerating the penalty process, etc.

Suggested indicators:

- Direct crash-related indicators: Proportion of alcohol-related crashes
- Observed safety performance indicator: Observed drink driving behavior
- Self-reported safety performance indicator: Self-reported drink driving behavior

PART II: EVALUATION IN PRACTICE

Publicity campaigns

These campaigns can either be nationwide or local, aimed at the general public or a special group (e.g., young drivers, professional drivers, a transport company).

Suggested indicators:

- Observed safety performance indicators: Observed drink driving behavior
- Self-reported safety performance indicator: self-reported drink driving behavior, opinions/attitudes

Education

These can be implemented in schools, related to driver/rider training, or rehabilitation/driver improvement courses for violators or repeat violators.

Suggested indicators:

- Observed safety performance indicators: Observed drink driving behavior
- Self-reported safety performance indicator: Self-reported drink driving behavior, opinions/attitudes

Introduction of alcohol interlock devices in vehicles

For violators, for repeat violators, for professional drivers

Suggested indicators:

- Safety performance indicators: Registered drink driving offenses
- Self-reported safety performance indicator: Self-reported drink driving behavior, opinions/attitudes

Step 2. Which indicators can you use?



We distinguish five types of indicators:

- Direct crash-based indicators
- Observed safety performance indicators
- Self-reported safety performance indicators
- Indirect crash-based indicators
- Surrogate indicators

Please note that it can be very helpful to use more than just one indicator.

Direct crash-based indicators

Direct crash-based indicators refer to actual alcohol-related crashes or casualties.

a. Crashes or casualties?

If an intervention specifically aims at the reduction of the number of crashes, the number of crashes should be addressed; if an intervention specifically aims at the reduction of the number of casualties, the number of casualties should be addressed. Otherwise, it is recommended to assess the effects on casualties rather than on crashes. In the end, it is the casualties that determine the psychological and financial burden of a lack of road safety.

b. All levels of injury severities, severe injuries, or fatal injuries?

It is recommended to look at severe and fatal injuries; information about number and characteristics of slight injuries is generally very unreliable.

c. Overall crashes or alcohol-related crashes?

We discourage looking at overall effects on casualties. Effects must be very large to have a perceptible effect on this overall level. It is recommended to look at injuries in crashes where one of the active road users (so excluding

PART II: EVALUATION IN PRACTICE

passengers) had consumed alcohol above the legal limit (or in specified BAC classes). Obviously, a prerequisite is that alcohol consumption is tested and registered systematically and reliably among all drivers in both severe and fatal road crashes.

d. Absolute numbers or proportions?

When looking at alcohol-related crashes, it is recommended to report the effects in terms of a proportion of the total number of crashes. That way the effects are independent of general road safety trends or developments in crash-reporting rates.

e. Body fluid samples or self-reported alcohol consumption?

It is recommended that blood samples or breath tests be used to determine the blood alcohol concentration in involved drivers. The use of breath tests is less invasive and less costly than blood samples, but some injured drivers may not be able to perform a breath test. If blood is routinely collected in an emergency room, an additional blood sample may be easier to take than a breath sample with a breathalyzer. It must be noted that asking crash-involved drivers whether they had consumed alcohol will most likely result in unreliable information about the role of alcohol in crashes. For testing for the presence of alcohol in killed drivers, it is recommended either to take a blood or a urine sample.

Suggested research method:

- Crash database study
- Hospital study

Observed safety performance indicators

Observed safety performance indicators are based on roadside surveys and registered drink driving offenses. They give an indication of the observed prevalence of drink driving.

a. Roadside survey or registered offenses?

It is recommended to use roadside surveys to assess the prevalence of drink driving in traffic. Using data from police offense registers is much cheaper, but the results are more difficult to interpret. The registers offer hardly any background data to understand the value of the indicator. For example, an increase in the number of offenders can be the result of an increase in alcohol consumption by drivers or the result of increased enforcement.

b. Any alcohol level or above the legal limit?

When observing alcohol consumption of drivers, we encourage recording the actual BAC level. This gives highest flexibility when using the data. Generally, it is best to use the legal limit as a reference point in analyses and subsequent communications, as well as limits related to specific types of punishment (e.g., rehabilitation course, alcohol interlock program, driver license suspension), for example, the proportion of BAC offenders above 0.8 mg/ml or above 1.2 or 1.3 mg/ml.

c. Aggregate or disaggregate data?

For evaluation studies, it is recommended to record data as disaggregate as possible (i.e., at the lowest possible level of detail). So, for example, register the exact BAC level (0.4 mg/ml) rather than a class (between 0.1 and 0.5 mg/ml) or the exact age (43 years) rather than the range (between 40 and 50 years). It is always possible to aggregate data to classes (e.g., for reporting purposes), but not the other way around.

d. Absolute numbers or proportions?

When looking at observed drink driving behavior, it is recommended to report the effects in terms of a proportion of the total number of offenders or drivers. That way the effects are independent of general exposure trends or effects of reducing or increasing the chance of being caught.

Suggested research methods:

- Roadside surveys
- Offender database study

Self-reported safety performance indicators

Self-reported safety performance indicators generally focus on the prevalence of drink driving or on knowledge, attitudes, and opinions about drink driving.

PART II: EVALUATION IN PRACTICE

- a. Self-reported or observed prevalence of drink driving?
For collecting information about the alcohol consumption of drivers, it is recommended to use data from observations rather than self-reported information. Self-reported alcohol consumption tends to underestimate the problem, since people may be less inclined to explicitly confess that they drink and drive. Obviously, information about attitudes and opinions about drink driving can only be collected through self-reports.
- b. Questionnaire or interview?
It is recommended to use questionnaires if the objectives and specific research questions can be clearly structured and specified in advance. This will be the case for most evaluation studies. Interviews can be helpful if the research area is still to be explored and defined; results can help to identify the exact questions of a subsequent questionnaire survey.
- c. Open-ended or multiple choice questions?
It is recommended to use multiple choice questions whenever possible. Multiple choice questions largely simplify analysis and prevent difficult interpretations of sometimes ambiguous information.

Suggested research method:

- Questionnaire and interview surveys

Indirect crash-based indicators

An indirect crash-based indicator looks at typical alcohol-related crashes. An example of a typical alcohol-related crash is a single vehicle weekend night time crash. However, an alternative explanation for this type of crash is fatigue. A difference is that with alcohol-related crashes there are generally more people in the car than with fatigue crashes. In general, the same recommendations can be given as for direct crash-based indicators:

- a. Crashes or casualties?
It is recommended to assess the effects on casualties rather than on crashes. In the end, it is the casualties that determine the psychological and financial burden on road traffic.
- b. All severities, severe injuries, or fatal injuries?
It is recommended to measure severe and fatal injuries; information about number and characteristics of slight injuries is generally very unreliable.
- c. Absolute numbers or proportions?
When looking at alcohol-related crashes, it is recommended to report the effects in terms of a proportion of the total number of crashes. That way the effects are independent of general road safety trends or effects of reducing or increasing crash-reporting rates.

Suggested research method:

- Crash database study
- Hospital study

Surrogate indicators

We do not recommend using surrogate indicators, because the relationship with drink driving is too indirect. However, sometimes they can provide some background information when searching for explanations of effects that were detected based on one of the other types of indicators. Typical surrogate indicators are alcohol consumption and alcohol sales. Both have a lot of flaws. Usually, these numbers come from production and import/export statistics, but illegal production and transit traffic are not accounted for. This illegal production may increase when taxes on alcohol are getting higher and consumption of legal alcohol beverages decreases.

Suggested research method:

- Other relevant databases or statistics

We distinguish between the following research methods:

PART II: EVALUATION IN PRACTICE

Step 3. Which research method do you consider?



- Roadside surveys
- Questionnaire and interview surveys
- Crash database studies
- Hospital studies
- Offender or other database studies

In advance, here is a general remark about the target groups and the sampling. If a measure targets a specific part of the population (e.g., young drivers, visitors to bars or sports club canteens, or drink driving violators), the evaluation should focus on these target groups as well. For example, only crashes with young drivers are examined, or questionnaires are distributed among visitors of (a random sample) of bars or sports club canteens, or the number of drink driving violations of previously convicted drink drivers are compared with those of (a matched sample of) not previously convicted drivers.

Roadside surveys

What and how in brief

A roadside observation study provides information about the prevalence of alcohol in actual traffic. It is a behavioral safety performance indicator. A random sample of drivers is stopped. The alcohol level of each driver is assessed by means of alcohol breath testing. Some basic information about the driver (e.g., age, gender) and the trip (e.g., length, motive) is observed or asked. A common approach is one of stratified multi-stage sampling: first study regions are selected, then the areas or roads within these regions, and finally the exact research locations.

Legal/ethical issues

Before starting a roadside study it is necessary to determine whether legal or ethical rules allow for a roadside study. Potential legal and ethical issues that interfere with a roadside study:

- In some jurisdictions, the police can only stop drivers when the driver is under suspicion of drink driving. This means that it is not possible to test a random sample of drivers. In that case, the study can only be based on voluntary participation. This will probably result in “non-response” bias, with drink drivers less likely to cooperate than sober drivers. Consequently, the result of the study will not give a good picture of the prevalence of alcohol in traffic.
- The data collection method is considered to be too invasive for the participant. In general, an alcohol breath test is not considered to be invasive. However, the collection of blood (or urine) is.

Cooperation of police and/or road authorities

If there are no legal and ethical issues or if they have been resolved, the police and/or the road authorities need to be involved. Often, only police officers are allowed to stop drivers from moving traffic. In addition, in many countries police agencies are the only body with the legal power to perform random breath testing. When the police do not cooperate, an alternative is to approach drivers when they are stopped (e.g., at gas stations or at parking places). Participation is then voluntary and likely to result in a non-response bias, strengthened by the fact that drivers are more likely to refuse cooperation without police being around. Hence, a roadside study without the cooperation of police forces is not recommended. Apart from the police, it may be necessary to hire research consultants who collect background information of the breath-tested drivers before they continue to drive again.

Sampling method

Ideally, the study should be based on a random sample of drivers (e.g., every fourth, 10th, 25th driver passing by). However, in practice this may not be feasible due to low numbers of vehicles, for example, during night time hours. As an alternative, random sampling can be realized by selecting the next driver as soon as there is capacity of the police team

PART II: EVALUATION IN PRACTICE

to test another driver, without taking account of special characteristics such as vehicle type, or the gender and age of the driver. The sample should also include a representative sample of locations (e.g., inside and outside built-up areas). For the police, it would feel more logical to set up a roadside study on roads near bars, discotheques, and restaurants. However, testing at these locations will most likely result in an overestimation of the proportion of drink driving. Therefore, it would be better to test drivers at main roads where the largest part of the drivers (both sober and under the influence of alcohol) will be passing. The sampling method (days of week, time of day, type of locations) should be comparable between different measurements (before-after, treatment group and reference group). For example, prevalence of alcohol in traffic in region A will be compared with that in region B. If in region A the breath tests were collected during weekend nights between 10:00 p.m. and 4:00 a.m., breath tests in region B should also be collected during this time interval. If, on the other hand, samples in region B were collected during daytime hours, the proportion of drink drivers is likely to be lower in region B, since in general more alcohol is consumed during night time hours than during daytime hours.

Size of the sample

The desired sample size depends on several factors such as the purpose of the study and the prevalence of alcohol. If, for example, the goal is to find a difference between two regions, it might be good to use a sample size calculation tool to define the sample size based on the expected difference in the prevalence of alcohol in traffic between these regions. If the difference is likely to be small, the sample needs to be larger than if the difference is likely to be large. Moreover, for the same difference, a prevalence closer to 50% requires larger samples than a prevalence closer to 0%. A standard roadside survey with eight to 10 police officers (with six of them having breath test devices) would probably result in 100 tested drivers per hour, if traffic volume is high. If traffic volume is low, and there are fewer than six police officers with breath tests, the total number of tested drivers will be much lower, probably around 30 to 50 per hour. In this case, more survey sessions need to be included to reach the target sample size. When information is gathered on different days and during different time intervals, at least one survey session needs to be available for each day and time period. This session needs to include sufficient tests to get the required power.

Before conducting a roadside survey, it is recommended to estimate the required sample size. The smaller the expected effect of an intervention is, the larger the sample should be to ensure that, if there is an effect, you actually detect it with a statistical test. It would be a pity if a study applied a very good design, but the effect was not detected because the sample was too small. In order to determine the required sample size, you can use a sample size calculator. Several examples of sample size calculators are available. These calculators generally require the following information:

- The expected proportions (of e.g., alcohol positives) in the study groups.
- Whether the test for differences between study groups should be one-sided (you are only interested in an increase or a decrease) or two-sided (you are interested in any change, whatever the direction).
- The level of statistical significance: in general, this is 0.05 (i.e., a 5% risk that we incorrectly conclude that an intervention had an effect).
- The power of the study: this value indicates the probability that the study will obtain a statistically significant effect. A standard value is 0.8, which means that if the study is repeated, significant findings will be found eight out of 10 times.

For example: A roadside survey is planned to compare the proportion of drink drivers in two regions, in one of which an intervention had been implemented. In region A the expected prevalence is 2% and in region B the prevalence is 3%. Based on the outcomes of a simple sample size calculator (for this example we used <http://www.stat.ubc.ca/~rollin/stats/ssize/b2.html>³) the sample size in each region needs to be 3,013 for one-sided testing and 3,826 for two-sided testing. On the other hand, if the expected prevalence in region A is 12% and in region B 13%, then the sample size for each region should be 13,524 for one-sided testing and 17,169 for two-sided testing.

Source: <http://www.select-statistics.co.uk/article/blog-post/the-importance-and-effect-of-sample-size>

3 Please note, to use that page, your browser must recognize JavaScript.

PART II: EVALUATION IN PRACTICE

Information to be collected

As an absolute minimum, the results of the breath tests should be registered. The exact BAC value should be registered, not in classes. Classes can be based on the exact BAC level afterwards. However, if only information on the BAC class is stored, data will be less usable for future analyses. It is strongly recommended to collect some additional information on aspects that are known to be related to the prevalence of drink driving (e.g., driver gender and age, day of week, and time of day) and items that are relevant for policy making (e.g., type of vehicle, trip motive, and type of origin and destination). If breath testing is not mandatory, information should be collected on the reason for refusal as well. Furthermore, it may be interesting to include some questions on knowledge and attitudes towards drink driving and enforcement activities. However, it should be kept in mind that there is a balance between the number of questions that is asked and the time spent per road user. [Annex II](#) provides examples of questions to ask in a roadside survey.

Information registration and storage

Recent breath testing devices often have an integrated data storage system with each breath test result being stored together with the testing time. If additional information is gathered, it is more practical to record the result of the breath test on a form along with this other information. For efficiency reasons, interviewing and recording of information can best be done by researchers rather than the police officer. The information registration form preferably contains a line for each driver, with per column the information items that need to be registered. Use tablets or waterproof ballpoint pens and clipboards for completing the questionnaires. An agreement with the police should be made that in case of an arrest of a driver, the interview can proceed to avoid bias in the results between the response group and the non-response group.

The observation locations

The choice of the exact observation locations is very important. They have to fulfill the following requirements:

- *Working conditions for police officers and field workers*
Working conditions should be safe and secure. This can be enhanced by ensuring vehicles slow down before they enter the test location. Methods include clear signalling and physical traffic calming measures such as making corridors. Furthermore, it is recommended to choose well lit areas with sufficient parking space for vehicles of drivers who are found to be driving with a BAC above the legal limit. Police officers and other persons at the scene need to wear clothes with reflecting parts to make them more visible.
- *Continuous availability of traffic*
The location should be so that sufficient vehicles pass by. The number of police officers and field workers or interviewers needs to be tuned to the supply of vehicles.
- *No or limited possibilities to turn and evade the test location*
To prevent drivers from turning and evading the test location, it is recommended to place a police car or motorcycle at side roads just before the test location. The longer a location is in use during a session, the larger the possibility that drivers become aware of the test location and successfully avoid it. For this reason, it is recommended to change the location regularly. Balancing this requirement with efficiency (it takes time to move the test location), it is recommended to maintain a test location for a minimum of 45 minutes and a maximum of 1.5 to 2 hours.

Data collection time periods

Data collection time periods should be selected in a way that they provide a representative picture of the alcohol consumption of drivers. It is preferred to include survey sessions during all periods of the day and all days of the week, since there may be large differences in prevalence between different time periods. However, in practice this is often not possible.

An alternative is to aggregate the days of the week and the times of the day and make a distinction between daytime and night time hours and between weekdays and weekend days, and have surveys sessions in each of these four periods. If this is not possible and if only one time period can be selected, it is recommended to choose a time period with relatively high prevalence of drink drivers, and with safe working conditions and continuous supply of vehicles.

PART II: EVALUATION IN PRACTICE

Data collection time period

Choosing an appropriate time period is often based on choosing the right mix of the theoretically optimal time period on the one hand, and the practical possibilities and limitations on the other. In Nigeria, roadside drink driving surveys are conducted as part of the Global Action program. When designing the study, it was assumed that many drink drivers drive in traffic before dusk. This is because night time traffic is very low in Nigeria for different safety and security reasons. Therefore, the study does not incorporate night time traffic, but focuses on traffic during daylight hours.

Source: Onigbogi, 2014 (personal communication)

Training and piloting

Drink drivers may be less willing to participate in the subsequent interview or even become aggressive. Therefore, interviewers should be trained to deal with reluctant or aggressive people.

Police officers are inclined to look for “interesting” suspect cases rather than select drivers randomly. It is recommended to brief them about the crucial importance of random selection before the start of each session. Furthermore, it is recommended that the study coordinator stays at the location of the roadside study to make notes of things that disturb the process of random selection.

A pilot study is a very good way to test whether the plans for a roadside study work in practice. In theory everything may look fine, but in practice there will always be unforeseen things. When no problems occur during the pilot study, the data may even be used as additional data points.

Staff and equipment requirements

When conducting a roadside study, sufficient police officers and field workers have to be available to stop, breath test, and interview the stopped drivers according to the agreed working program. Police officers should have time for taking care of violators. Just as an indication, on average and on roads with relatively high traffic volumes, a team of eight to 10 police officers can handle around 100 drivers per hour.

The BAC level is to be assessed by means of breathalyzers. Breathalyzers should be calibrated on a regular basis to prevent measurement errors. After each test a new mouth piece is required for hygiene reasons. Breathalyzers need to present the exact BAC level measured rather than a BAC class (e.g., 0.7 mg/ml, rather than the class 0.5-0.8 mg/ml).

Choosing a breathalyzer

Breathalyzers can be equipped with a semiconductor sensor or with a fuel cell sensor. Breathalyzers with a semiconductor sensor are less reliable and therefore not of use for roadside testing. Some breathalyzers with fuel cell technology allow for passive tests (blow-over) and others for active tests (with a mouth piece). Passive testing will result in a qualitative result: a BAC below or over the legal limit. Active breathalyzers will result in an exact alcohol level and are regarded as more usable for research purposes.

The time needed to take a breath sample differs per device. It is recommended to use a fast breathalyzer with a strong battery so that more samples can be taken during a roadside survey session.

Source: Houwing, 2013

PART II: EVALUATION IN PRACTICE

Data analysis

In general, the result of a roadside study will be a database including records for each driver with information in columns on the different items. This data can be analyzed by means of simple frequency tables, showing, for example, the percentages of drivers in different BAC classes before and after an intervention. The next question is whether a difference is statistically significant. Depending on the type of data and the research question there is a wide variety of statistical analysis techniques available, including chi-square and logistic regression analysis (see [Step 4](#)).

Interpretation of the results

When interpreting and communicating the results, it is important to take account of potential biases. The most common source of bias in roadside studies is the selection bias, in particular the non-response bias (see [Chapter 6](#)). It is recommended to study these sources of bias and their possible effect on the results. Furthermore, it is important to be aware that the results are only valid for the types of locations and times included in the survey. Generalizing the results to other types of locations or other times is not valid unless they are a random sample from the population.

Questionnaire and interview surveys

What and how in brief

Questionnaire and interview surveys are means to get information about road users' self-reported behavior, behavior intentions, attitudes, personal and social norms, opinions, etc. Generally, a representative sample of the population of interest is invited to participate. This section describes different types of surveys and the planning of surveys.

Planning a survey

Planning a survey involves many strategic and operational decisions. These decisions depend upon theory, resources, stakeholder interests, and simple opportunities. Some of the main survey decisions revolve around the following issues:

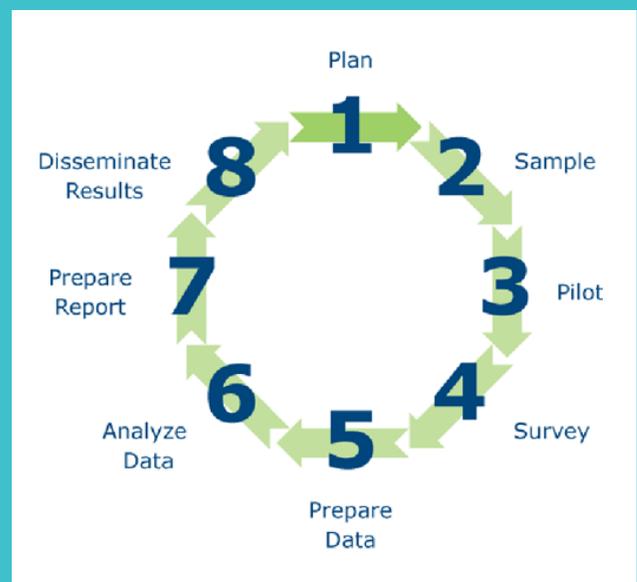
- What are the main objectives of the survey?
- Based on the main objectives, what is the target population of the survey?
- What type of surveys can we use?
- What must be the size of the sample?
- How are we going to sample?

With planning through actual sampling, piloting, data collection, data analysis until and including reporting (see Box), the complete survey cycle may take from four to nine or 10 months.

Eight steps for performing a survey

These eight steps (see [Annex III](#) for more details about the steps in a mail survey) clarify that planning and performing a survey require several skills: planning and organizational skills, expertise in sampling procedure, competence in questionnaire construction, knowledge of data collection and analysis, and skill in report writing. It also shows that the scientific quality of the survey can be compromised by major or small errors occurring at each of these steps.

Source: http://www.aoa.acl.gov/Program_Results/POMP/Chapter5.aspx



PART II: EVALUATION IN PRACTICE

Types of Surveys

Questions can be multiple choice with a restricted number of predefined answer categories, or open-ended where respondents give their own answer. It is recommended to limit the number of open-ended questions. Open-ended questions have to be coded for analyses. This can be a time consuming process that often requires a subjective interpretation of the answers. Questions can be presented as a paper-and-pencil questionnaire, an online questionnaire, a telephone interview, or a face-to-face interview. Interviews in turn can be structured or open.

Sampling method

A survey only involves a sample of the population of interest. However, generally, the aim is to be able to say something about the complete population of interest. Hence, the sample must form a good representation of the complete population. This makes the sampling method crucially important. Sampling is preferably done in such a way that each unit in the population has an equal chance of being selected for the sample: probability or random sampling. On the other hand, non-probability or informal sampling is based on subjective or common sense arguments rather than mathematical reasoning. This is easier and faster, but will result in non-generalizable results.

There are different techniques to realize probabilistic sampling. Simple random samples, stratified random samples, systematic samples, and cluster samples are all examples of probability sampling techniques. These all require fairly in-depth knowledge about sampling and probability theories.

Size of the sample

The appropriate sample size is based on several factors, including the main type of analysis, variability of the variable under study, desired confidence level, and desired margin of error. This requires complex statistical skills. In general, the larger the sample, the more precise the estimate of the variable under study. Just as an indication, a national representative survey among drivers or road users often has a sample size of between 500 and 1,500. If you want to distinguish between specific subgroups (e.g., specific age groups), it is important that the subgroups are sufficiently represented in the survey. As an indication, it is recommended to strive for a minimum size of 50 to 100 respondents per subgroup. In more qualitative oriented studies, (e.g., interview studies) the numbers of respondents can be lower, since the results need not be statistically generalizable.

Maximizing the response rate

The response rate of a survey refers to the percentage of people who were invited to participate and actually completed the survey. The higher the response rate, the more representative the findings for the target population. This is particularly true if the people who do not respond (e.g., young and male) are different from those who do respond. In that case, there is a non-response bias (see [Part I, Chapter 6](#)).

Some tips to increase your response rate:

- Make the survey as short as possible by removing marginal questions.
- Make the survey as interesting as possible to the respondents.
- Offer an incentive or reward.
- Make an appeal to altruism: "I need your help."
- Pre-contact participants to inform them about the survey.
- Send reminders to people who did not respond within a certain time. Multiple follow-ups may be needed.
- Be wise in the timing: avoid sending questionnaires or arranging interviews during Christmas, Easter, and other holidays.

Sources: Smith & Albaum, 2012; DAA, 2014

Response rates are affected by several factors: questionnaire format and length, endorsements, type of presentation, personalization and type of cover letter, specifications about anonymity and confidentiality, availability of rewards or incentives, perceived time for task, and the use of a follow-up reminder (see Box).

PART II: EVALUATION IN PRACTICE

The questionnaire/interview questions

Almost any question can be asked in more than one way and one of the skills in questionnaire design is deciding how many questions to ask, what questions to ask, and how to formulate them.

- A survey should not be too long; an indication of the length must be specified in the introduction of the questionnaire. If it is too long, the response rate will go down. As a very crude indication, questionnaires should not exceed a length of 20 minutes to complete. Interviews should not be longer than 45 minutes. These concern questionnaires and interviews completed/held at home. Questionnaires and interviews at the roadside must be much briefer.
- Be very specific in what information you absolutely need to get out of the questionnaire or interview survey and limit the questions to this. It is tempting to include all sorts of interesting questions just because you have the opportunity.
- Questions should be formulated in simple, unambiguous language; leading questions should be avoided (see Box).
- Make sure to plan a thorough pilot of the draft questionnaire/interviews among the intended target group.

[Annex IV](#) gives further information about formulating questions and response categories.

Some general advice for formulating multiple choice questions:

- 1) Use simple, familiar words (avoid technical terms, jargon, and slang);
- 2) Use simple syntax;
- 3) Avoid words with ambiguous meanings (i.e., aim for wording that all respondents will interpret in the same way);
- 4) Strive for wording that is specific and concrete (as opposed to general and abstract);
- 5) Make response options exhaustive and mutually exclusive;
- 6) Avoid leading or loaded questions that push respondents toward an answer;
- 7) Ask about one thing at a time (avoid double-barreled questions); and
- 8) Avoid questions with single or double negatives.

Source: Krosnick & Presser, 2010

Questionnaire vs. interview

Written questionnaires are more standardized, more objective, easier to implement, more straightforward to analyze, less time-consuming and costly, and less amenable to socially desirable answers than interviews. On the other hand, interviews are more flexible, allowing respondents to bring in their own points of interest and interviewers to ask for further clarifications.

It is recommended to use questionnaires if the objectives and specific research questions can be clearly structured and specified in advance. This will be the case for most evaluation studies. Interviews can be helpful if the research area is still to be explored and defined; results can help to identify the exact questions of a subsequent questionnaire survey.

Interviewer effects

In an interview, the interviewer plays a critical role. Interviewers will handle a survey situation differently and as such influence respondents' answers, thereby producing an "interviewer variance" effect. For example, interviewers may ask questions with slightly different wordings, or help respondents in phrasing their answers. Some interviewers may be more inclined than others to ask for clarification, and without being aware, may vary their tone of voice or facial expressions with certain types of answers. Interviewer effects can be reduced by giving clear instructions and definitions, and by training interviewers to follow the instructions and an agreed protocol.

Analyzing the data

Questionnaire surveys result in frequencies and percentages per response category. Several statistical analysis techniques are available, sometimes available as open source statistical packages, to test whether differences, for

PART II: EVALUATION IN PRACTICE

example, between groups or between two time periods, are statistically significant (i.e. whether the difference is likely to reflect a real difference or is just the result of chance). See [Step 4](#) for some more information about statistical analyses. The analysis of qualitative data, as resulting from open-ended interviews, is less straightforward and less suitable for strict statistical analyses. Responses need to be interpreted and subsequently coded and structured (see Box).

Analysis of qualitative data: deduction or induction

There are two fundamental approaches to analyzing qualitative data:

1. The deductive approach uses a predetermined framework, generally based on an existing theory, to code and analyze the responses. This approach requires that the likely participant responses are already known. This approach is relatively quick and easy, but fairly inflexible and not open to unforeseen outcomes, potentially biasing the analysis.
2. The inductive approach codes and analyzes the response with little or no predetermined framework. In an iterative process, the responses themselves determine the structure of analysis. This approach is comprehensive, but time consuming. It is most suitable when little or nothing is known about the phenomenon under study.

Source: Burnard et al., 2008

Interpretation of the results

In interpreting the questionnaire and interview survey results, the following issues should be kept in mind:

- Self-reported behavior is not the same as actual behavior, and expressed opinions and attitudes may not always reflect the actual opinions and attitudes. Generally, the discrepancy is larger when the topic under study is the subject of strong public opinions and consequently amenable to socially desirable responses. In many countries, drink driving is such a subject.
- General data quality: the credibility of the findings and conclusions is directly related to survey quality. Survey quality is not simply high or low. High quality surveys may have some weak spots and, on the other hand, some surveys may offer quite interesting information despite low quality.
- Specific biases: even when general survey quality is high, there may be specific findings that may raise some doubt as to whether a question has been poorly understood or truthfully answered.
- Comparability with other findings: when a survey generates findings that strongly differ from those of other or earlier research, or from popular perception, this calls for a well-grounded explanation.
- Going beyond the data: just reporting on statistical results often does not help the interested reader or the interested stakeholder very much; the researcher has to conclude what the data mean for policy making or policy performing organizations (relevance data).

Crash database studies

What and how in brief

A crash database study uses reported road crash data to assess the number or proportion and type of relevant crashes and related casualties. For the evaluation of drink driving interventions or for monitoring development in drink driving over time, relevant crashes are alcohol-related crashes (as a direct indicator) or “typical” alcohol-related crash types (e.g., single vehicle weekend night time crashes, indirect indicator).

Purpose of reporting crashes

A crash database generally consists of information about crashes reported to or by the police. The main purpose for reporting a crash is the need to assess whether the law has been violated and, subsequently, who is liable for any damage or injury caused by a crash. The resulting database can be used for epidemiological crash analysis for prevention purposes, but this often is not the primary goal of the data collection effort. As a consequence, the database has several limitations when used for this latter goal.

PART II: EVALUATION IN PRACTICE

Completeness of crash databases

Crash databases do not contain all crashes. Generally speaking, more severe crashes and crashes involving motorized vehicles are more often included in the database than less severe crashes, crashes between non-motorized vehicles, and single vehicle crashes. So the database is not only incomplete, but the missing cases are also a specific type of crash. This results in biased information: the database is not a representative sample of reality; a simple multiplication factor cannot compensate for this. Moreover, reporting rates are not stable over the years. They vary, for example, related to registration method and available traffic police capacity. This type of unreliability of crash databases needs to be noted when using reported crashes to evaluate drink driving interventions or monitor alcohol-related crashes. Analyzing the proportion of alcohol crashes instead of the absolute number may, under some circumstances, help minimize the effect of changing database completeness.

Furthermore, it needs to be noted that there is generally a substantial time lapse (typically between four and eight months) between the at-the-spot crash registration by the police and the record being consolidated and available in the database.

Information in crash databases

Databases contain various characteristics (or variables) of the crash, including severity of the crash, number of injured people and their gender and age, their vehicle type/role, the crash location and road type, main cause(s) of the crash, etc. Usually, the number and type of crash characteristics to be collected is fixed. Increasingly, however, the crash registration by the police is digitized allowing for more flexibility, for example, requiring or providing an option to register additional variables for particular types of crashes.

Information on alcohol in crashes

For the direct indicator of the proportion of alcohol-related crashes or casualties, the database needs to include a variable that describes the presence or absence of alcohol in a crash. This is not as straightforward as it may seem. It requires that all “active road users” who are involved in the crash (driver, rider, and pedestrian, independent of legal liability) are tested for the presence of alcohol. In most countries, alcohol tests are only applied if there is a specific suspicion that alcohol played a role. Road fatalities are tested even less on the presence of alcohol, or test results are destroyed if the injured person dies. This has to do with ethical considerations (integrity of the human body).

For the indirect indicator (e.g., the proportion of single vehicle night time and/or weekend crashes), it is of importance to report the number of vehicles involved and the number of persons or casualties per vehicle. From the reported date/time of the crash it should be possible to identify the night time and/or weekend crashes. Special attention should be given to the unknown/missing time. If this is coded as 0, these crashes should not be taken as equivalent to midnight.

The need for a complete code book

A crash is described in a series of variables. Some variables have a free text entry, others are just binary Yes or No (or unknown). Many variables are described as a code. Codebooks are needed to help the database users to fully understand the meaning of different codes for variables. For example, information about the presence of alcohol can be binary: yes or no alcohol. It has to be made very clear what “yes” means: any alcohol level or an alcohol level above the legal limit. If it means above the legal limit, it has to be specified whether it is the general legal limit or the legal limit for a young driver or a professional driver (if different from the general limit). Alcohol presence can also be described by using different codes and, in that case, specifications are needed, for example:

- 1 = alcohol level not checked
- 2 = checked, but no alcohol observed
- 3 = checked and alcohol observed, BAC level below the legal limit
- 4 = checked and alcohol observed, BAC level above the legal limit

The code book can also require that the exact observed alcohol level is registered or put in classes.

PART II: EVALUATION IN PRACTICE

Linking police and hospital data

It can be useful to link police recorded casualties to hospital data about road casualties. This gives an opportunity to get more information about the nature and severity of the injuries sustained in the crash. Linking police and hospital data also gives an indication of the casualties from road crashes that were not recorded in the police database and hence allows for a more realistic estimate of the true casualty numbers.

Reliably linking police and hospital data requires information in both databases to match persons. Preferably, records are matched by ID number or, in case this number is not available, by the following:

- Date of birth
- Gender
- Date and time of crash/hospital admittance
- Severity

Data analysis

Crash or casualty data analysis is preferably performed by time series analysis (see [Annex I](#)). This means that there is a series of data points (monthly or annual) available of the variable of interest (e.g., alcohol-related road fatalities or single vehicle car-crashes). Ideally, there is a series of data points before and after the intervention, and even better, for both a treatment group and a comparable reference group that was not exposed to the intervention (see [Chapter 2](#)).

As an alternative, one could analyze differences between single data points; again, preferably before and after the intervention and for a treatment group and a reference group. Suitable analysis techniques are presented in Step 4.

Interpretation of the results

When interpreting the results there are two important things to keep in mind. First, the absolute number of crashes or casualties may be rather low. As a consequence, there will be much fluctuation in number over the different measurement periods just because of chance. Differences have to be substantial in order to be able to classify them as an effect of the intervention or, in other words, to be statistically significant. The statistical analyses take this into account.

Secondly, it is not uncommon, and actually very understandable, that road safety interventions are implemented at locations or in regions that have a high number of crashes. Statistically, however, this introduces the phenomenon "regression-to-the-mean." This is because the number of crashes fluctuates randomly. For example, in a particular period, the number of crashes in a region may be much lower or higher than in another region just by chance. However, in the next period, the number of crashes will automatically tend to shift in the direction of the average number of crashes in all regions due to chance alone, rather than due to the intervention (see Box). Time series research designs are less amenable to a regression-to-the-mean effect because they use multiple data points.

Regression-to-the-mean

"Regression-to-the-mean denotes the tendency for an abnormally high number of accidents to return to values closer to the long term mean; conversely, abnormally low numbers of accidents tend to be succeeded by higher numbers. Regression-to-the-mean occurs as a result of random fluctuation in the recorded number of accidents around the long-term expected number of accidents. Regression-to-the-mean threatens the validity of before-and-after studies, but is, at least in large samples, perhaps a less serious threat to validity in cross-sectional studies."

Source: Elvik et al. 2004, page 20

PART II: EVALUATION IN PRACTICE

Hospital studies

What and how in brief

In a hospital study, road traffic casualties who are admitted to a hospital are tested for the presence of alcohol in the blood as an indication of the proportion of alcohol-related crashes. It gives information about the percentage of the crash-involved road users who were under the influence of alcohol at the time of the crash. This type of study can be used as an alternative for crash database studies to collect data on the proportion of alcohol-related crashes in traffic. It can also be used to calculate the relative risk of alcohol, namely by comparing the proportion of alcohol-related casualties in crashes with the proportion of drink drivers in real traffic (by means of a roadside observation study).

Who to be tested?

In principle, only the casualties who at the time of the crash participated in traffic as an active road user (driver, rider, pedestrian) need to be tested, assuming that drunk passengers do not increase the risk of a crash. However, since tests have to be taken as soon as possible after arrival at the hospital (see below), and the role of the victim may not be immediately clear, it is recommended that all admitted road crash victims are tested. In a large hospital in a region with many crashes, random sampling of casualties to be tested is a possibility. Otherwise, it is better to test all road traffic casualties entering the hospital.

When to test?

The alcohol test has to take place as soon as possible after admittance to the hospital. The longer the time between crash and alcohol test, the lower the BAC level will be. It is recommended that a maximum time period is set between crash and the alcohol test sampling. The time period can be reduced when tests already take place at the emergency department. In most cases, blood is taken anyway at the emergency department, so no extra interference is required.

When tests are taken or analyzed at another location, transportation and storage of the blood samples should be arranged. Transportation and storage temperatures should be as cold as possible and at least between 1° and 6° Celsius (34° and 44° Fahrenheit) to keep the blood and alcohol from degrading. For improved preservation, glass tubes with sodium fluoride can be used to collect blood samples.

DRUID: a maximum of three hours between crash and test

The European DRUID project studied the prevalence and risks of different psychoactive substances, including legal and illegal drugs and alcohol in several European countries. Among other things, the project measured alcohol levels of road casualties when admitted to a hospital. The study protocol stipulated that the time between the crash and the testing of the casualties should not exceed three hours. Tests taken longer than three hours after the crash were excluded from the analyses.

Source: Assum et al., 2007

Which hospital(s) to include?

When there is more than one hospital in a region (treatment and reference region), and not all of them can be included in the study, a representative sample has to be chosen. Criteria may include, if relevant:

- Spread over the region of interest in proportion to the population size;
- Spread proportionally over private and public hospitals;
- Spread proportionally over general and specialized hospitals;
- Spread proportionally over religious identities of the hospitals.

More practical criteria to include a hospital are that it receives a substantial part of seriously injured drivers in the region and that it has good facilities for analyzing blood samples.

Ethical permission

Many hospitals require formal permission for this type of study from their boards and/or a medical ethics committee. Generally, this is a time consuming procedure. Hence, it is recommended to start making arrangements with the hospital long before the start of the actual data collection.

PART II: EVALUATION IN PRACTICE

Juridical and financial consequences: informed consent

It should be made clear, both when applying for ethical permission and for the patient's (or his relatives') permission, that the test results do not have juridical or financial consequences, but that they just support fact finding research. It must be ensured that analyses will be anonymous and not traceable to individual persons. This means that some study data need to be aggregated to prevent accidental recognition. Particularly in small hospitals, this can be an issue. The patient (or his relatives) has to give informed consent for the use of the results of the alcohol test; the consent can be obtained after the test has been taken. In case of no-consent, test results have to be destroyed.

Information to be recorded

Preferably, the absolute value of the measured blood alcohol level is recorded together with the time of the crash and the time of the blood test. In addition, some basic information about the patient and the crash is needed, notably, as a minimum:

- Gender
- Age or date of birth
- Date of the crash
- Transport mode
- Active road user or passenger

As background it may be interesting to also register:

- Trip motive
- Single or multiple vehicle crash
- In case of multiple crash: other (type of) vehicles involved
- Use of seatbelt, child restraint, or helmet

This additional data provide more insight into the characteristics and circumstances of drink driving, to allow for adjustment for possible confounding factors, and for linking police and hospital data (see Box on [page 35](#)). Each patient should get a unique respondent code in order to link the alcohol test results and the background information. Generally, to ensure that the hospital study results are anonymous, it is not allowed to use the hospital's patient IDs.

A trial nurse for organizational aspects

When conducting a hospital study, it is recommended to appoint a "trial nurse," a person working at the hospital who is responsible for the testing, ensuring that sampling is random and as complete as possible, collecting or supervising the collection of background information about the patient and the crash, and for administrative operations, including data registration and handling of informed consent.

Data analysis

In general, the result of a hospital study will be a database including records for each driver with information in columns on the different items. This data can be analyzed by means of simple frequency tables, showing, for example, the percentages of drivers in different BAC classes before and after an intervention. The next question is, whether a difference is statistically significant. Depending on the type of data and the research question, there is a wide variety of statistical analysis techniques available, including chi-square and logistic regression analysis. See [Step 4](#) for some more information about statistical analyses.

Interpretation of the results

When interpreting the results of hospital studies, a few things need to be taken into account:

- Hospital studies are often based on a fairly small number of patients. When looking at subsamples (e.g., a particular age group), the numbers are even smaller. This affects the reliability of the analyses.
- Furthermore, the time between crash and sampling is very relevant when interpreting the results. The longer the time, the larger the difference between the actual BAC level at the time of the crash and the measured BAC level in the hospital.
- If blood samples are analyzed for other substances as well, the administration of medicinal drugs should be reported to avoid false positive findings for medicinal drugs.

PART II: EVALUATION IN PRACTICE

Offender database and other database studies

What and how in brief

To evaluate the effect of a drink driving intervention, it is an option to analyze the number and/or the severity of registered drink driving offenses. This requires an accessible database covering the treatment areas and the reference areas for the time period(s) of interest. Drink driving offender databases may just contain offenses detected during targeted enforcement activities, or also include offenses detected because the driver was involved in a crash. It is important to make that distinction. Also, surrogate indicators (e.g., alcohol consumption) could be analyzed by information in databases.

Caveats of offender database research

When conducting an evaluation based on number of offenders, it should be noted that the number of offenders is not just an indication of the prevalence of drink driving. It is also an indication of the effort and efficiency of police enforcement: the higher the level of police enforcement, the more likely that drink drivers will be caught. In other words, an increase in the number of drink driving offenders could reflect an increase in actual drink driving, but also an increase of police enforcement capacity. Additionally, changes in the proportion of offenders in a database (e.g., the number of offenders per 1,000 breath tests) may be the result of differences in enforcement strategies (e.g., more testing during daytime hours), rather than that they are related to interventions. This means drink driving interventions that are known to have an effect on either the enforcement capacity (e.g., a publicity campaign that is accompanied by increased police enforcement) or the enforcement efficiency (e.g., the introduction of evidentiary breath testing) should not be evaluated based on offender database information.

Database research for surrogate indicators

It is not recommended that surrogate indicators be used to assess the effects of drink driving interventions. If they are used, it is very important to verify whether there are alternative explanations for identified changes. For example, an increase in alcohol tax can result in a decrease of alcohol sales, but this does not necessarily affect alcohol consumption and drink driving if many people decide to distill their own alcohol.

Relevant variables

When performing analyses on a drink driving offender database, it is useful to have some background information about, for example, age and gender of the convicted (or accused) driver, the level of alcohol that was detected, and the circumstances of detection (during random breath testing enforcement, selective enforcement actions, crash-involvement). It is recommended to include this type of information in the offender database. This allows for more targeted analyses in case an intervention focused on a specific group of drivers (e.g., young males).

Data analysis

Data analysis of offender and other database research is generally rather straightforward, comparing numbers or proportions before and after an intervention, preferably in the treatment area and a reference area (see [Step 4](#)).

Interpretation of the results

For the interpretation of the results, it should be carefully checked if there are no other explanations for identified differences as mentioned previously.

PART II: EVALUATION IN PRACTICE

Step 4. How to analyze the data



When data are collected or selected, data need to be electronically stored in spreadsheet programs like Excel and SPSS. Subsequently, statistical analyses have to be performed in order to evaluate and quantify the potential effects of interventions on the chosen road safety indicators. This section provides a very brief overview of the main statistical analysis techniques for different types of data. The aim is to give you a taste of this specific aspect of scientific research; without a statistical background, the information presented here is far too limited to allow you to understand the exact ins and outs to conduct your own statistical analyses. This requires in-depth methodological and statistical knowledge and experience, which goes beyond the scope of the current guidelines. So, if you are not a statistician, we advise asking assistance from experts when analyzing data. For those interested in more information on this issue, we refer to some basic text books, for example:

- [Fundamental Statistics for Behavioral Sciences by Robert B. McCall \(2000\)](#)
- [Statistics by William Hays \(2007\)](#)

Contingency tables (Chi-square tests)

Contingency tables, or chi-square tests, can be used to investigate the relationship between two nominal variables (i.e., a variable with categories that have no numerical value), such as group (with the categories treatment group and reference group) and time (with categories pre-test and post-test). The cells of such a table contain frequencies or counts. For example:

	PRE-TEST	POST-TEST
Treatment Group	225	340
Reference Group	241	283

A chi-square test can then be used to investigate whether the two variables are related or not. If the p-value⁴ of the test is smaller than 0.05 (or 0.01) we may conclude that there is a statistically significant relation between the two nominal variables. It is not possible to correct for confounding factors, if information on these should be available. To correct for confounding factors, we need to perform a logistic regression.

Analysis of Variance (ANOVA)

In a quasi-experimental study, there are independent and dependent variables. An independent variable, sometimes called an experimental or predictor variable, is a variable that is being manipulated in order to observe the effect on a dependent variable, sometimes called an outcome variable, which is a variable that is not being manipulated. For example, in a study assessing the effect of higher fines on drink driving, the amount of the fine is the independent variable and the prevalence of drink driving is the dependent variable.

4 The p-value is a concept used in statistical analyses that describes the probability that an identified difference (e.g., between two groups or between the situation before and after an intervention) is a real difference or just based on chance. A p-value of 0.05 means that there is 5% probability that the difference is based on chance; a p-value of 0.01 that there is a 1% probability that the difference is just chance.

PART II: EVALUATION IN PRACTICE

Analysis of Variance (ANOVA) can be used when the dependent or outcome variable is continuous (e.g., the number of crashes). The independent variables are then nominal (e.g., time and group in a randomized controlled trial, or only group in a design without a pre-test, or only time in a design without a reference group). In an ANOVA, F-tests are used to investigate whether mean values in the cells of the design are different or not. Tests with a p-value smaller than 0.05 indicate that there are significant differences between the means of the cells in the design. If you want to test more than one dependent variable simultaneously, you use Multivariate Analysis of Variance (MANOVA). If information on confounding variables was collected or happens to be available in a non-randomized design, then these variables can be incorporated in the analysis as covariates, in which case we statistically control for the confounding variables with a covariance analysis of variance (ANCOVA).

Logistic regression analysis

Logistic regression analysis is used when the dependent variable is nominal and only consists of two categories (i.e., is a dichotomous variable). The independent variables can be both nominal and continuous. This way, the possible relation between an independent nominal variable and the dependent variable can be statistically controlled for confounding variables. The effects of the independent variables are expressed as adjusted odds ratios. If the statistical test for an odds ratio is significant (with a p-value smaller than 0.05), then the odds ratio will differ from 1. This can be either a value between 0 and 1, in which case the corresponding independent variable is associated with a decrease in the likelihood of the outcome, or value larger than 1, in which case the corresponding independent variable is associated with an increase in the likelihood of the outcome. An odds ratio close to or equal to 1 always implies that the corresponding independent variable is not related to the dichotomous outcome variable.

Time series analysis

A time series is a series of repeated observations of one and the same phenomenon over time. Examples are the annual or monthly number of fatalities observed in a region, or the annual or monthly number of motor vehicle kilometres travelled by cars in a region. Time series analysis is concerned with the description, explanation, and forecasting of such developments over time. Time series analysis can indicate the possible effect of an intervention by investigating whether the introduction of the intervention is associated with a break in the general trend of the time series, correcting for possible seasonal variations. It also allows for the evaluation of the effect of an intervention on the general trend while simultaneously correcting for confounding variables (which are then time series also). If a reference time series is available (e.g., the annual or monthly number of fatalities observed in another, but similar, region where the intervention was not introduced) then it becomes possible to compare the change in the “treatment series” with that in the “reference series,” providing an idea of what happens when the intervention is not introduced. For further details, see [Annex I](#).

Empirical Bayes analysis method

The Empirical Bayes analysis method was developed to evaluate the effect of interventions in single data point studies with a before and after measurement, but without a proper reference group. The Empirical Bayes method aims to estimate the effect of regression-to-the-mean and confounding variables like changes in traffic intensity, weather conditions, and, consequently, the remaining effect that is likely to be caused by the intervention. This estimate is based on crash data and relevant characteristics obtained from previous studies in similar type of situations. Supposing that the indicator used in the study is the number of fatalities, the Empirical Bayes method generally consists of the following steps:

- Observe the number of fatalities for the objects to which the intervention is applied in the pre-test and in the post-test periods;
- Determine the expected number of fatalities for these same objects in the pre-test period by determining the number of fatalities and its variance in a reference population that is as similar as possible to the objects to which the intervention has been applied;
- Correct the number of observed fatalities in the pre-test period based on this expected number of fatalities;
- Compare the latter corrected number of fatalities in the pre-test with the observed number of fatalities in the post-test and use this difference to quantify the effect of the intervention.

For further details, see [Annex I](#).

PART II: EVALUATION IN PRACTICE

Statistical significance

In all these types of analysis, we noted that differences and effects are considered to be statistically significant when the p-value of the appropriate statistical test is smaller than 0.05. This implies that we are willing to take up to a 5% risk that our conclusion from the test is incorrect. However, in this context it is important to keep in mind that statistical tests can become significant even when observed differences or effects are very small, especially when samples are very large. On top of statistical significance, we should always also consider the practical significance of the numerical value of an observed difference or effect.

Case studies

What follows are some examples of evaluation studies.

A drink driving survey in China

Type of study: Roadside survey

Method: Survey locations were randomly selected by means of a multi-step approach. In total, 10,685 drivers were tested between December 2006 and March 2007. Six time periods were included in the study: weekdays, weekend days, and holidays, each in night time and daytime hours in two regions. Within each region both urban and countryside areas were included.

Analysis: The BAC distribution was weighted for differences in traffic flow during the roadside survey sessions. Data on traffic flow were gathered by a person who counted traffic during each session.

Of interest: Sites of the roadside survey were selected with care, taking into account safe working circumstances and traffic flow. Samples were weighted based on traffic counts to adjust for differences in time period and type of location. Sample size was predetermined by means of a power study.

Highlight: 4.6% of all drivers were driving with BACs over the lowest legal limit (0.2 mg/ml).

More information: Yuan, A., Li, Y. & Zhang, J. (2013). The result of a baseline survey on drink driving in Nanning and Liuzhou of Guangxi Province, China. *Traffic Injury and Prevention*, 14(3), 230-236.

The impact of alcohol and road traffic policies on crash rates in Botswana

Type of study: Time series analysis on crash rates

Method: Crash rates were calculated by dividing crashes (overall, fatal, single vehicle, night time) by fuel sales on a monthly base in the period January 2004 to December 2011. Fuel sales were used as a surrogate measure for vehicle miles travelled.

Analysis: Time series analyses were used both with data before and after three interventions (introduction taxes on alcohol in October 2008, increase in fines and penalties in April 2009, increase of alcohol taxes in November 2010). Results were presented in figures with observed crash rates, modelled rates, and expected rates based on the pre-intervention period.

Of interest: Different types of indirect crash indicators were used together with a surrogate exposure measure. Data were available on a monthly basis for a period of seven years, which included three interventions.

Highlight: Crash rates were found to decline, but it cannot be excluded that other interventions or the general trend contributed to this finding.

More information: Sebego, M., Naumann, R.B., Rudd, R.A., Voetsch, K., Dellinger, A.M. & Ndlovu, C. (2014). The impact of alcohol and road traffic policies on crash rates in Botswana, 2004-2011: a time-series analysis. *Accident Analysis & Prevention*, 70, 33-39.

PART II: EVALUATION IN PRACTICE

Effect of new traffic law on behavior of drivers at alcohol outlets in Brazil

Type of study: Questionnaire study combined with breath test results of drivers

Method: In 2008, new legislation was introduced in Brazil that made drink driving above the 0.2 mg/ml level a misdemeanour and above 0.6 mg/ml a criminal offense. Samples were collected from drivers who left alcohol outlets. They answered a questionnaire, were breath tested, and an oral fluid sample was taken to detect the use of psychoactive substances other than alcohol. In total, 3,118 individuals were approached between April and December 2009.

Analysis: Frequency analysis was used to determine the proportion of drivers who had used alcohol or other psychoactive substances. Chi-square testing was used to determine differences between the people who had indicated that they had changed their behavior after the new law and the people who did not. Logistic regression analysis was used to determine the relative importance of these distinguishing factors.

Of interest: Samples of drivers were collected without roadside enforcement activities, resulting in a low non-response rate.

Highlight: Awareness of the law had not been enough to bring about behavioral change.

More information: De Boni, R.B., Pechansky, F., Vasconcellos, M.T. & Bastos, F.I. (2014). Have drivers at alcohol outlets changed their behavior after the new traffic law? *Revista Brasileira de Psiquiatria*, 36(1), 11-15.

Combining police and self-reported data to estimate drink driving in Brazil

Type of study: Questionnaire study and offender database study

Method: Between August 2011 and February 2012, questionnaires were taken from 805 drivers passing a sobriety roadblock of the police. Results were combined with the offense results of the police. In total, four roadside survey sessions were conducted in two state capitals.

Analysis: Three correction techniques were used to adjust for differences between self-reported prevalence of alcohol and prevalence of alcohol based on police offender databases.

Of interest: Detailed information is given on the procedure of having interviewers at police roadblocks. Traffic counts per type of vehicle were made by one of the researchers. In Brazil, a driver has the right to refuse a breath test, which is one of the major reasons why official police data need to be adjusted before used.

Highlight: Prevalence in traffic in the two Brazilian state capitals is estimated to be 10% to 30% depending on the method for adjustment.

More information: Sousa, T., Lunnen, J.C., Gonçalves, V., Schmitz, A., Pasa, G., Bastos, T., Sripad, P., Chandran, A. & Pechansky, F. (2013). Challenges associated with drunk driving measurement: combining police and self-reported data to estimate an accurate prevalence in Brazil. *Injury*, 44(4), 11-16.

PART II: EVALUATION IN PRACTICE

The effects of lowering the legal limit in North Carolina, USA

Type of study: Crash database study

Method: The effect of lowering the legal drink driving limit from 1.0 mg/ml to 0.8 mg/ml in North Carolina, USA, was measured by monitoring the alcohol-related motor vehicle crashes between 1991 and 1995. Data for both before and after the change (October 1993) were used and results were compared with the trend in 37 other states in which the legal limit remained 1.0 mg/ml.

Analysis: Time series intervention analysis was used to determine the effect of the lower legal limit. Furthermore, the slope of the trend was compared with that in other states where the legal limit had not changed.

Of interest: It describes the results of several direct and indirect crash indicators and provides a good discussion of possible reasons for not finding an effect by means of this study design.

Highlight: No effect was found when controlling for the overall downward trend, both before the change of the legal limit in North Carolina and the trend as found in the reference states.

More information: Foss, R.D., Stewart, J.R. & Reinfur, D.W. (2001). Evaluation of the effects of North Carolina's 0.08% BAC law. *Accident Analysis & Prevention*, 33, 507-517.

Effects of a lower legal limit in five states in the USA

Type of study: Crash database study

Method: The effect of lowering the legal drink driving limit in five different states was measured on six different types of alcohol-related motor vehicle crashes in those states. Data for both before and after the reduction of the legal limit were used, but no reference group was included.

Analysis: The percentage change before and after the change in legislation was compared for six different crash indicators. Chi square tests were used to analyze differences. For this study, the percentage change was considered statistically significant at 0.10. The time periods were chosen based on four considerations: (1) using the latest available year of crash data at the time of the analysis (1992); (2) using at least two years of crash data before vs. after, where possible; (3) the earliest available year of crash data for which it was possible to make reliable estimates of alcohol involvement (1982); and (4) the effective date of the legislation.

Of interest: The analysis method for comparing the crash data between the before and after situations is different when compared with other studies that often use time series analysis.

Highlight: The results suggest a significant effect of the new legislation. However, possible confounding factors were not taken into account in this study.

More information: Johnson, D. & Fell, J. (1995) The impact of lowering the illegal BAC limit in five states in the US. 39th Annual Proceedings of the Association for the Advancement of Automotive Medicine. Chicago, IL, 16–18 October 1995.

PART II: EVALUATION IN PRACTICE

Effects of a drink driving program for recidivists in Australia

Type of study: Offender database study

Method: Two groups of, in total, 3,096 drink driving recidivists who completed the Sober Driver Program (SDP) were compared with 1,405 drink driving recidivists who did not follow the program. All were monitored during a period of 2,000 days and the rates of non-offending were compared.

Analysis: A survival analysis regression model was used to find differences between the treatment group and the reference group and to assess whether or not factors such as gender and being an Aboriginal had an effect on the results.

Of interest: The outcomes were presented in the number of prevented reoffenses on an annual basis.

Highlight: At the end of the 2,000 days, approximately 15% of the treatment group reoffended, in comparison with approximately 20% of the reference group.

More information: Mazurski, E. Withaneacchi, D. & Kelly, S. (2012). The NSW Sober Driver Program: recidivism rates and program parameters, Roads and Traffic Authority, New South Wales, Australia.

Effects of a therapeutic drink driving program in Australia

Type of study: Questionnaire study

Method: In Australia, 150 participants of the Under the Limit (UTL) program agreed to be followed up to evaluate this drink driving program. A questionnaire was sent to them by post or by e-mail. Due to the low non-response rate, different follow-up actions were taken to increase the number of respondents: by postal mail, e-mail, and telephone. All in all, the response rate was 20%.

Analysis: Results on demographic information and knowledge and attitudes towards drinking and drink driving were reported by frequency analysis, but the study did not include a before measurement nor a reference group.

Of interest: The discussion clearly states some points for approval, including using a study design with a pre-assessment and post-assessment.

Highlight: The participants had poor knowledge of drinking and drink driving, which led to changes to the program.

More information: Sheehan, M. C., Fitts, M. S., Wilson, H. & Schramm, A. J. (2012). A process and outcome evaluation of the Under the Limit (UTL) therapeutic drink driving program for recidivist and high range offenders. Centre for Accident Research and Road Safety - Queensland, Brisbane, QLD.

PART II: EVALUATION IN PRACTICE

Monitoring drink driving rehabilitation courses in the UK

Type of study: Offender database study with questionnaire to drink driving offenders

Method: Between April 2000 and March 2005, 92,697 drink driving offenders were recorded and included in the database study. In all, 44% attended a drink driving rehabilitation course, 54% did not, and 1% did not follow the course, but were still eligible to participate.

Analysis: Reconviction rates were calculated for participants and non-participants by means of survival analysis. An invitation for a questionnaire assessment was sent to 10,028 drink driving offenders. However, the response rate was only 8.4%.

Of interest: The six principal reasons for drink driving were: I thought it was safe (33%), I thought I was under the legal limit (26%), I did not think about whether I was under the legal limit (23%), I did not have far to travel (23%), I thought I would get away with it (19%), and I had to go somewhere unexpectedly (17%).

Highlight: Non-participants were 44% more likely to reconvict within a period of five years than participants.

More information: Inwood, C., Buckle, G., Keigan, M. & Borrill, R. (2007). Extended monitoring of drink-drive rehabilitation courses. Final Report. TRL Report 662. Crowthorne, Transport Research Laboratory.

Drinking and driving in Vietnam: Prevalence, knowledge, attitudes, and practices in two provinces

Type of study: Roadside survey and questionnaire study

Method: Between November 2010 and December 2011, breath alcohol concentration levels were collected from 8,404 drivers by the police in two provinces. Information was collected on knowledge, attitudes, and practices (KAP) by asking 1,661 randomly selected drivers at gas stations at quarterly intervals.

Analysis: Poisson regression models were used to assess differences in prevalence of drink driving over time. For the questionnaire, cross tabulations were made and analyzed by means of chi-square calculations to assess associations.

Of interest: The study was conducted in a period with enhanced mass media campaigns and increased drink driving enforcement. Furthermore, this study provides a baseline for future studies. The Vietnamese questionnaire on KAP is an example of a well-designed questionnaire.

Highlight: The results of the roadside survey and the questionnaire show that there is a need to further improve drink driving enforcement.

More information: Bachani, A.M., Jessani, N.S., Pham, V.C., Quang, L.N., Nguyen, P.N., Passmore, J., Hyder, A.A. (2013). Drinking & driving in Viet Nam: Prevalence, knowledge, attitudes, and practices in two provinces. *Injury*, 44(4), 38-44.

REFERENCES

1. Abiona, T.C., Aloha, O.O. & Fatoye, F.O. (2006). [Pattern of alcohol consumption among commercial road transport workers in a semi-urban community in south western Nigeria](#). *East African Medical Journal*, 83, p. 494-499.
2. Assum, T., Frison, G., Hels, T., Houwing, S. & Mathijssen, R. (2007). [Uniform design and protocols for carrying out case-control studies](#). *DRUID Driving Under the Influence of Drugs, Alcohol and Medicines*. D2.1.2.
3. Assum, T. & Sørensen, M. (2010). [Safety Performance Indicator for alcohol in road accidents; international comparison, validity and data quality](#). *Accident Analysis & Prevention*, 42(2), p. 595-603.
4. Australian Transport Council (2011). [National Road Safety Strategy 2011-2020](#).
5. Blomberg, R.D., Peck, R.C., Moskowitz, H., Burns, M. & Fiorentino, D. (2005). [Crash risk of alcohol involved driving: A case-control study](#). Dunlap and Associates Inc., Stamford.
6. Burnard, P., Gill, P., Stewart, K., Treasure, E. & Chadwick, B. (2008). [Analysing and presenting qualitative data](#). *British Dental Journal*, 204, p. 429-432.
7. Cestac, J. & Delhomme, P. (Eds.) (2012). [European road users' risk perception and mobility: the Sartre 4 survey](#). Brussels, European Commission.
8. DAA (2014). [Response Rates](#). Data Analysis Australia, Broadway, Nedlands, West-Australia.
9. EMCDDA (2012). [Driving under the influence of drugs, alcohol and medicines in Europe](#). Findings from the DRUID project. Luxembourg, Publications Office of the European Union, 2012.
10. Elvik, R., Christensen, P. & Amundsen, A. (2004). [Speed and road accidents. An evaluation of the Power Model](#). TØI report 740/2004. Institute of Transport Economics TOI, Oslo.
11. European Commission (2014). [Study on the prevention of drink-driving by the use of alcohol interlock devices](#).
12. Houwing, S. (2013). [Estimating the risk of driving under the influence of psychoactive substances](#). PhD Thesis University of Groningen. SWOV Institute for Road Safety Research, Leidschendam.
13. Houwing, S., Hagenzieker, M., Mathijssen, R., Bernhoft, I.M., Hels, T., Janstrup, K., Van der Linden, T., Legrand, S.-A. & Verstraete, A. (2011). Prevalence of alcohol and other psychoactive substances in drivers in general traffic. [Part 1: General results](#) and [Part 2: Country reports](#). *DRUID Driving Under the Influence of Drugs, Alcohol and Medicines*, D2.2.3.
14. H.M. Treasury (2011). [The Magenta Book: guidance for evaluation](#). London, H. M. Treasury.
15. IARD (2015). [A guide to evaluating prevention programs](#). Washington, DC, International Alliance for Responsible Drinking.
16. Krosnick, J.A. & Presser, S. (2010). [Question and Questionnaire Design](#). In: P.V. Marsden & J.D. Wright (Eds.). *Handbook of Survey Research 2nd Edition*. Bingley, UK, Emerald Group Publishing, p.263-313.
17. Mock, C.N., Asiamah, G. & Amegashie, J. (2001). [Random roadside breathalyzer survey of alcohol impaired drivers in Ghana](#). *Journal of Crash Prevention and Injury Control*, 2, p.193-202.
18. Moskowitz, H. & Fiorentino, D. (2000). [A review of the literature on the effects of low doses of alcohol on driving related skills](#). Washington, DC, National Highway Traffic Safety Administration.
19. Ngoc, L.B., Thieng, N.T. & Huong N.L. (2012). [The drink driving situation in Vietnam](#). *Traffic Injury Prevention*, 13, p. 109-114.
20. NHTSA (2012). [Traffic Safety Facts, 2010 Data: Alcohol-Impaired Driving](#). Washington, DC, National Highway Traffic Safety Administration.
21. Odero, W.O. & Zwi, A.B. (1995). [Alcohol-related traffic injuries and fatalities in LMICs: a critical review of literature](#). In: C.N. Kloeden & A.J. McLean (Eds.) Proceedings of the 13th International Conference on Alcohol, Drugs and Traffic Safety, Adelaide, 13–18 August 1995. Adelaide, Road Accident Research Unit, p. 713 - 720.
22. Shults, R.A., Elder, R.W., Nichols, J.L., Sleet, D.A., Compton, R. & Chattopadhyay, S.K. (2009). [Effectiveness of multicomponent programs with community mobilization for reducing alcohol-impaired driving](#). *American Journal of Preventive Medicine*, 37(4), p. 360-71.
23. Smith, S.M. & Albaum, G.S. (2012). [Basic Marketing Research: 1. Handbook for Research Professionals](#). Qualtrics Labs Inc.
24. THCU (2007). [Evaluating health promotion programs](#). Toronto, University of Toronto, Department of Public Health Sciences, The Health Communication Unit, 2007.
25. Valencia-Martin J.L., Galan, I., Rodriguez-Artalejo, F. (2008). [The joint association of average volume of alcohol and binge drinking with hazardous driving behaviour and traffic crashes](#). *Addiction*, 103(5), p. 749-57.
26. WHO (2007). [Drinking and Driving: A road safety manual for decision-makers and practitioners](#). Geneva, Global Road Safety Partnership.
27. WHO (2014). [Global status report on road safety 2013: Supporting a decade of action](#). Geneva, World Health Organization.
28. World Bank (2008). [Overview: Impact evaluation, 2008](#).

ANNEX I

ANNEX I: Study design and time series analyses

By Prof. Dr. Jacques J.F. Commandeur

This annex is a more elaborate discussion on study designs with a special focus on time series analysis. It first starts by discussing how the evaluation of interventions on drink driving should ideally be performed in such a way that we may conclude that the intervention has actually caused the effect (if any is found). In short, such an ideal evaluation can only be achieved if we are able to experimentally manipulate the circumstances during the introduction of the intervention. However, in practice it is usually impossible to exercise experimental control over these circumstances. Apart from the strict experimental evaluation of interventions, we also discuss two other methods typically suited for the evaluation of interventions in observational studies: time series analysis and the Empirical Bayes method.

1. The ideal study design – a randomized controlled trial

The ideal study design is a randomized controlled trial (RCT). The reason that we first discuss the ideal study design is that the principles underlying the RCT design can teach us a lot about the best way to proceed when we need to evaluate the effect of interventions on drink driving in a non-experimental observational setting.

The ideal study design for the evaluation of interventions requires the following experimental manipulations:

- A random draw of the sample of objects from the target population (i.e., the population on which the intervention is supposed to have an effect);
- A random assignment or allocation of the study objects in the thus obtained sample to one of the following two experimental conditions: one where the objects are subjected to the intervention under investigation (this is called the experimental or treatment group of objects), and another where the objects are not subjected to the intervention (this is called the reference group of objects).

The first experimental manipulation allows us to generalize the effect of the intervention (whatever it may be) found in the sample to the total target population because a random draw guarantees that the sample is representative of the target population. The second experimental manipulation allows us to conclude that the intervention has caused the effect (if any) because a random assignment of the sample to the two experimental conditions guarantees that possible confounding variables are equally distributed over the two conditions. Furthermore, in a randomized controlled trial the road safety indicator at hand must be measured during a period of time that precedes the introduction of the intervention (this is called the pre-test) as well as during a period of time that follows on the introduction of the intervention (this called the post-test), and this should be done both for the treatment group and for the reference group. The change found between the pre-test and the post-test in the treatment group – compared to the change found between the pre-test and the post-test in the reference group – then gives us the direction and the size of the effect of the intervention under investigation.

At first reading, this may all sound rather abstract, so we will now illustrate the use of a randomized controlled trial in the evaluation of the effect of an intervention on drink driving. Suppose we are interested in the effect of increasing penalties and fines for alcohol-impaired driving on the number of fatal crashes in a country. In the ideal study design, we would randomly select a number of regions from the country, 60 let's say, and then randomly assign 30 regions to the treatment group and 30 regions to the reference group. We would register the number of fatal crashes in the 60 regions during one year before the increase in penalties and fines for alcohol-impaired driving, then increase the penalties and fines for alcohol-impaired driving in the 30 regions belonging to the treatment group only, and finally register the number of fatal crashes in the 60 regions during one year after the increase in penalties and fines for alcohol-impaired driving.

We would then use a repeated measures ANOVA (Analysis of Variance) procedure to investigate whether the change in fatal crashes between pre-test and post-test in the 30 regions of the treatment group was different from the change in fatal crashes between pre-test and post-test in the 30 regions of the reference group (see, for example, Kirk (2012), and many other statistical text books for details on the analysis of variance for experimental designs). Suppose that the mean number of fatal crashes for the treatment group and reference group in the pre-test and post-test are those displayed in *Figure I-1*. Here we see that the number of fatal crashes in the treatment group has on average decreased between pre-test and post-test, while it has remained more or less the same in the reference group. Such an effect is known as an *interaction effect*.

ANNEX I

If the statistical test for the interaction between time (pre-test and post-test) and group (treatment versus control) is found to be significant, we would be in the position to conclude that the introduction of higher penalties and fines for alcohol-impaired driving had caused the reduction in fatal crashes, and that this effect not only applies to the 60 specific regions involved in the study, but to the country as a whole as well.

In practice, of course, it would be very difficult if not impossible to perform such a study in the way that we just described. It would probably be unacceptable, for example, to drivers in the treatment regions to have to pay higher fines and to get heavier penalties for drink driving than drivers in other regions of the country.

We will next present three alternative experimental designs, and discuss the implications these have for the conclusion of the evaluation of an intervention. These are designs without a pre-test, designs without a reference group, and designs lacking both a pre-test and a reference group.

1.1. No pre-test

If there is no pre-test, the registration of the number of fatal crashes in the treatment and reference group only starts after the introduction of the intervention. It is now still possible to test with an ANOVA whether there is a significant difference between the number of fatal crashes in the treatment group and in the reference group after the intervention has been introduced. Should this be the case, however, then we are only allowed to conclude that this was the result of the intervention if the objects in the study were randomly assigned to the treatment group and the reference group. Without random assignment, an alternative explanation for the observed difference would be that it was the result of systematic differences between the treatment and reference groups also related to the number of fatal crashes. Perhaps the regions in the reference group happened to have much larger amounts of traffic than the regions in the treatment group. Or maybe the drivers in the regions ending up in the treatment group happened to have higher incomes than those in the reference group, thus allowing them to buy better and more protective cars, etc.

This is why random assignment of objects to the treatment group and the reference group is so important: it guarantees that characteristics like differences in the amount of traffic in a region and the income of the drivers are evenly distributed over the two groups. Such characteristics that, if not appropriately handled, affect the outcome of an intervention study are called confounding variables. By far the best way to avoid such variables affecting the outcome of a study is to exert experimental control by using random assignment. However, should it not be possible to assign the objects randomly to the treatment group and the reference group, and should we happen to have data on such confounding variables, then the next best thing we can do is to apply statistical control. Suppose, for example, that we have data on the amount of traffic circulating in the 60 regions, and it was not possible to assign the regions randomly to the treatment group and the reference group, then we could at least statistically correct for possible systematic differences in the amount of traffic in the 60 regions by performing an analysis of covariance (ANCOVA) where the amount of traffic is handled as a covariate in the analysis. We again refer to Kirk (2012) and many other statistical text books for details on the analysis of covariance.

1.2. No reference group

In this case we would typically select a number of regions, register the number of fatal crashes during a year before the introduction of the intervention, then apply the intervention to all of these regions, and again register the number of fatal

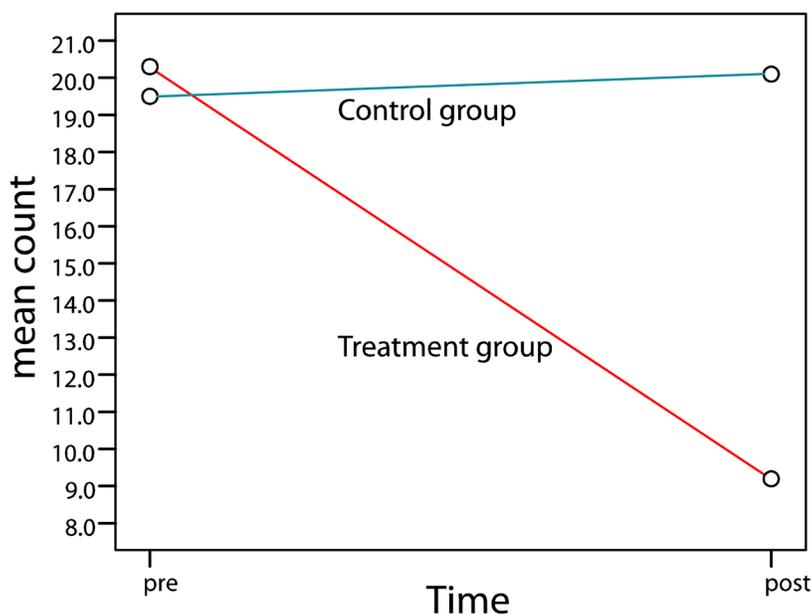


Figure I-1. Mean numbers of fatal crashes in a randomized controlled trial.

ANNEX I

crashes during a year after the introduction of the intervention. So, we have no reference group. In this situation, it is again possible to statistically test with ANOVA whether there is a significant difference between the number of crashes in the pre-test and in the post-test periods.

Should we find a significant difference, however, then we simply cannot conclude from this intervention study that the intervention caused the change in the number of fatal crashes, because these are the possible alternative explanations for the observed change:

1. There was a general decreasing trend in the number of fatal crashes.
2. There is statistical regression-to-the-mean. Interventions are often applied to the most dangerous locations or regions. In that case, the number of crashes in the post-test will automatically tend to shift in the direction of the average number of crashes in all regions due to chance alone.
3. There has been a change in the registration of the number of crashes.
4. Other external events occurred at the same time as the introduction of the intervention that are also responsible for the effect.
5. A combination of all of these.

In this design, which is also called a simple before-after design, the general problem is that we have no idea what would have happened had the intervention not been introduced. The assumption that nothing would have changed between the pre-test and post-test periods had the intervention not been introduced is rather hard to defend.

This again points out the importance of a reference group in the study design, and of the random assignment of objects to the two groups. Should there be a decreasing (or increasing) trend, statistical regression-to-the-mean, a change in registration, other external events, or a combination of all four, then the random assignment will guarantee that the objects in the treatment group and the reference group will be equally influenced by all these factors between the pre-test and the post-test periods.

1.3. *No pre-test nor a reference group*

The worst case scenario is the situation where we only have data for the treatment group after the introduction of the intervention. Unfortunately, in this case nothing can be done to evaluate the effect of an intervention.

2. *Practice – observational studies*

When evaluating the effects of interventions on road safety, it is often not possible for ethical and/or practical reasons to randomly select research objects from a population and/or to randomly allocate these objects to a treatment group and reference group. As an extreme example, consider the research question whether drivers are more in danger of becoming involved in a crash when being under the influence of alcohol. In this situation, it is not possible to draw a random sample of drivers and assign them randomly to a treatment group and a reference group, and then let the drivers in the treatment group circulate in traffic after drinking alcohol. More generally, it is of course not possible for ethical reasons to randomly assign people to circumstances that may result in human suffering or danger to human lives.

Most research into the effects of interventions in road safety is therefore not based on experimental designs, but is purely observational. The effects of the introduction of an intervention are simply being monitored without any experimental manipulation. In contrast with the ideal study design, it is then no longer possible to conclude that the intervention caused an effect.

Even in these situations, however, the challenge is to try to emulate the ideal study design as best as we can by:

1. Trying to establish what the situation in the post-test period would have been had the intervention not been introduced;
2. Controlling statistically instead of experimentally for confounding variables;
3. Comparing the (sample from the) target population with a (sample from a) reference population that is not expected to be influenced by the intervention, but is as similar as possible to the target population in all other respects.

We will now discuss two methods for the analysis of data obtained in non-experimental, observational studies: time series analysis and the Empirical Bayes method.

ANNEX I

3. Time series analysis

Generally, time series analysis is concerned with the description, explanation, and forecasting of developments over time. When these developments have been quantified in terms of annual or monthly counts of accidents, fatalities, seriously injured, etc., then we need special statistical techniques to properly analyze these time series data. The reason why we need special techniques is that time series are a very special type of data. They are repeated and sequential measurements over time of one and the same phenomenon. An important property of such data is that they are usually not independent of one another: the observed number of fatal crashes for the last year in a country, for example, is often a pretty good indicator for the number of fatal crashes to be observed in the same country this year.

Standard analysis techniques typically assume that the observations are independent. This assumption is crucial for the correctness of statistical tests and of confidence intervals. Dedicated time series techniques, on the other hand, are well equipped to handle the dependencies in time series data. Here we will illustrate the analysis of time series with structural time series models, also known as state space models (Harvey, 1989; Commandeur and Koopman, 2007; Durbin and Koopman, 2012). Older alternatives for the analysis of time series are the ARIMA family of models developed by Box and Jenkins (see Box et al., 2008). Unlike structural time series models, ARIMA models usually require that time series are pre-processed before the actual analysis can start, and we therefore do not discuss them here.

Besides their application for simple descriptive time series analyses, structural time series models can also be used to:

1. Evaluate the effects of interventions and of other explanatory variables on the development of road safety;
2. Investigate whether recently published road safety statistics are or are not in line with what could be expected based on the past;
3. Predict future developments in road safety, based on the past.

We will now explain some important methodological aspects of the evaluation of the impact of interventions with time series analysis using the following observational study as an illustration (see Harvey and Durbin, 1986). In the United Kingdom, car drivers and passengers sitting in the front seat were obligated to wear a seatbelt starting on January 31, 1983. The percentage of people wearing a seatbelt was 40% in December 1982; in February 1983 this percentage had risen to 90% and it stabilized at approximately 95% in March 1983. The wearing of a seatbelt is only supposed to affect the number of victims of car crashes and their severity, not the number of car crashes themselves. Possible confounding variables that also influence the severity of injury of victims of car crashes are speed, type of car, type of crash, and the number of passengers.

Restricting ourselves to car drivers for the moment, the monthly numbers of car drivers killed or seriously injured (KSI) for the years 1969 through 1984 are those given in *Figure I-2*.

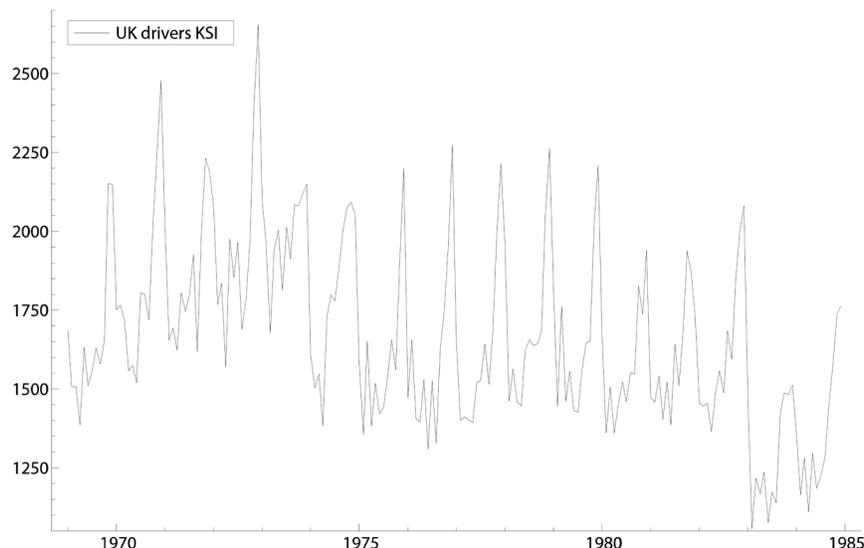


Figure I-2. Monthly numbers of car drivers killed and seriously injured in the United Kingdom in the period 1969-1984.

ANNEX I

The most simple but also the most naive evaluation of the effect of the mandatory use of the seatbelt would be to add up the monthly number of drivers KSI in the year before the introduction of the seatbelt law (the pre-test), for example, and then to compare this sum with the sum of the monthly data in the year after the introduction of the seatbelt law (the post-test). Adding the monthly data of February 1982 through January 1983, and also adding the monthly data of February 1983 through January 1984, yields the numbers given in *Table I-1*.

Table I-1. Total number of UK drivers KSI in the year before and in the year after the introduction of the seatbelt law in the United Kingdom in February 1983.

1982(2)-1983(1)	1983(2)-1984(1)	TOTAL
19498	15335	34833

The F-test for the difference between these two frequencies is:

$$F = \frac{N_1}{N_2 + 1} = \frac{19498}{15335 + 1} = 1.27$$

with $2(N_2 + 1) = 30672$ degrees of freedom in the nominator, and $2N_1 = 38996$ degrees of freedom in the denominator (see Kanji, 1993, p.51).

This test is very significant ($p < 0.01$), which means that the null-hypothesis of equal frequencies in the pre-test and the post-test periods can be rejected. We would then conclude that the introduction of the seatbelt law is associated with a $100((19498-15335)/19498)=21.35\%$ decrease in the number of car drivers KSI.

Alternatively, we could add all monthly data before the introduction of the seatbelt law (there are 169 of them), also add up all monthly data after the introduction of the seatbelt law (there are 23 of them), and then test the difference between these two sums while correcting for the different time periods to which the two sums apply:

$$F = \frac{\frac{1}{t_1}(N_1 + 0.5)}{\frac{1}{t_2}(N_2 + 0.5)} = \frac{\frac{1}{169}(290308 + 0.5)}{\frac{1}{23}(30339 + 0.5)} = 1.30$$

with $(2N_1 + 1)$ degrees of freedom in the nominator, and $(2N_2 + 1)$ degrees of freedom in the denominator (Kanji, 1993, p.51).

This test is also very significant ($p < 0.01$), indicating a reduction of 23% in the number of UK drivers KSI.

The question is now: is this a convincing result? All the objections that we mentioned in *Section 1* apply to this naïve approach in evaluating the impact of an intervention: the just mentioned reduction could as well have been caused due to a general decreasing trend in the data, due to statistical regression-to-the-mean, due to changes in the registration of drivers killed and seriously injured in crashes, or due to other external events than the introduction of the seatbelt law occurring around February 1983 and also affecting the number of drivers KSI.

Applying time series analysis to the data in *Figure 2* using state space models, also known as structural time series models (see Harvey, 1989; Commandeur and Koopman, 2007; Durbin and Koopman, 2012), opens up the possibility to estimate what the development in the number of UK drivers KSI would have been had the seatbelt law not been introduced in February 1983. This is the subject of *Section 1.3.1*. Furthermore, with time series analysis we can also evaluate the impact of an intervention corrected for a general decreasing or increasing trend, for monthly variations and for confounding variables about which information is available (see *Section 1.3.2*). Finally, in *Section 1.3.3* we discuss how to use something resembling and akin to a reference group in a randomized controlled trial in order to implicitly correct for possible confounding variables about which no information is available.

ANNEX I

3.1. Forecasting developments had the intervention not been introduced

In order to obtain an estimate of what would have happened if the introduction of the seatbelt law had not been introduced, we first apply a time series analysis to the UK drivers KSI series restricting ourselves to the monthly data in *Figure 2* corresponding to January 1969 up to and including January 1983, the 169 observations just before the introduction of the seatbelt law. This is done using a state space model including a trend in order to capture the general long term change in the series, and a seasonal pattern in order to capture the monthly variation in the data. This model is found to yield a good representation of the observed development up to and including January 1983. For the technical details of such state space models we refer to the earlier mentioned literature. The results of this analysis are shown in *Figure I-3*.

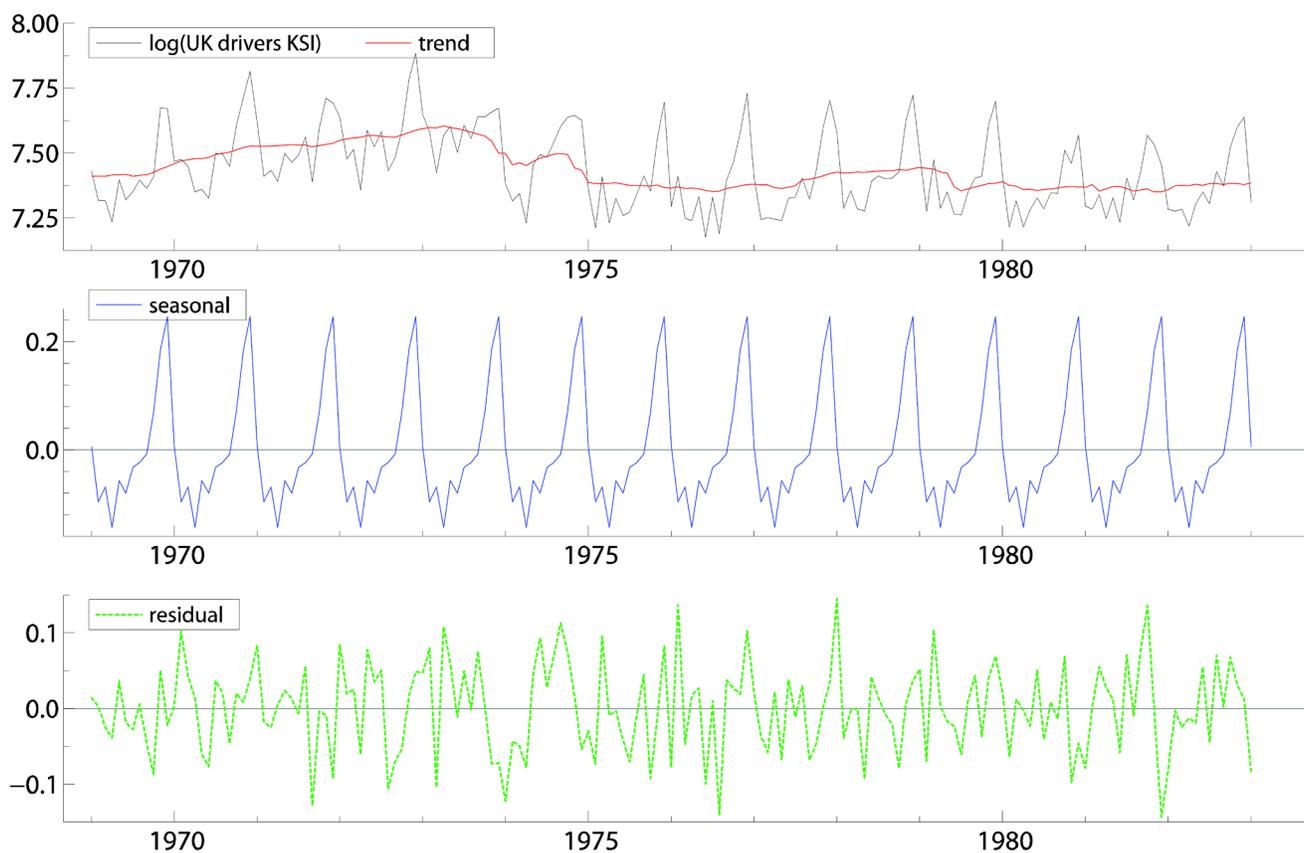


Figure I-3. Results of the time series analysis of the logarithm of the number of UK drivers KSI in the period 1969 up to and including January 1983.

The top graph in *Figure I-3* displays the trend in the series (including the observations), and the middle graph shows the seasonal pattern. Zooming in on the seasonal pattern reveals that the highest number of drivers KSI is always observed in the months of November and December, while the month of April always results in the smallest number of drivers KSI. The bottom graph in *Figure I-3* contains the residuals of the model: these are the parts of each observation that are not accounted for by the model.

By projecting the trend and the seasonal pattern in *Figure I-3* into the future, we obtain forecasts for the 23 time points of February 1983 through December 1984. We can then compare these forecasts with the observed number of victims after the introduction of the seatbelt law. The result is shown in *Figure I-4*.

ANNEX I

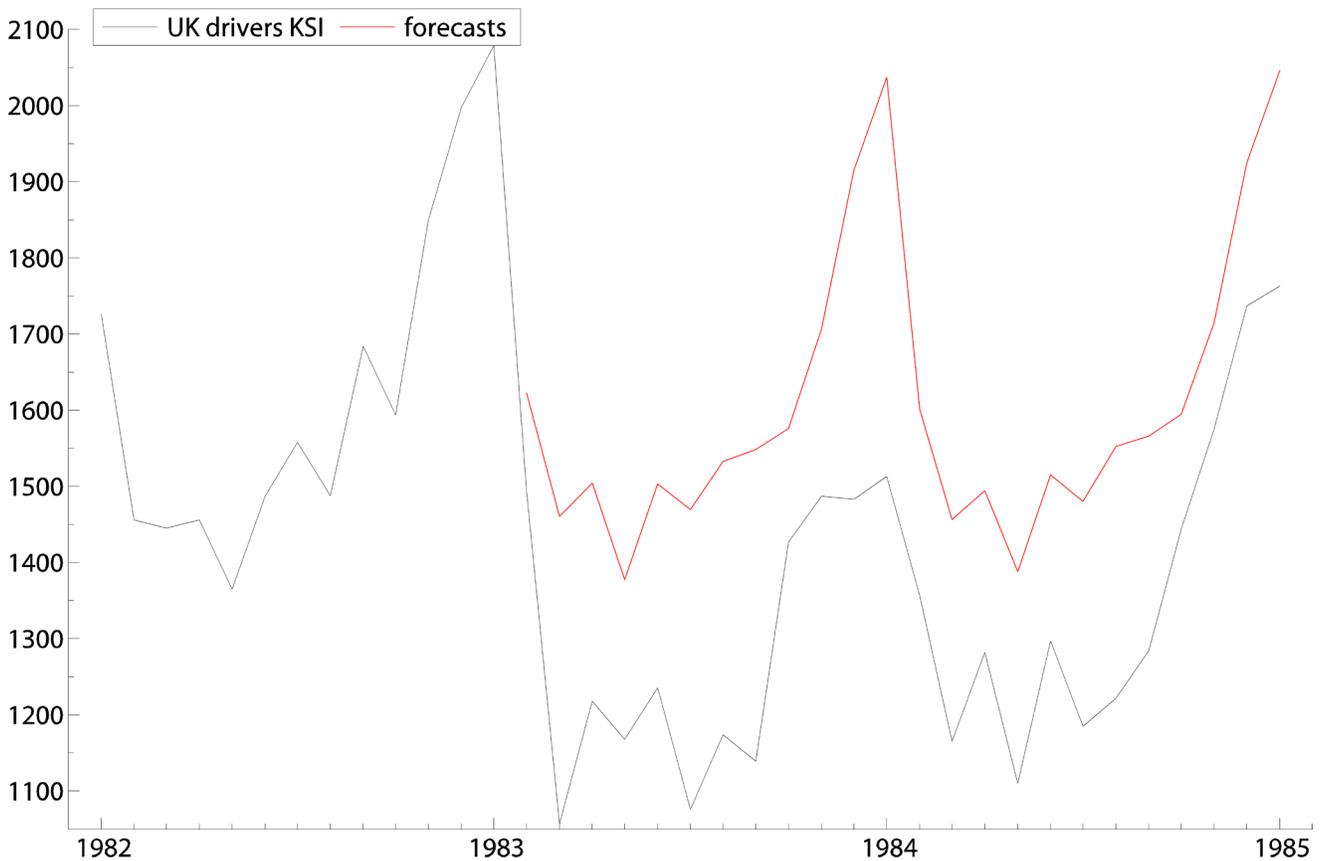


Figure I-4. Forecasts of the number of drivers KSI after February 1983 (red line), only using the model estimated for the period before the introduction of the intervention. The actual observations are indicated by the black line.

As the Figure indicates, the observed number of drivers KSI in the period February 1983 through December 1984 (black line) is clearly smaller than the numbers we would have expected based on the past up to and including January 1983 (red line).

The total numbers of observed and expected (i.e., forecasted) drivers KSI in the period February 1983 through December 1984 are given in Table I-2.

Table I-2. Total expected and observed number of drivers KSI for the period February 1983 through December 1984.

FORECASTS	OBSERVATIONS
36970	15335

ANNEX I

The F-test for the difference between the frequencies in *Table I-2* equals:

$$F = \frac{N_1}{N_2 + 1} = \frac{36970}{15335 + 1} = 2.41,$$

with $2(N_2 + 1) = 30672$ degrees of freedom in the nominator, en $2N_1 = 73940$ degrees of freedom in the denominator, and is very significant ($p < 0.01$).

This is a first indication that an important change occurred in the development of the number of drivers KSI in February 1983.

3.2. Intervention analysis including confounding variables

In the next step, we analyze all 192 observations in the series of *Figure I-2*, adding an intervention variable to the structural time series model. Intervention variables generally allow us to model temporary and structural changes in a time series. In principle, interventions may have several different types of effects, some of which are displayed in *Figure I-5*. An intervention may have the effect that the number of crashes or victims changes only very temporarily after which the series immediately returns to its previous level. This is called a pulse. An intervention can also result in a sudden and permanent reduction, or in a gradual and permanent reduction in the number of crashes or victims. These are called level breaks. An intervention may finally also have the effect that the increase or the decrease in the number of crashes or victims suddenly and permanently becomes more or less pronounced. This is known as a *slope break*.

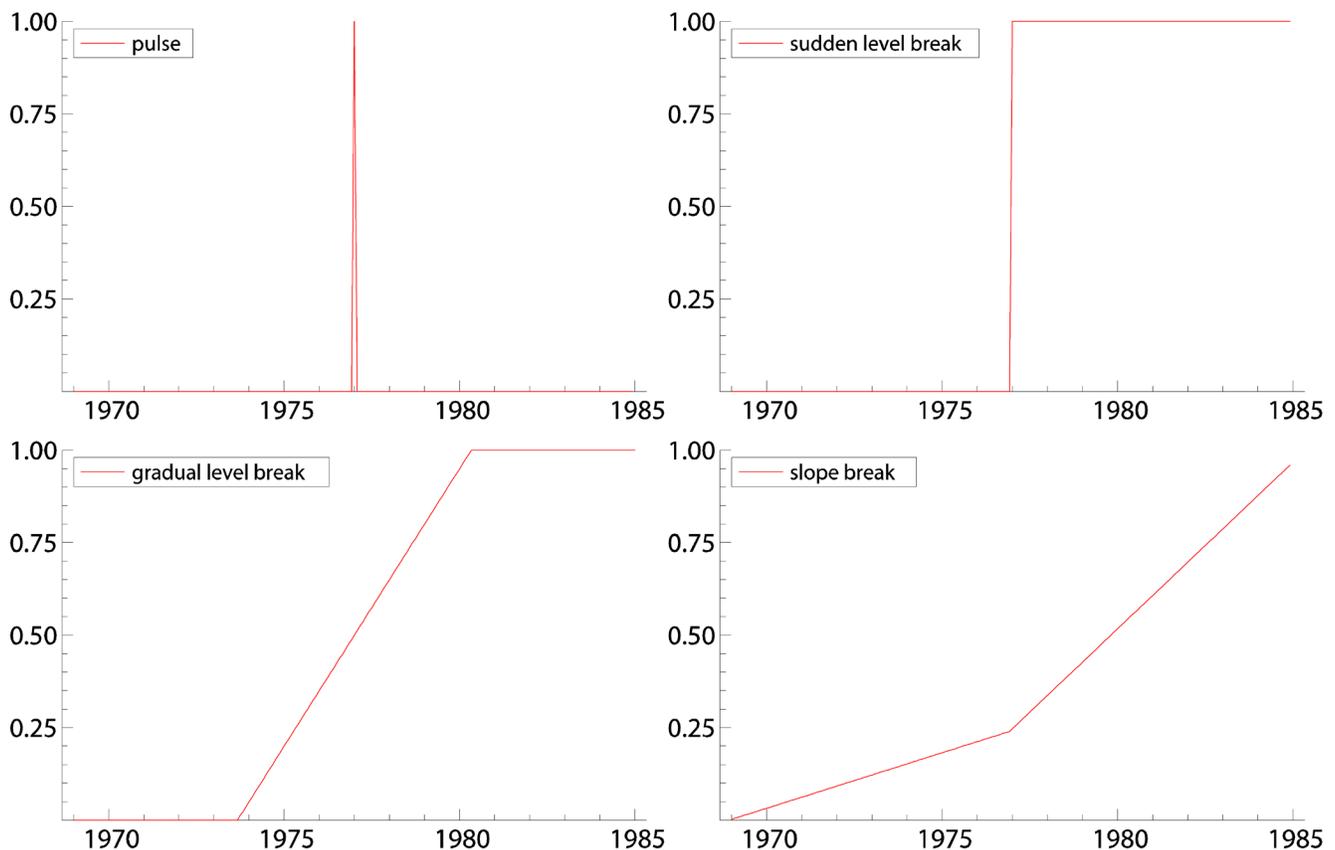


Figure I-5. Four types of intervention.

ANNEX I

In order to evaluate the effect of the seatbelt law, we opt for a sudden and permanent change and add a level break variable to the time series model. Such a variable typically contains all zeroes in the first 169 months of the series (before February 1983) and all ones in the last 23 months of the series (at and after February 1983). We further also add the logarithm of the monthly price of petrol in the period January 1969 through December 1984. The idea is that we may consider the price of petrol to be a proxy for the mobility, thus allowing us to evaluate the effect of the seatbelt law after correcting for changes in mobility, which can be considered as always a potential confounding variable in observational intervention studies. The latter procedure is a typical example of applying statistical control instead of experimental control.

The graphical result of this analysis is shown in *Figure I-6*.

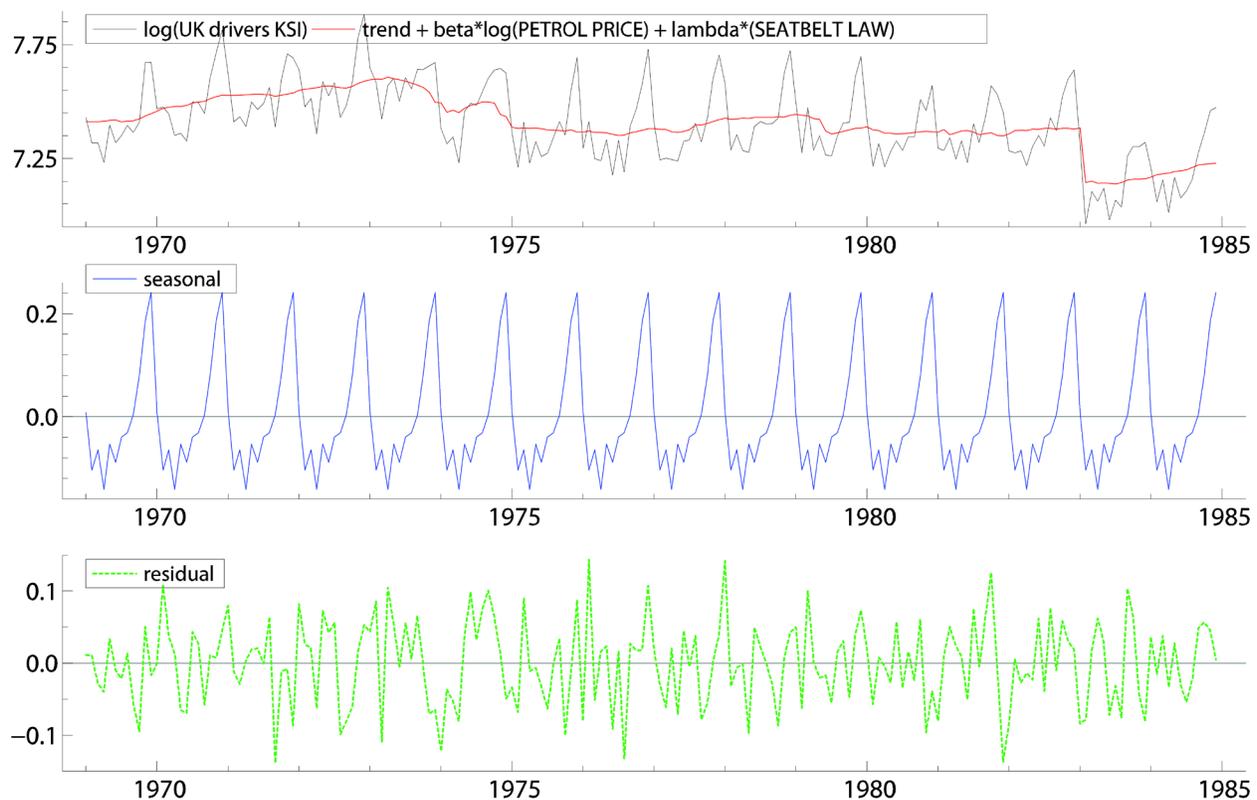


Figure I-6. Results of the time series analysis of the logarithm of the number of drivers KSI in the period January 1969 through December 1984, including a level break intervention variable (SEATBELT LAW) and the logarithm of petrol price.

The regression coefficient for the logarithm of the petrol price (called “beta” in the legend of the top graph in *Figure I-6*) is found to be -0.2767, and significantly deviates from zero. This implies that a 1% increase in petrol price in the period 1969-1984 is associated with a 0.28% decrease in the number of UK drivers KSI. The regression coefficient for the intervention variable equals -0.2376 and is also significant. This means that the introduction of the seatbelt law in February 1983 is associated with a sudden $100(\exp(-0.2376) - 1) = -21.1\%$ change in the number of UK drivers KSI. This sudden drop in February 1983 is clearly visible in the top graph of *Figure I-6*.

All the previous results are graphically summarized in *Figure I-7*. In *Figure I-7* we not only see that the model forecasts based on the data up to February 1983 reflects a very different development than what was actually observed (as was already made clear in *Figure I-4*), but also that the addition of an intervention variable to the time series model applied to the complete series results in a good representation of the observed developments both before and after February 1983. All these results make the conclusion more likely and convincing that the introduction of the seatbelt law indeed had an effect on the number of UK drivers KSI, and was associated with an approximate 21% decrease.

ANNEX I

We certainly have not proven that this reduction was caused by the seatbelt law. But we have at least tried to establish what would have happened had the law not been implemented, and have been able to establish the magnitude of its effect after correcting for seasonal effects, for fluctuations in the petrol price, and for a possible general decreasing trend in the development of the number of UK drivers KSI.

Finally, by using a long series of 169 observations before the introduction of the intervention we also have shown that the observed reduction could not have been the result of a coincidental very high number of drivers KSI in the pre-test period, but that these numbers reflected the actual and structural situation of the group of road users under study. This makes it very unlikely that statistical regression-to-the-mean was (even only partly) responsible for the estimated effect.

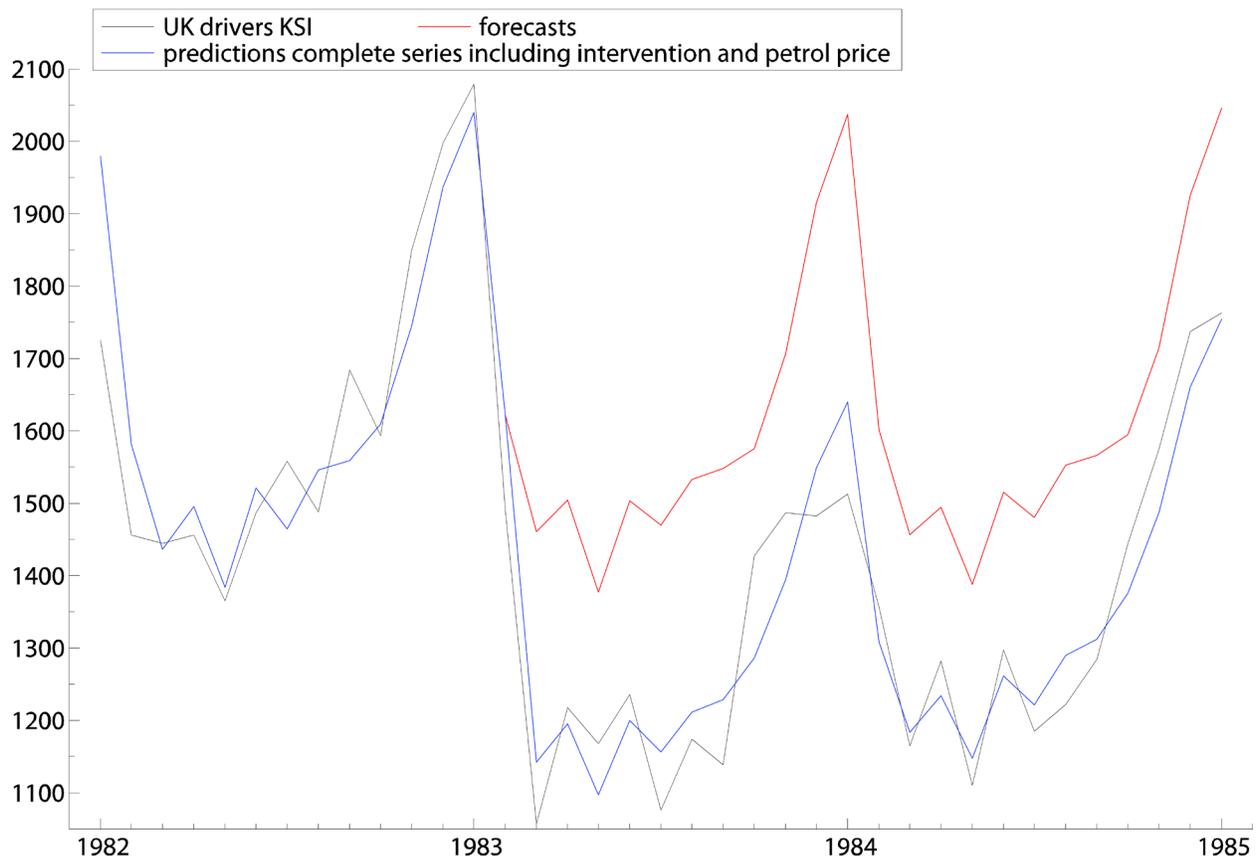


Figure I-7. The red line denotes the forecasts after the introduction of the seatbelt law in February 1983 obtained from the time series analysis only using the observations of January 1983 through January 1983. The blue line denotes the predictions using the complete series including the intervention variable for the introduction of the seatbelt law and the logarithm of petrol price. The observed number of UK drivers KSI is displayed as a black line.

3.3. Intervention analysis with a reference group

The evaluation of the effect of an intervention can be made even more convincing if we can show that the intervention had an effect on the target population, but not on an alternative population that is as similar as possible to the target population, except that the intervention under study is not supposed to apply to the members of that alternative population. If there is such an alternative population, and if we have road safety data, then this allows us to approximate the ideal study design even further by using them as a reference group. In order to avoid confusion with a reference group obtained by random allocation of the objects of the study (see Section 1) we will denote such a group in observational studies by the term reference group.

ANNEX I

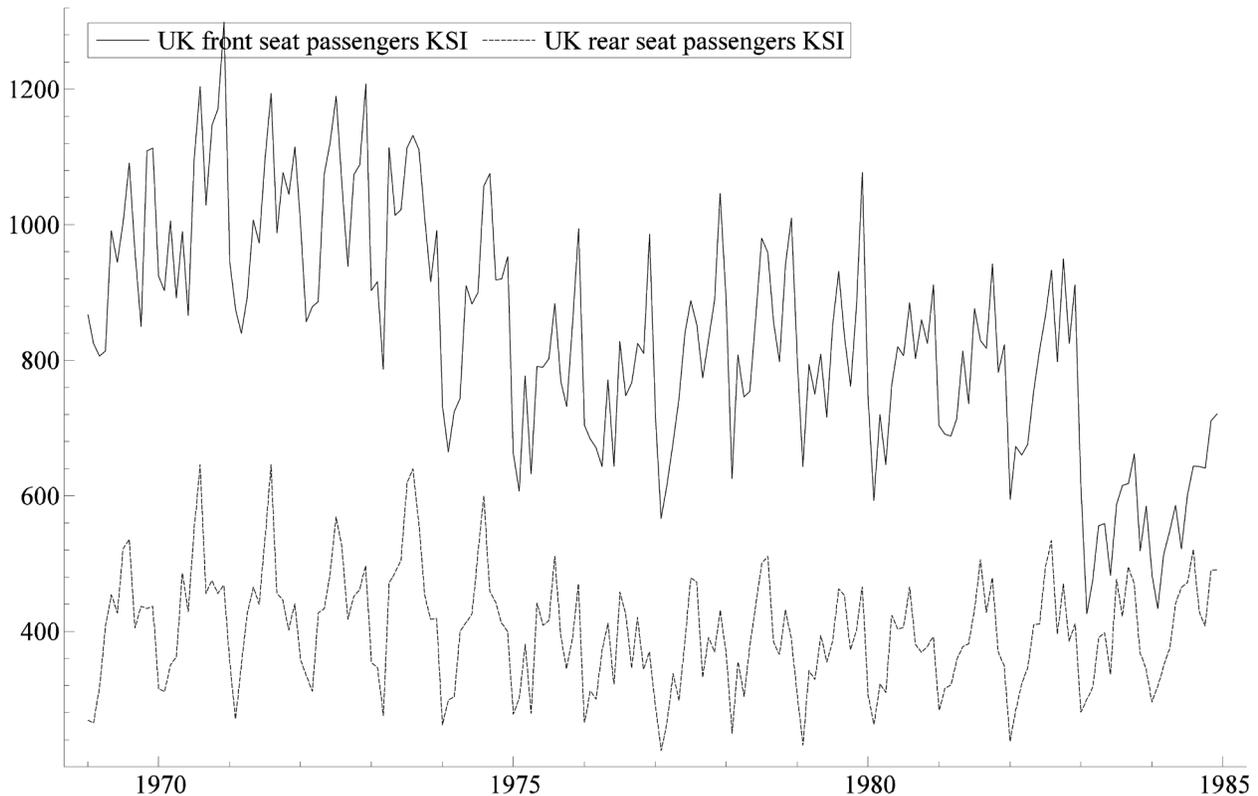


Figure I-8. Monthly observations of the number of UK front seat passengers KSI and the number of UK rear seat passengers KSI in the period January 1969 through December 1984.

The introduction of the seatbelt law in February 1983 in the United Kingdom only applied to drivers and front seat passengers, not to rear seat passengers. There is not only information available on the monthly number of drivers KSI, but also on the monthly number of front and rear seat passengers KSI. The latter two time series are shown in *Figure I-8*. With such data it becomes possible to emulate something akin to or resembling a quasi-experimental design. First of all we may now investigate the frequencies at pre-test and post-test in the treatment group and in the reference group as displayed in *Table I-3*.

Table I-3. Total numbers of front and rear seat passengers KSI in the year before and after the introduction of the seatbelt law.

	1982(2)-1983(1)	1983(2)-1984(1)	TOTAL
Front seat passengers	9482	6568	16050
Rear seat passengers	4749	4618	9367
Total	14231	11186	25417

ANNEX I

The Chi-square of this contingency table is 168.5, indicating that there is a very significant interaction between time (pre-test and post-test) and group (front and rear seat passengers KSI) ($p < 0.01$). Assuming that the change in the number of front seat passengers KSI would have been identical to the change in the number of rear seat passengers KSI had the seatbelt law not been introduced, then we would expect $9482 (4618/4749) = 9220$ front seat passengers KSI at post-test. This would suggest a $100[(6568-9220)/6568] = -40\%$ reduction in the number of front seat passengers KSI as a result of the introduction of the seatbelt law. The objection to this analysis is that we did not correct for possible trends in the treatment and reference groups, neither were we able to correct for any confounding variables about which information may be available.

However, both the latter two requirements can be fulfilled by performing a bivariate time series analysis of the two series depicted in *Figure I-8*. Such an analysis where we add both the logarithm of petrol price and a level break intervention variable for the introduction of the seatbelt law as explanatory variables to both series, we obtain the results shown in *Table I-4* which contains the estimated regression coefficients for all four explanatory variables and their significance tests.

Table I-4. Effects of log (petrol price) and of the intervention variable "seatbelt law" on the number of front and rear seat passengers KSI.

	1982(2)-1983(1)	1983(2)-1984(1)	TOTAL	
Front seat passengers				
log (petrol price)	-0.3071	0.1067	-2.8791	0.0045
Seatbelt law	-0.3372	0.0495	-6.8157	1.2e-010
Rear seat passengers				
log (petrol price)	-0.0836	0.1123	-0.7443	0.4576
Seatbelt law	0.0021	0.0518	0.0405	0.9677

These results indicate a very significant $100(\exp(-0.3372) - 1) = -28.6\%$ reduction in the number of front seat passengers KSI associated with the introduction of the seatbelt law, while the effect of the intervention is completely non-significant for the rear seat passengers KSI. Since there is no evidence that the reference group was affected by the law, this makes the estimated reduction in front seat passengers KSI of 28.6% after correction for petrol price even more convincing. For more details on how to apply multivariate times series models with state space methods we refer to Durbin and Koopman (2012) and Commandeur and Koopman (2007).

ANNEX I

4. *The Empirical Bayes analysis method*

The Empirical Bayes (EB) analysis method was developed by Hauer (1997) to evaluate the effect of interventions in observational before-after studies. A central feature of the EB method is the understanding that objects subjected to an intervention in observational studies often are not a random sample from the target population, but have been selected for the intervention because they are relatively very unsafe (in terms of crashes and/or victims). As a result of the fact that there is then a clear relation between the number of crashes and/or victims observed in the study objects and the reason for selecting them for the intervention, we have what is called selection bias which may well result in the regression-to-the-mean effect that we already discussed in [Part 1](#).

Furthermore, Hauer also emphasizes that in a naive before-after study, the observed change (if any) need not necessarily be the result of the intervention under study, but also of confounding variables like changes in traffic intensity, the weather, the registration of the number of crashes, etc. In short, and as already mentioned before, in a naive before-after study we have no idea what would have happened had the intervention not been introduced.

Supposing that the indicator used in the study is the number of fatalities, for example, the EB method for the evaluation of the effect of an intervention in an observational before-after study generally consists of the following steps:

- Observe the number of fatalities for the objects to which the intervention is applied in the pre-test and in the post-test periods;
- Determine the expected number of fatalities for these same objects in the pre-test period by determining the number of fatalities and its variance in a reference population that is as similar as possible to the objects to which the intervention has been applied;
- Correct the number of observed fatalities in the pre-test period based on this expected number of fatalities;
- Compare the latter corrected number of fatalities in the pre-test with the observed number of fatalities in the post-test and use this difference to quantify the effect of the intervention.

The EB method thus attempts to correct for the regression-to-the-mean effect when it is suspected that there is a relation between the intervention under study and the reason why the intervention was applied to specifically selected objects and not to others. Obviously, the EB method can only be applied if we have at our disposal estimates of the expected number of crashes and/or victims obtained in another previous study of the relation between the number of crashes and/or victims and relevant characteristics of the objects in a reference population. For further details on the EB method we refer to Hauer (1997).

5. *Conclusions*

Obtaining hard evidence for the effect of an intervention in observational studies is clearly not a trivial matter, especially when it comes to establishing whether the intervention actually caused an effect. Still, we hope to have shown that the evaluation of the effect of an intervention in an observational setting can be made considerably more convincing by, whenever possible:

1. Using longer time series of observations in the pre-test period in order to be able to correct for general trends, seasonal patterns, and regression-to-the-mean effects;
2. Correcting explicitly for confounding variables that are known and have been measured in the pre-test and post-test periods by including them as covariates in the intervention analysis;
3. Correcting implicitly for confounding variables that are unknown and/or have not been measured by using information on a reference group or population in the analysis;
4. Applying different types of analysis in order to cross-validate the estimated effects of the intervention.

ANNEX I: REFERENCES

1. Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (2008). *Time Series Analysis: Forecasting and Control*. Fourth Edition. Wiley.
2. Commandeur, J.J.F. and Koopman, S.J. (2007). *An introduction to state space time series analysis*. Oxford: Oxford University Press.
3. Durbin, J. and Koopman, S.J. (2012). *Times series analysis by state space methods*. Oxford: Oxford University Press.
4. Hauer (1997). *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Pergamon, New York.
5. Kanji, G.K. (1993). *100 statistical tests*. London: Sage Publications.
6. Kirk, R.E. (2012). *Experimental Design: Procedures for the Behavioral Sciences*. Fourth Edition. London: Sage Publications.

ANNEX II

ANNEX II. Example of roadside survey questionnaire

During a roadside survey, drivers are stopped by the police and breath tested to assess their alcohol consumption. For more information about the personal characteristics of the drivers in relation to their alcohol use in traffic, several questions can be asked. Below you will find a short and simple standard questionnaire to be used in a roadside survey. For information on how to formulate questions, see also [Annex IV](#) on how to design a standard questionnaire survey on drinking and driving.

General information:

ID number: _____ Survey location: _____
Date: _____ Time: _____

Standard questions for each breath tested driver:

1. Gender Male Female
2. Age (in years) _____
3. Type of vehicle
 - Minibus
 - Person car/van
 - Tanker/truck
 - Bus
 - Motorcycle
 - Other (please specify) _____
4. Number of passengers
5. Breath alcohol concentration (BAC) _____

A standard form can be developed to register the information, for example:

ID-NR	GENDER 1 = MALE 2 = FEMALE	AGE (IN YEARS)	MINIBUS PERSON CAR/VAN TANKER/TRUCK BUS MOTORCYCLE OTHER	NUMBER OF PASSENGERS	BREATH ALCOHOL CONCENTRATION (READ FROM BREATHALYSER)
1					
2					
3					
4					
5					
6					
7					

ANNEX II

General additional questions

Besides the standard questions listed above, it may be interesting to collect additional information concerning:

- Trip distance (This information can best be asked by providing some answer options: 0-10 miles, 10-20 miles, 20-50 miles, 50 miles or more)
- Being a professional driver or not
- Being a novice driver or not
- Experiences with police checks on drinking and driving
- Knowledge and attitudes concerning drinking and driving
- Education level
- Occupation

However, it is recommended asking only a subgroup of all drivers these additional questions. For example, a random selection of passing drivers. Asking all drivers would probably take up too much time.

Additional questions for violators

Furthermore, information from drink driving offenders on the following items may be interesting:

- Place and time of last beverage
- Occasion for drinking
- Information on recidivism (yes or no)

ANNEX III

ANNEX III. A work plan for a mail survey

Source: http://www.aoa.acl.gov/Program_Results/POMP/Chapter5.aspx.

Mail survey work plan:

1. Plan
 - Determine purpose of the survey ○
 - Select appropriate survey instrument
 - Determine the methodology
 - Design data collection procedures
 - Develop cost estimate
 - Determine staff needs
2. Sample
 - Identify population
 - Specify sample selection criteria
 - Identify sampling procedure
 - Identify source of information for sampling
 - Draw sample
3. Pilot
 - Test procedures and revise according to the results of testing
4. Survey
 - Notify selected agencies (if sample is drawn from another agency)
 - Prepare materials: cover letter, copies of survey instruments, envelopes for mailing and returning survey instruments
 - Mail survey materials
 - Enter data into an electronic data entry utility as completed instruments are returned
 - Contact non-respondents at designated time periods
5. Prepare Data
 - Enter remaining data
 - Check for accuracy of data entry
 - Review database for outliers and anomalies in the data
 - Code responses to open-ended questions
6. Analyze Data
 - Run descriptive statistics
 - Analyze responses to open-ended questions
7. Prepare Report
 - Prepare draft report
 - Submit report to reviewers
 - Revise report
8. Disseminate Results
 - Develop plan for sharing results
 - Disseminate report

ANNEX IV

ANNEX IV. Formulating questions about drink driving

By Dr. Charles Goldenbeld

This annex describes the process of composing a good questionnaire and provides examples of survey questions on drinking and driving.

1. Setting up a questionnaire and defining its scope

There is no standard way to ask questions on drinking and driving. The composition of a questionnaire depends upon:

- Specific aims of the evaluation;
- Chosen subjects;
- Characteristics of your target population.

Composing a questionnaire is a stepwise process:

- First, decide on which topics you want to include in the questionnaire;
- Then decide on the sequence of topics in the questionnaire;
- After that, decide on how you want to formulate specific questions per topic;
- After having formulated one or more example questions per topic, decide on the specific answer scales;
- After you have a first concept questionnaire, organize critical feedback on the questionnaire from experts (e.g., psychologists, marketing specialists, social science researchers, opinion researchers);
- After a revision of the questionnaire, organize a pilot or try-out of the questionnaire among respondents of your target population.

In a pilot questionnaire, the try-out of the questionnaire should provide information on the following:

- Do respondents understand the questions well or are some questions misunderstood or understood in more than one way?
- Do respondents find that the answers they want to give are not listed among the answer categories?
- Do respondents find questions interesting or boring?
- Do respondents find questions shameful or offensive?
- How much time do the respondents need to fill in the questionnaire?
- Is there a need to clarify certain questions with either some additional verbal explanation or some pictures?

Subjects for the questionnaire (or questionnaire themes) include a number of subjects that seem relevant for many drink driving evaluations:

- Self-reported frequency of drinking and driving
- Places and times of drinking and driving
- Occasions and reasons for drinking and driving
- Circumstances surrounding drinking and driving
- Experiences with police checks on drinking and driving
- Knowledge and attitudes concerning drinking and driving

The first question tells you something about the extent of the problem. The last five questions may help you to better understand the total drinking and driving problem and to give recommendations concerning countermeasures in terms of publicity, enforcement, or behavioral interventions. The next section gives example questions for each of these questionnaire themes.

ANNEX IV

2. Example questions for a drinking and driving questionnaire**2.1. Self-reported frequency of drinking and driving**

An example of questions about the frequency of drink driving is:

- *In the past four weeks, how often have you driven a motor vehicle after drinking a small amount of alcohol?*

Consideration: Asking questions concerning behavior in the past week or the past four weeks decreases the chances of certain memory bias. Of course, the disadvantage is that the reported behavior cannot be automatically generalized to the past year.

The answer scale to this question could be the following:

- Not once
- Once
- Twice
- Three times
- Four times
- Five times
- More than five times

Consideration: This answer scale does not differentiate between drivers who drink alcohol more than five times in four weeks; they all fall within one category (more than five times).

The same question with a better differentiating scale could be the following:

- In the past four weeks, how often did you drive a motorcycle or a car after drinking even a small amount of alcohol?*

- Not once
- Once
- Twice
- Three times
- Four times
- Five to 10 times
- 10 to 20 times
- More than 20

Consideration: This answer scale presupposes that respondents are able to count drinking and driving occasions over a period of four weeks. However, in certain cultures persons may not be inclined to count in such a way. Persons may be more inclined to remember on how many days per week they drink and drive. In that case, the answer scale below may be preferred:

- In the past four weeks, how often have you driven a motorcycle or a car after drinking a small amount of alcohol?*

- Not once in four weeks
- On one day in four weeks
- On several days in four weeks
- On one day every week
- On several days every week
- On most days every week
- Everyday every week

ANNEX IV

Consideration: This answer scale presupposes that respondents are better able to think in terms of number of days per week than in terms of total times over a four week period. However it should be checked in a pilot whether this is the case. In order to identify the heavy drinkers, you could add a question on drinking and driving like the following:

- In the past four weeks, how often have you driven a motor vehicle after drinking at least six glasses containing alcohol?

Or:

- *In the past four weeks, how often have you driven a motor vehicle after drinking at least eight glasses containing alcohol?*

Consideration: In different countries or regions a glass may hold different amounts of alcohol. A large glass may perhaps contain a half a liter or even a liter. The question can be improved if you show pictures of glasses and then ask the respondent to indicate the glass from which he or she drinks. Also, the beverages the respondent drinks may contain different amounts of alcohol. Local beers or wines may vary in alcohol content.

2.2. Places and times of drinking and driving

Example questions are:

- *During the most recent occasion when you drank and drove, where were you when you did most of your drinking?*
 - At home, for example, your house, apartment, condominium, or dorm room
 - At another person's home
 - At a restaurant or banquet hall
 - At a bar or club
 - At a public place, such as at a park, concert, or sporting event
 - At a workplace
 - Other
- *During the most recent occasion when you drank and drove, which day was it?*
 - Monday
 - Tuesday
 - Wednesday
 - Thursday
 - Friday
 - Saturday
 - Sunday
- *During the most recent occasion when you drank and drove, at which time did you drive?*
 - Midday between 12:00-16:00 hrs
 - Afternoon between 16:00-18:00 hrs
 - Early evening 18:00-20:00 hrs
 - Evening 20:00-22:00 hrs
 - Late evening 22:00 – 24:00 hrs
 - Night between 00:00-2:00 hrs
 - Night between 2:00 – 6:00 hrs
 - Early morning between 6:00 – 9:00 hrs

ANNEX IV

2.3. Occasions and reasons for drinking and driving

- *During the most recent occasion when you drank and drove, why did you drink? (More than one answer may be given.)*
 - Boredom, nothing to do
 - Joined friends in drinking/ friends invited me to drink
 - Public festival or event where everybody drank
 - Needed to drink to release tension
 - Needed to drink to overcome tiredness
 - Habit, I drink alcohol (nearly) everyday
 - Other
- *During the most recent occasion when you drank and drove a motor vehicle, why did you decide to use your car? (More than one answer can be given.)*
 - I had to, could not leave car unattended
 - No other transportation available
 - I had to, distance to home too great
 - I had to transport friends/family members with the car
 - Other

2.4. Circumstances surrounding drinking and driving

The circumstances surrounding drinking can tell you about the social nature and social norms concerning drinking. If social norms play an important role, this could have important consequences for the type of countermeasures chosen. An example of a question that attempts to explore the circumstances of drinking and driving is given below.

- *At times when you drink a lot of alcohol before driving, which of the circumstances below apply? (More than one answer can be given; only check the circumstances that are most important for you.)*
 - I am mostly alone when I am drinking a lot
 - All my friends drink
 - The whole village drinks
 - There is free or cheap beer/wine
 - There is a holiday, a festival, or a party
 - There is a weekend or a day off from work
 - All my colleagues at work drink
 - I want to forget problems
 - The drinking is necessary to join a celebration
 - Others expect me to drink
 - Others encourage me to drink
 - I always drink whenever I get a chance

2.5. Experience with police checks

In questions concerning experience with police enforcement, asking for experiences in the past year may increase errors in memory. However, asking about experiences in the past week or past four weeks is a fairly short reference period, given that experiences with police enforcement are not that frequent in many countries. A compromise could be to ask for experiences with police enforcement in the last two or three months.

ANNEX IV

An example question would be:

- *In the past three months, while driving your car or bike, have you been stopped by the police who checked your drinking and driving?*

The answer scale to this question could be the following:

- Not once
- Once
- Twice
- Three times
- Four times
- Five times
- More than five times

Or perhaps a scale such as:

- Not once in three months
- Once in three months
- A few times in three months
- Nearly once every week in the past three months
- Once or twice every week in the past three months
- On most days every week in the past three months

- *On the last occasion when you were stopped and checked for drinking and driving, please indicate the location, day, and time*

Location: road//district/city

Day: Monday/Tuesday/Wednesday/Thursday/Friday/Saturday/Sunday

Time: early/late morning, early/late afternoon, early/late evening, early/late night

- *In the past three months, while being a passenger in somebody else's car, has the car you travelled in been stopped by the police and the driver checked?*

- Not once
- Once
- Twice
- Three times
- Four times
- Five times
- More than five times

- *In the past three months, have you seen other car drivers in traffic being stopped by the police and checked for alcohol?*

- Not once
- Once
- Twice
- Three times
- Four times
- Five times
- More than five times

ANNEX IV

It is interesting to have some information on whether drivers think they are able to outsmart police to avoid drinking and driving checks. Below we list some statements about police enforcement that could be used to elicit this information.

- *How much do you agree or disagree with statements below:*
 - I know in advance when and where police will check on traffic and drinking and driving
 - The police traffic checks can be seen in advance and can be easily avoided
 - It is easy to avoid police traffic checks
 - My friends tell me how to avoid police drinking and driving checks
 - The police do not seriously check for drinking and driving
 - When stopped by the police, I can hide my drinking and driving
 - When I drive after drinking, I choose roads where police are not present

The answer scale to these statements could be a 5-point Likert scale such as

- Very much disagree
- Disagree a little bit
- Neither agree nor disagree
- Agree a little bit
- Very much agree

Alternatively, a 7-point semantic differential scale is also often used. This scale looks as follows

Very much disagree 1 -----2-----3-----4-----5-----6-----7 Very much agree

2.6. Knowledge and attitudes concerning drinking and driving

In order to achieve further understanding of knowledge and attitudes concerning drinking and driving, you could ask a person to state his agreement or disagreement with a number of statements. Below we give examples of these statements:

- A person can be a safe driver even after drinking a large amount of alcohol
- Drinking alcohol impairs the ability of a person to control a car in a safe way
- The penalties for drinking and driving in our country should be made more severe
- In our country, drinking and driving is a major problem for safety on the roads
- Everybody should be free to decide for himself how much alcohol he wants to drink before driving
- In our country, a lot of road accidents are caused by drinking and driving

The answer scale to these statements could again be a 5-point Likert scale

- Very much disagree
- Disagree a little bit
- Neither agree nor disagree
- Agree a little bit
- Very much agree

Or the 7-point semantic differential scale:

Very much disagree 1 -----2-----3-----4-----5-----6-----7 Very much agree

ANNEX IV

Specific examples of knowledge questions could be the following:

- Driving after consuming alcohol increases your accident risk
- Driving after drinking alcohol is prohibited in our country
- In our country, it is prohibited to drive with a blood alcohol concentration larger than 0.2%
- If you wait 2 hours after drinking, the effects of alcohol have disappeared
- If you drive carefully, combining alcohol and driving is not a problem

The answer scale to the knowledge questions above could be the “very much agree – very much disagree” scale listed above or also a scale denoting whether the statement is true or untrue:

- Completely true
- A bit true
- Neither true nor untrue
- A bit untrue
- Completely untrue

Or:

Completely true 1 -----2-----3-----4-----5-----6-----7 Completely untrue

3. Final remarks

Very often questionnaires are not set up from scratch, but are based on other research. It could be practical to take a questionnaire from another country as the starting point. However, it must be noted that copying a questionnaire from another country does not guarantee that you will have a good questionnaire for your country. There may be large or subtle differences between countries or cultures in how questions are understood. Also, the traffic system, drinking and driving laws, alcohol policies, and specifics of police enforcement may differ from one country to another. Hence, the questions on knowledge and attitudes need to be tailor-made for each country. The best way to do this is to organize feedback on the draft questionnaire from local experts and to organize a try-out of the draft questionnaire among a small sample (e.g., eight to 12) of the target population.

Toolkit

IARD Toolkits provide an overview of key topics, including approaches to developing and implementing alcohol interventions, policies, and situation assessments. While the *Toolkits* provide a comprehensive overview we recommend they be used in conjunction with IARD's other, more comprehensive resources, including *IARD Policy Reviews*.

© International Alliance for Responsible Drinking (IARD) 2016

IARD is a not-for-profit organization dedicated to addressing the global public health issue of harmful drinking. Our mission is to contribute to the reduction of harmful drinking and promote responsible drinking worldwide. This is a problem that requires new insights, urgent action, and open dialogue. Central to IARD's work is our role as Secretariat of the Beer, Wine and Spirits Producers' Commitments to Reduce Harmful Drinking.



IARD
INTERNATIONAL ALLIANCE FOR
RESPONSIBLE DRINKING
action on alcohol and global health

The International Alliance for
Responsible Drinking (IARD)
The Jefferson Building
1225 19th Street NW, Suite 500
Washington, DC 20036

Phone: 1.202.986.1159
Fax: 1.202.986.2080
Email: info@iard.org
Web: www.iard.org