Ref. Ares(2018)2710958 - 25/05/2018

# Delivery Report for

# MeBeSafe

# Measures for behaving safely in traffic

| | |
|---|---|
| Deliverable Title | Coaching methodology |
| Deliverable | D4.2 |
| WP | WP4<br>Driver Coaching |
| Task | Task 4.2<br>Research methodology |

This project (MeBeSafe) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723430.

## Disclaimer

## Copyright

| | |
|---|---|
| Grant Agreement No. | 723430 |
| Project Start Date | 01/05/2017 |
| Project End Date | 31/10/2020 |
| Duration of the Project | 42 months |
| Deliverable Number | D4.2 |
| Deliverable Leader | Bram Bakker, Cygnify |
| Dissemination Level (Confidentiality) | PU: Public |
| Nature | Report |
| Status | Final |
| Due Date | M13 (May 2018) |
| Deliverable Leader/ Responsible Organization | Bram Bakker, Cygnify |
| Contributing Author(s) | Anders af Wåhlberg, Cranfield University<br>Bram Bakker, Cygnify<br>Saskia de Craen, SWOV/Shell<br>Felix Fahrenkrog, BMW<br>Teri Lillington, Shell<br>Mikael Ljung Aust, Volvo Cars |
| Reviewer(s) | Olaf Op den Camp, TNO |
| Formal review and upload to the EU by Coordinator | Institute for Automotive Engineering - RWTH Aachen University (ika) |

## Abstract

This document describes the research methodologies employed within Work Package 4. WP4 focuses on the development of driver coaching schemes, supporting coaching software/apps, and evaluations of such systems. WP4 consists of various tasks. Some of these are directly related to each other (e.g. concerning coaching of Heavy Goods Vehicle drivers), such that their research methodologies are closely aligned; whereas others can and will be executed more independently (e.g. concerning coaching private vehicle drivers on the use of ACC) and therefore have their own methodology. For all tasks, we distinguish between the methodology for development and the methodology for evaluation, each of which is described in some detail.

Each of the research methodologies has at this point been sufficiently defined and where necessary aligned, such that we can move forward with the development and small-scale evaluation of the coaching methods and apps within WP4, and subsequent larger-scale evaluation in the field test of WP5. A point of concern is that within WP4 we will not be able to do pilot testing with many drivers, meaning that we cannot come to statistically sound results to guide decisions about the final coaching schemes and apps in WP5; however, the pilots and simulations will give sufficient insights into whether the apps and coaching schemes are ready for use in WP5.

| Version | Date | Comment |
|---|---|---|
| 1 | 28-2-2018 | Template and outline |
| 2 | 6-3-2018 | Updated template/outline after SdC comments |
| 3 | 20-4-2018 | Complete draft |
| 4 | 26-4-2018 | Updated complete draft after internal WP4 review |
| 5 | 11-5-2018 | Updated complete draft after MeBeSafe-internal review |
| 6 (final) | 20-5-2018 | Final revisions made based on second MeBeSafe-internal and WP4-internal review |

*table of document history*

# Table of Contents

## List of Figures

## List of Tables

# Acronyms

| Acronym | Explanation |
|---|---|
| **ACC** | Adaptive Cruise Control |
| **ADAS** | Advanced Driver Assistance System(s) |
| **HGV** | Heavy goods vehicle |
| **IVMS** | In Vehicle Monitoring System |
| **FOT** | Field Operational Test |
| **KPI** | Key performance indicator |
| **MeBeSafe** | Measures for Behaving Safely in Traffic |
| **ND** | Naturalistic driving |
| **UDRIVE** | European naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment |
| **WP** | Work Package |

## Glossary

| Term | Definition/explanation |
|---|---|
| Coaching | A collaborative solution-focused, results-oriented systematic process, used with normal, non-clinical populations, in which a coach facilitates the enhancement of work performance and the self-directed learning and personal growth of a coachee. |
| Competences | Dispositions that allow an individual to master variable situations successfully and responsibly and can be seen as fundamentals for learning. This includes motivational and violational aspects. |
| Driver profiling | Profiles based on driving behaviour in order to distinguish between different styles of driving (risky versus safe) or between driving behaviours in different situations. |
| KPI variables | The variables selected for measuring driving behaviour. The Key Performance Indicator (KPI) variables in the current study are: harsh braking, speeding, distraction, drowsiness, close following, harsh cornering, lane departure and possibly fuel consumption. |
| Naturalistic driving | A research method wherein every day trips by drivers are recorded by unobtrusive data acquisition systems with the aim of providing insights into actual driver behaviour. |

# 1 Introduction

## 1.1 MeBeSafe and work package 4

The aim of the MeBeSafe project is to develop, implement and validate measures that direct road users towards safer behaviour in common traffic situations. MeBeSafe is planning to do this by changing habitual traffic behaviour using 'nudging' and coaching, with the aim of improving driving behaviour. Nudging is a technique that subconsciously stimulates drivers to drive safer, while with coaching, drivers are given feedback on their driving behaviour by a coach in order to learn about their own driving behaviour and enhance driving performance. The work described here is focussed on coaching, in particular of heavy goods vehicle (HGV) drivers (employees of companies contracted to work for Shell), but also private consumer vehicle drivers (recruited by Volvo Cars).

MeBeSafe is organised in altogether six work packages (WPs), as shown in *Figure 1.1*. The coaching of drivers on their driving behaviour is part of Work Package 4.

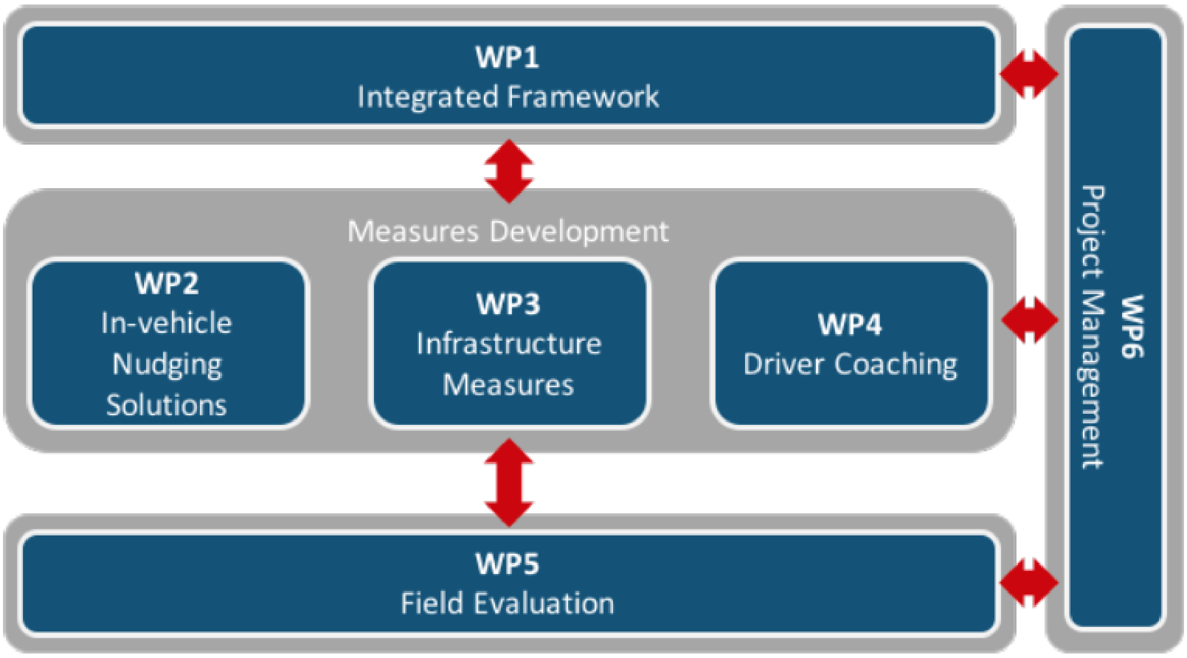

*Figure 1.1 Work packages in MeBeSafe.*

Work Package 4 focuses on the development of driver coaching schemes, software/apps that can be used to support the coaching of drivers, and evaluations of the coaching schemes. HGV drivers that will be approached to participate in the study are working for hauliers contracted by Shell; private (consumer) vehicle drivers are recruited by Volvo Cars. The end result of WP4 consists of coaching schemes and a coaching app or in-vehicle software that can be used for a large-scale field evaluation in WP5. WP4 consists of multiple tasks (see chapter 2 for details); in this deliverable the work in *task 4.2* is described, which is concerned with the research and coaching methodologies used in various parts of WP4.

## 1.2 Objective of task 4.2 and this deliverable

The objective in task 4.2 is to *define* and *align* the research methodologies used in the various parts of the work package that are concerned with the development and evaluation of driver coaching and feedback measures. This deliverable document is both a progress report and a reference document, aimed at clarifying and solidifying the connections between the various parts in the work package, and where necessary identifying possible gaps or mismatches, so they can be resolved in time.

WP4 is made up of different parts (tasks and subtasks; see chapter 2), each of which is concerned with a separate "component" in the overall work package (e.g.: development and evaluation of the HGV driver app), and which has its own research methodology. The research methodologies between different parts may (and will) differ; but it is important that where necessary they are sufficiently *aligned*. In WP4, we strive towards coherence between the tasks and subtasks, such that the "output" of one task or subtask is used properly as the "input" of another task or subtask, and the overall output of the work package is the result of the collaborative effort of the tasks and subtasks. If that is the case, the end result is, in a way, 'more than the sum of its parts'. Task T4.2 and this deliverable document D4.2 aims to facilitate this process.

As an example of such alignment, when a simulation study (T4.4) is performed to assess likely safety effects given a future large-scale roll-out of the HGV driver app and associated coaching scheme (T4.3), and in parallel a small-scale real-world pilot study (T4.3/4.4) is performed to obtain preliminary estimates of likely behaviour changes and safety effects, it is important that the methodologies are aligned, to some extent. They should be aligned in the sense that they use evaluation measures which can be compared in a meaningful way, and that likely behaviour changes as estimated by the real-world study are taken into account in the modelling for virtual simulation. Moreover, the real-world pilot and simulation should both represent sufficiently accurately the actual app and coaching measures which have been developed (T4.3). And the app and coaching measures, should also be aligned with other tasks, i.e. be based on and take into account driver behaviour profiling data (T4.1).

On the other hand, some (sub)tasks are relatively independent of others, and their research methodologies can be relatively independent, and they cannot and do not need to be aligned so much with the other tasks. This is the case in particular for the subtasks concerned with Adaptive Cruise Control (ACC) coaching for private drivers (headed by VCC), which has its own chapter.

"Research methodology" is a broad concept. In this work we distinguish two separate aspects of research methodology, each of which is addressed:

- **Development methodology**: the methods and approaches used in *developing* a certain coaching scheme or system used in WP4. This may include:
    - descriptions of scientific literature or previous studies or projects on which a coaching scheme or app design is based;
    - descriptions of how the app or in-vehicle software or additional components such as back-end software or additional components such as cameras are built and included in the overall system;
    - descriptions of how driver behaviour profile data is used in the app or in-vehicle software or coaching scheme;
    - descriptions of simulations or simulators built, adapted and used.

- **Evaluation methodology**: the methods and approaches used in *evaluating* a certain coaching scheme or system used in WP4. This may include:
    - descriptions of a small-scale real-world pilot study, participants in that study, experimental vs. control group allocation, intended statistical tests (if any);
    - descriptions of intended simulator or simulation scenarios and parameters, intended evaluation measures in simulator or simulation studies, experimental vs. control conditions, intended statistical tests (if any) on simulator or simulation results;
    - descriptions of initial, roughly outlined plans for how the larger-scale real-world study in WP5 will be performed and will build on the small-scale real-world studies and simulator or simulation studies.

## 1.3 Report structure and contribution by partners

In this deliverable we focus on the development and evaluation methodologies of:

- ACC coaching for private vehicle drivers (see chapter 3)
- Coaching of truck (HGV) drivers using an app and face-to-face coaching (see chapter 4)
- Simulation studies to estimate safety effects given large-scale roll-out, (see chapter 5)

Complete and separate methodologies have been developed for all three in parallel. The first two are concerned with the actual development of coaching apps/schemes. The third one, the simulation studies, will focus on simulation of safety effects, given large-scale roll-out of the HGV driver coaching, which is a sufficiently large and comprehensive task to warrant its own chapter. Development and evaluation of the coaching method and associated app for HGV drivers is the largest task in WP4, thus receiving most attention in this report.

The chapters are written by different organisations and authors. The structure of the deliverable is as follows (see *Table 1.1* for an overview). The current chapter, Chapter 1, provides the introduction. The next chapter (Chapter 2) gives an overview of and describes relationships between the different tasks. The core of this report consists of the chapters on ACC coaching for private vehicle drivers, coaching of truck (HGV) drivers using an app and face-to-face coaching, and simulation studies to estimate safety effects, respectively. Finally, Chapter 6 provides concluding remarks, focusing in particular on how this work in WP4 leads into and carries over to the large(r) scale study in WP5.

| Chapter | Title | Main focus | | Author |
|---|---|---|---|---|
| | | Development Methodology | Evaluation Methodology | |
| **Ch. 1** | Introduction | | | Cygnify |
| **Ch. 2** | Overview of and relationships between tasks in WP4 | | | Cygnify |
| **Ch. 3** | ACC coaching research methodology | Coaching scheme designed to stimulate ACC use | WP4 small-scale pilot with private vehicle drivers; WP5 large-scale study with private vehicle drivers | Volvo Cars |
| **Ch. 4** | Truck app & coaching research methodology | App and face-to-face coaching, in various versions, designed to improve behaviour on various monitored KPI driving behaviour variables | WP4 small-scale pilot with professional HGV drivers; WP5 large-scale study with professional HGV drivers | Cranfield University/Shell /SWOV |

| | | | | |
|---|---|---|---|---|
| **Ch. 5** | Simulation-based effectiveness research methodology | Simulation of traffic agents and safety effects, based on driver behaviour change modelling in computer simulations | Evaluation of simulated behaviour changes and safety effects under assumptions of large-scale roll-out | BMW |
| **Ch. 6** | Concluding remarks | | | Cygnify |

*Table 1.1 Overview of the chapters, the focus of the chapter, and their author*

# 2 Overview of and relationships between tasks in WP4

## 2.1 Introduction

Work Package 4 focuses on the development of driver coaching schemes, software/apps that can be used to support the coaching of drivers, and evaluations of the coaching schemes. This is done for HGV (truck) drivers, which is the main focus of this work package, but also for private vehicle drivers. Within WP4, smaller scale tests and piloting are done to arrive at the end result of WP4: coaching schemes and coaching app(s) or in-vehicle software that can be used for a large-scale field evaluation in WP5.

## 2.2 Overview and relationships

The work package consists of the following 4 tasks:

1) Driver and situation profiling for the coaching schemes (task 4.1 – reported in deliverable D4.1);
2) Defining and aligning the research methodologies used in various parts of WP4 (task 4.2);
3) Development of the coaching scheme and coaching app, including a small-scale pilot (task 4.3);
4) Evaluation of the coaching scheme and app based on the pilot results (task 4.4).

The relationships between the tasks are visualised in the schematic picture of Figure 2.1. Task 4.1, on driver and situation profiling, provides input on which drivers, driver behaviors, and situations to focus on in task 4.3, by giving information on driver types, driver behaviour & situation classifications, and driver KPI variables to focus on.

The KPI variables to focus on consist of:

- Harsh braking

- Harsh cornering
- Speeding
- Close following
- Lane departures
- Distraction
- Fatigue and drowsiness
- Optionally: fuel consumption

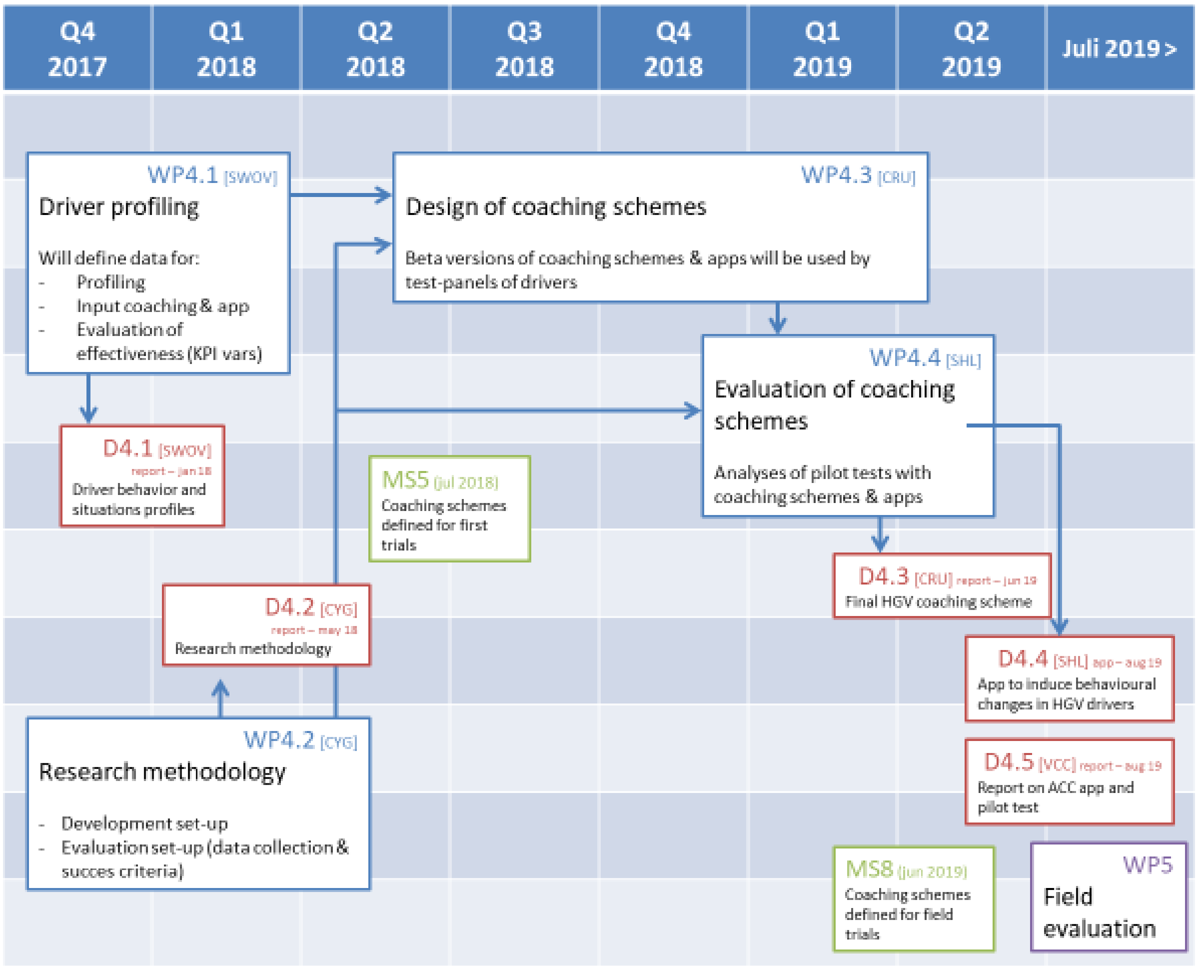

*Figure 2.1 Schematic overview of the various tasks in Work Package 4 and their relationships.*

Note that these KPI variables are not independent from each other, and some can be measured more directly than others (see Deliverable Document 4.1, which describes this and the main results and approaches developed within task 4.1 in detail). For example, fatigue, which by itself is not easy to measure (although we will use methods based on inward-facing cameras and face/eyes analysis), may lead to more harsh braking and lane departure events. However, together these KPI variables are the variables we wish to improve on, even if they are correlated.

Task 4.3 is concerned with producing the main results of WP4. This task results in the coaching methods and the apps, which are to be evaluated in the project. These result are first used within WP4, i.e. in task 4.4, for evaluation on a small scale as well as in simulation. Subsequently the coaching methods and apps are transferred to WP5, in which a larger-scale test and evaluation is performed.

Finally, task 4.2, concerning Research Methodology (the focus of this document), has an overall supporting, defining and monitoring role, relative to the other tasks; giving input to the tasks of designing the app and coaching schemes and the evaluation of them.

Within the tasks, there are some parts (or subtasks) which can be, and are, executed largely or completely independently (this is not visualised in detail in Figure 2.1). Most notably, the work on HGV drivers (the main focus of WP4) is done independently from the work on private vehicle drivers (ACC coaching, by VCC). Furthermore, evaluation is done both by real-world pilot tests and by computer simulations (by BMW). These different parts are described separately in the following chapters.

# 3 ACC coaching research methodology

## 3.1 Introduction

In order to crash into a lead vehicle (have a rear end collision); the accident causation research shows that generally two things are required:

1) A distracted driver, and…
2) … a lead vehicle.

It may seem irrelevant to bring up the second condition, but in fact, this is very important. The reason is that many proposed solutions to the problem of rear end crashes only focus on the first issue (the distracted driver), but do not address the second one (close following). However, research on this crash type clearly shows that the risk of crashing is highly influenced by how far behind you are when the unexpected happens (Victor et al., 2015). If you're not following a lead vehicle very close, you're much more able to resolve the conflict once it arises. Rear end crashes can therefore most likely be addressed just as well by avoiding close following as by avoiding distracted drivers.

Given this, the next question becomes how to make drivers (of private vehicles) avoid close following. Given that driving is largely automatized and habitual, nudging a particular driver to alter his/her distance keeping behaviour is challenging.

However, there is a simpler way. As illustrated in the figure below, when ACC is on, people get into short time headway situations much more seldom (results from euroFOT, Malta et al., 2012, where driving with and without ACC on in lead vehicle following situations was compared).

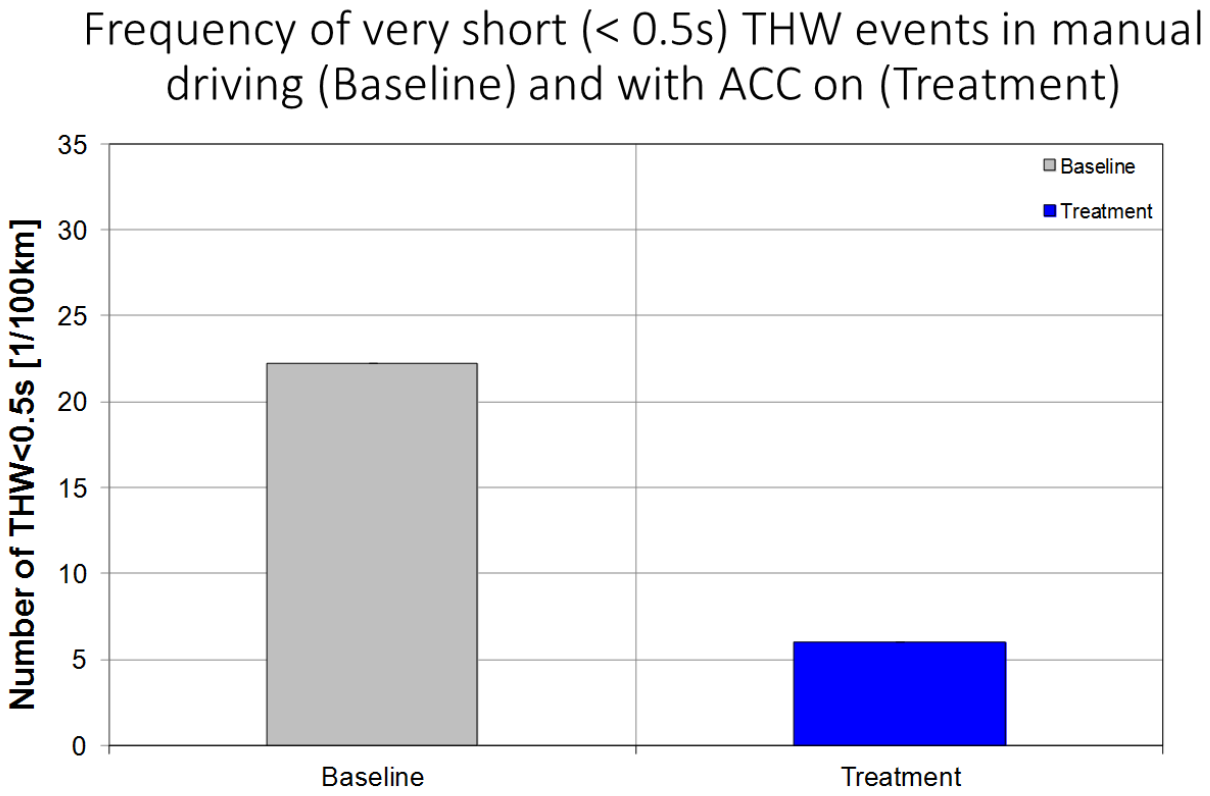

*Figure 3.1 Frequency of short time headway events per 100 km of driving as a function of whether ACC was active or not. Source: euroFOT (Malta et al., 2012).*

One of the two main use cases that VCC has in MeBeSafe is therefore to nudge drivers of private vehicles to use ACC more in their everyday driving.

The main part of that use case, i.e. developing and installing an in-vehicle feedback mechanism that reflects drivers' ACC usage in such a way that they become more inclined to use it, is handled in WP2, and thus not the topic here. Similarily, developing and installing the incentive mechanism intended to increase compliance with the VCC drowsiness alert system will be studied within WP2 rather than WP4.

However, there is another aspect to it, which is where WP4 and coaching comes into play. Our analysis of ACC usage levels in baseline (i.e. no nudging/coaching applied) suggests that ACC usage is not uniform across the population. Rather, fairly clear user groups can be identified, and their ACC usage levels suggest different approaches are required to increase their overall ACC usage. In the Figure below, ACC usage data that was collected over a 6 month period from 19 Swedish company car drivers (VCC employees) has been compiled.

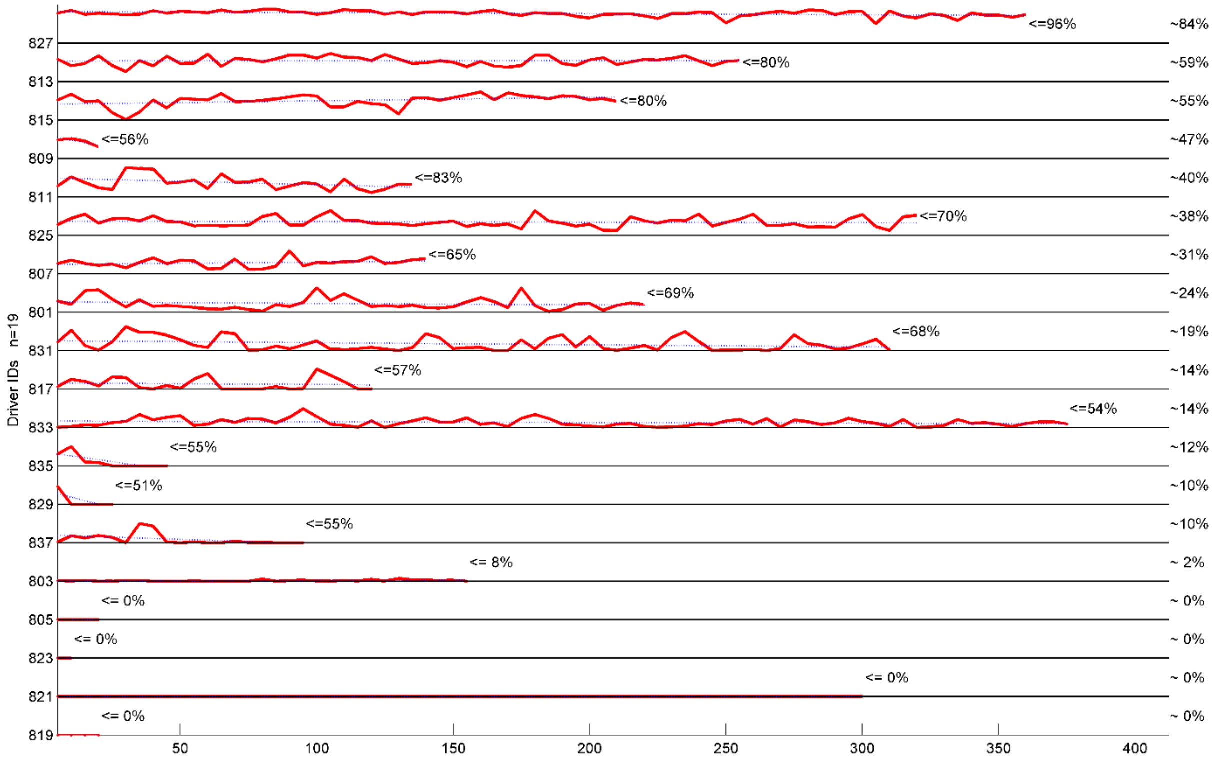

*Figure 3.2: ACC usage ratio per 5 hours of driving at speeds above 30 kph. Source: internal VCC research.*

This Figure tells us three things. First, on the individual level, drivers seem fairly consistent in ACC usage levels over time. Second; the range of usage spans from 0 % to 64 %, so not everybody uses ACC in the same way. Third, there seems to be an overall pattern to the usage level. There are those who use it a lot (say above 30 %, exactly where to draw the line can of course be discussed), those who use it a little (say 10-29 %), and those who do not use it at all.

Based on driver interviews, one way of labelling these three groups is to call them the *tech happy*, the *modest users*, and the *non-users*. This is where coaching comes in. The first group do not really need coaching or nudging, since they are already using ACC to a large extent. The second group are the ones who might be susceptible to nudging, because they already know how to operate the system and use it from time to time. These will be targeted in WP2.

The third group however, need to start using the system in the first place before any nudging can be applied. Moreover, experiences at VCC tell us that these people need to be able to test drive under controlled circumstances with someone at their side instructing them what to do and what to expect, before they will consider using the system on their own.

Since this is very labour intensive to do manually, the idea is to let these drivers test ACC under coaching from an app that they bring into the vehicle on a smartphone or tablet. The app will be tied to the vehicle data streams and can thus determine when it is a suitable time to test ACC, and then instruct the driver how to engage and disengage, as well as change time headway.

## 3.2 Development methodology

For coaching first time ACC usage in MeBeSafe, we will further extend the capabilities of a Test Drive Prototype App that currently is being developed within VCC. The app will be integrated in the main infotainment system Sensus. This means that interactions between driver and system can be voice controlled. It also means that it is possible to send the driver out on a specific route, with visual and auditory guidance, and set specific interaction points where it is suitable to test ACC for the first time.

Current testing of the prototype app indicates that it is well received by test participants, and that trying systems out in traffic under guidance is a highly appreciated feature.

## 3.3 Evaluation methodology

To evaluate the effectiveness of coaching through the test drive app, we first need to identify a suitable group of non-users. These will be approached by analysis of our Customer Field Data Fleet, which currently contains 350 drivers of various vehicles in various configurations. A select group of non-ACC-users will be invited to test ACC using the Prototype Test Drive App. Their initial response to the experience will be recorded, and then we will follow up on their actual ACC usage in their own vehicles over a period of 2-3 weeks after the test drive has taken place.

Assuming this approach is efficient; the drivers who are to take part in the larger field study will be monitored through a first 1 month baseline phase. Those who fall into the non-ACC-user group will then be identified and put through the test drive program. If their usage does not increase after the test drive, a currently open question is whether to bring them in to do the same thing once more, or whether a test drive with a human rather than an app should be tried.

## 3.4 Conclusions and next steps

Currently, development of the ACC specific application within the test drive app is starting up. Once ready, it will be field trialled (WP5) on a selected group of non-ACC-users, and iteratively modified to a point where the User Experience is sufficiently good.

## 3.5 References

Malta, L., Ljung Aust, M., Faber, F., Metz, B., Saint Pierre, G., Benmimoun, M., & Schäfer, R. (2012). EuroFOT final results: impacts on traffic safety. Retrieved from http://www.eurofot-ip.eu/download/library/deliverables/eurofotsp620121121v11dld64_final_results_impacts_on_traffic_safety.pdf. Source Date: 11/21/2012.

Victor, T.; Dozza, M.; Bärgman, J.; Boda, C.-N.; Engström, J.; Flannagan, C.; Lee, J.D.; Markkula, G. (2015). Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention and Crash Risk; Transportation Research Board: Washington, DC, USA, 2015; Available online: http://onlinepubs.trb.org/onlinepubs/shrp2/SHRP2_S2-S08A-RW-1.pdf

# 4 Truck app & coaching research methodology

## 4.1 Introduction

This chapter describes the development and evaluation of a coaching system for truck drivers (or HGV drivers) in MeBeSafe. In addition to coaching in the traditional sense (face-to-face), this includes a supporting app (which also functions as a measurement tool for the evaluation). These two features of the coaching system are strongly interconnected.

The main objective of coaching HGV drivers is to improve their harsh braking safety performance (Objective 5 of the MeBeSafe project). In addition, we aim to improve several driver behaviours that are related to traffic safety, such as harsh cornering, close following, lane deviations, drowsiness/fatigue, distraction, speeding, and optionally fuel consumption (Dingus et al., 2016; FMCSA, 2006; Hanowski, Perez & Dingus, 2005; Olsen et al., 2009; Sagberg et al., 2015; SWOV, 2016). Together these are the KPI variables in WP4 (see also chapter 2, where these KPI variables are discussed as well).

In this chapter the methodology of the development of the coaching system will be discussed first (section 4.2). This consists of three overlapping phases and methodologies. First, a literature review undertaken mainly within WP1, second a feasibility inquiry with HGV drivers and hauliers, third, a development phase, in which different versions of the app (with increasing sophistication) are tested or tried out by a small group of 'early adopters'.

Section 4.3 describes the methodology of the evaluation phase. The version of the app and coaching program that seems most promising at the end of the developing phase will be pilot tested with a somewhat larger group of drivers in a more controlled setting, within task 4.4. The results of this pilot test will be used to define the set-up for the larger field tests in WP5.

## 4.2 Development methodology

### 4.2.1 Literature review; extracting the basic principles for coaching

Within WP1, literature on coaching and similar approaches (developmental relationships in general, therapy and teaching) was searched for methods which can be applied in a transportation setting, as well as evidence of their effects. The results of this work has been described in detail in D1.1, and will here only be summarized and related specifically to the development process and methodology.

The basic method applied in the literature study was to look for intervention principles, to be implemented in the coaching scheme and app, which could work for professional drivers, i.e. it should be possible to combine the intervention with their normal work situation. When different approaches to inducing behavioural change had been found, evidence for their effects was analysed. Mainly, effect sizes were used as the criterion for effectiveness. It was found that the best evidence was for cognitive behavioural therapy and the techniques applied within this approach. Specifically, the most efficient methods would seem to be goal-setting, instruction, feedback, social support, self-monitoring and information about consequences, and are all highly applicable within an app+coaching scheme.

The basic assumptions in WP4 are that:

1) truck drivers are skilled drivers but do not always use their skills when driving;

2) providing a social setting for their driving as well as feedback on the quality of their driving will improve performance;

3) drivers' integrity needs to be respected to achieve buy-in.

The coaching intervention consists of two parts; an app which measures driver behaviour and provides feedback of this information, and a coaching scheme. Both of these are situated after driving, i.e. there is no in-cab, instantaneous feedback (apart from a possible real-time warning signal—see the description of different app versions below). The reasons for not using in-vehicle feedback is the risk of cognitive overload (Donmez, Boyle & Lee, 2007), as well as the lack of effect for this type of feedback shown in some studies (e.g. Bell et al., 2017).

### 4.2.2 Feasibility

To assess the feasibility of our proposed measures in terms of the possibility of implementing them in a professional driving environment, we discussed these with the participating companies Gasnor/Litra (Norway) and van Waveren (the Netherlands) and their drivers. Also, trips with some drivers to study their work environment, have been undertaken. This work has indicated that the planned app and coaching schemes can be implemented, although certain features and information in the app are probably less important in this specific environment than others. These contacts indicate that truck drivers have rather different work and driving environments in different countries and companies. Therefore, their coaching and information needs and how they can use the app differ. The development and implementation of the system must therefore take this into account.

Our approach for coaching initially met some resistance (i.e. “difficult to organise”, “drivers don’t want to be coached by their peers”), but with further talk with hauliers[1] and their drivers, it was found to be feasible. Resistance against (peer-to-peer) coaching is mainly amongst company executives, while drivers have a more positive attitude to this method.

### 4.2.3 Coaching

#### 4.2.3.1 Peer-to-peer coaching

Using peer-to-peer instead of supervisor coaching would seem to be preferable especially when the workers are not part of a natural work group, but mainly work

---

[1] Litra, Norway; van Waveren and Nijhoff-Wassink, the Netherlands

alone. Another advantage of this method is that it is less difficult and costly than assigning supervisors as coaches. Also, drivers might simply be more accepting of other drivers' advice, as supervisors rarely drive trucks themselves, and therefore lack the experience of professional truck drivers.

However, to ensure that the drivers do their coaching, they are to report to a supervisor about once every three months. This meeting will also be a coaching session, where the drivers report how they have worked, what obstacles they have encountered and what progress has been made. The supervisor should offer praise and advice, and help out with any practical problems that are perceived to influence driving (e.g. fatigue due to driving schedule). However, data on individuals should not be made available to the supervisors, only the aggregated results for the company, to respect the drivers' privacy and increase compliance with the program.

#### 4.2.3.2 Training of coaches

Drivers who have agreed to coach and be coached[2] will attend one training session (2 h) on how to coach other drivers, using the principles detailed under coaching methods. The drivers will also be trained in the use of, and interpretation of data from the app. The app will provide a more comprehensive manual on coaching techniques and the use of the app.

#### 4.2.3.3 Frequency of coaching meetings

Coaching meetings should preferably take place once every two weeks, but when improvement is visible, this could be extended to longer time periods. The length of each meeting should be 15-30 minutes. Meetings should be in person, but if not possible, video or telephone meetings can also be used.

---

[2] Because of the peer-to-peer coaching, participating drivers are coach and coachee at the same time

In WP4, a limited time for the intervention is available. In actual use, however, the coaching and feedback programme is expected to be a permanent part of employment, as effects of driving behaviour interventions often rapidly fade when the intervention is removed (af Wåhlberg, 2007; Hickman & Geller, 2005). However, it is also expected that the effect will stabilize after a few months. At that point, the frequency of meetings can be reduced.

#### 4.2.3.4 App

A smart-phone-based app will be used to gather data for feedback and evaluation (see also section 4.2.4 – Developing the safe driving app). The variables used in the app should create an output about once a week for each driver (the frequency can be set by the drivers), which is the information to be used for the coaching discussion. Also, the mean for all drivers on the same variables for the same period (properly anonymised) should be fed back to all drivers. This way drivers can compare themselves with the average driving behaviour of all drivers in the company.

#### 4.2.3.5 Coaching methods

Concentrating on the positive

The standard for telematics feedback and coaching is negative feedback, meaning that drivers are usually only informed when they have done something wrong (e.g. Adell, Várhelyi & Hjälmdahl, 2008; Albert et al., 2011; Boodlal & Chiang, 2014; Duarte, Gonçalves & Farias, 2013; Hickman & Geller, 2005; Musicant & Lampel, 2010). Although errors in driving must inevitably be discussed, an important part of the MeBeSafe coaching is to point out improvements and situations well handled, i.e. positive feedback.

Running a coaching session

The drivers will be instructed in how to behave in a meeting with their peer, to make the encounter as smooth and positive as possible. Furthermore, they will be taught to use behavioural change techniques.

### 4.2.4 Developing the safe driving app

The app that will be developed in the MeBeSafe project will give feedback on driver behaviour. This feedback can be useful by itself, to motivate drivers to improve their performance, but also serves as important input for face-to-face coaching.

As discussed above, we are aiming for the app to provide feedback on: harsh braking, harsh cornering, close following, lane deviations, drowsiness/fatigue, distraction, speeding, and optionally fuel consumption. Research has indicated that these behaviours are related to traffic safety (Dingus et al., 2016; FMCSA, 2006; Hanowski, Perez & Dingus, 2005; Olsen et al., 2009; Sagberg et al., 2015; SWOV, 2016) and are therefore relevant for coaching.

After having reviewed the area of telematics (IVMS) and driver behaviour, it was decided that the MeBeSafe app, apart from delivering information and feedback, should also be used as a platform for data collection (also see Deliverable 4.1 – driver profiling). This decision was due to the fact that different hauliers use different IVMS providers, whose systems are different from each other. Thus, the data would not necessarily be comparable, and could also be difficult to get access to for analysis. Also, the use of an app developed within the project as the sole technical system involved would make further development and application in new fleets easier.

The development of the app and coaching involves the roll-out of different versions. Each version will be tested by four drivers from Litra in Norway[3]. After a version works satisfactorily, a new version will be introduced. See section 4.4. for a time schedule for the implementation of these versions. Each version of the app will have more sophisticated features (e.g. the use of cameras is introduced as an option in V2), and will be more informative for coaching. In the most complete form the app will include information on the complexity of the traffic situation (see next section 4.2.5); and positive feedback on driving performance is possible. The following versions will be rolled out:

---

[3] 3. We are aiming to have these versions tested simultaneously by four drivers from van Waveren in the Netherlands.

**V0** In this version the app measures drivers' behaviour, but no feedback is given nor is there any coaching taking place. This provides the baseline measurement.

**V1** This version includes basic features of the app (straightforward IVMS/GPS data only, no cameras are used) and provides information on speeding, harsh braking and harsh cornering. The drivers start using the coaching scheme. The figure below illustrates how the in-vehicle app is connected to (secure) offboard servers in the cloud, which take care of data collection and storage, map matching, and the like. These data can subsequently be analysed further by MeBeSafe researchers, and data can be fed back to the app for peer-to-peer coaching.

**V2** In this version cameras (both outward-facing and inward-facing) are optionally installed, for those drivers who volunteer for this. When using cameras, short 20 to 40 second videoclips surrounding certain events derived from the standard IVMS/GPS sensors (e.g. a harsh braking event) will be recorded and uploaded to the offboard servers (see figure below). There they will be analysed by specialised software using dedicated high performance computing hardware, and analysis results returned to the app together with the videoclip itself, where they can be used for coaching purposes and the videoclips can be watched. Furthermore, information on route traffic complexity (see section 4.2.5) is used to provide better, more contextualised feedback to drivers. For those drivers who have cameras onboard, route traffic complexity will in part be computed by analysing outward-facing videos. In any case (also without cameras), route traffic complexity will be estimated based on map, traffic, weather data and the like. When using cameras, more KPI variables can be included in the feedback (e.g. close following, lane deviations, drowsiness/fatigue, distraction).

**V3** Version 3 is the final target version that has the inclusion of journey management features and checklist integration. Furthermore, this version also includes an optional real-time camera and computation system to warn drivers for potential real-time risks (close following, distraction, drowsiness/fatigue, lane departures, animal detection) using continuous video stream analysis, resulting in a sophisticated state-of-the-art demo-system, to be rolled out to only 1 or 2 trucks. In contrast to version 2, the real-time camera and computation system requires a separate onboard computer, next to the app, as depicted in the figure below.

*Figure 4.1 Description of the roll-out versions of the app and coaching scheme*

The final development phases for the app and coaching systems will be based upon qualitative feedback from the drivers about how they use the system and quantitative behaviour variables measured by the app (see section 4.3 Evaluation methodology).

### 4.2.5 Road traffic complexity, driver behaviour and performance

In WP4 we make a distinction between driver behaviour (what a driver does in a given situation) and performance (the quality of the behaviour in relation to the driving environment). This distinction is useful because it distinguishes between what a driver is capable of doing and the demands of the road environment. This can also be phrased as a need to hold the effect of the environment constant when driver behaviour is measured. In standard telematics technology, this does not happen (e.g. Albert et al., 2011; Boodlal & Chiang, 2014; Duarte, Gonçalves & Farias, 2013; Hickman & Geller, 2005; Lotan, Toledo & Prato, 2009; Musicant & Lampel, 2010; Musicant, Lotan & Toledo, 2007; Soleymanian, Weiberg & Zhu, 2016), and results for different drivers may therefore not be comparable, due to their driving in different environments. The feature of the driving environment which influences driving behaviour will henceforth be called road traffic complexity.

The central idea is to use estimated road traffic complexity to be able to provide better, more fine-grained, more contextualised feedback to the driver. For instance, it is to be expected that in an urban environment with many other road users and many intersections with crossing road users the amount of harsh braking is relatively large, compared to quiet, straight rural roads with very few intersections and road users. In such a case, we cannot directly use the measured amount of harsh braking as a measure for ‘how good’ a driver was driving (and as such that is an imperfect measure for coaching), but instead need to take the (environmental) traffic complexity into account; for example by scaling harsh braking relative to such road traffic complexity.

Road traffic complexity does not seem to have been studied in its own right, but instead has been accepted as an unproblematic feature. This is apparent because whenever complexity is used in some way, it is not discussed or measured by the researchers using it (e.g. Cantin et al., 2009; De Craen et al., 2009; Jahn et al., 2005; Rudin-Brown, Edquist & Lenné, 2014). This approach by traffic safety researchers is very different from that of air traffic control, where complexity of the environment is an important topic with several different methods developed for measuring it (Hilburn, 2004; Mogford et al., 1995).

This section describes the various features involved in measuring road traffic complexity, driver behaviour and performance, and the methodology which is to be used for further development of these concepts in WP4.

#### 4.2.5.1 Complexity

The ultimate goal of including complexity in the MeBeSafe measurements and feedback is providing more relevant data about performance for the drivers, i.e. variables like harsh braking are held constant for the difficulty of navigating the environment. To handle complexity of the environment in the app, information and feedback systems, it was decided to use two different variables, henceforth called macro- and micro-level complexity. These are independently derived from two different sources, and used to hold complexity constant in two different situations; macro when cameras are not available and micro from camera data (thus, only available from versions v2 and v3 of the app, for those driver who opt in to cameras)[4]. This separation is due to three different facts: 1) there are two main possible sources of data which can be used within this project, databases and similar sources (for road layout, weather etc.) and videos captured from the trucks; 2) complexity is not well understood today as a factor in traffic safety, and validation is needed, where the correlation between the two complexity variables is one type

---

[4] The terms "macro" and "micro" traffic complexity derive from similar uses of those words in computational traffic models and simulations. In our case, "micro" refers to taking into account individually perceived road users and their trajectories (as detected using cameras) and deriving complexity based on that; whereas "macro" refers to more 'high-level' estimates of complexity, at the level of relatively large stretches of roads and without knowledge of individual road users, using map database features such as functional road class, highway vs rural vs urban roads, speed limits, road curvatures and slopes and the like; as well as weather- and traffic data (e.g. from Google, Here or TomTom).

of test; 3) due to the different levels of technical sophistication in the different roll-out versions, as well as the anticipated need to be able to operate the app without the camera option, a macro- database-based complexity variable is necessary.

#### 4.2.5.2 Development of complexity variables

The complexity variables will need to be developed, since there is no such variable available yet. The complexity variable needs to include different factors influencing the complexity of the driving situation forming a level of complexity at a certain time. The complexity variable using data sources external to the vehicle (databases etc.), what will now be called macro-level complexity, will be developed from truck crash data available from published sources (e.g. Blower & Campbell, 1998; Blower, Campbell & Green, 1993; Blower et al., 1990; Braver et al., 1997; Campbell et al., 1988; Graf & Archuleta, 1985) (Jovanis, Chang & Zabaneh, 1989; Jovanis & Delleur, 1983). The basic principle of development will be to start from what databases are available for road traffic features which can be said to be part of complexity. Currently, this includes weather, speed limits and time of day. Additional features could be for example the load of the vehicle and type of terrain.

Micro-level complexity will be developed from the output of computer vision-based image-processing software, using video images collected by cameras. In principle, it adds the number of road users within a certain distance to the number of junctions, their trajectories, and possibly the degree of curvature of the road etc., to form an estimate of complexity at this micro-level of analysis. Further refinements within this model will be different weights for different speeds and trajectories of the road users (where, for example, a stationary road user has the lowest weight, and a road user with an intersecting trajectory will have the highest weight). To capture the micro-level of complexity cameras need to be installed, so this is only possible in versions v2 and v3 of the app/coaching scheme set-up.

### 4.2.5.3 Validation of complexity variables

As traffic complexity is a rather untested concept, it is necessary to undertake some validation within the MeBeSafe project. This will be done using three different approaches.

First, validation against subjective driver judgements can be used. This can be undertaken using Naturalistic Driving (ND) data available to the consortium, in particular UDRIVE data (see Deliverable Document 4.1 for more information on ND and UDRIVE data), and/or new data can be gathered from the piloting undertaken in Norway. The validity coefficient would be the correlation between the average complexity rating indicated by drivers for different traffic scenes versus the computed complexity variables.

Second, the correlation between the two (micro- and macro-level) complexity variables in itself would be an indicator of validity, which is the reason for using totally distinct data sources for the variables. This principle avoids the problem of common method variance. The validation might be possible to perform in UDRIVE, if database data can be extracted for the scenes recorded in that dataset. If not, data from the pilot study in Norway will be used.

Third, both complexity variables can be correlated with the celeration variable (af Wåhlberg, 2006; 2007; 2008) averaged over drivers. The celeration variable (the average of absolute positive and negative acceleration when speed is >0) yields a measure of total risk of crash, i.e. it takes into account both the environment and the behaviour of the individual (Af Wåhlberg, 2008). Averaging over individuals will even out the individual differences, and the differences over locations (segments of road) will therefore reflect the effect of the environment.

Sizeable correlations between the two complexity variables, subjective complexity and average celeration, measured over different environments, will indicate that the complexity variables have some validity. These correlations should increase as longer sections of road are used. A final value of >.5 should be reached for acceptable validity to have been achieved. Also, if the complexity index is held constant in between drivers reliability calculations (of, for example, harsh braking), the correlations should increase.

## 4.3 Evaluation methodology

This section of D4.2 describes the methods for the evaluation of a pilot test in WP4. It describes the research design, the intervention, the variables and the methods used for analysing data.

The aim of the pilot field trial in WP4 is to determine whether the app and coaching schemes are ready for use in WP5. This pilot test will (at least in part) be conducted with 'new' drivers, who were not involved in earlier development phases of the app and coaching, and will start using these measures for the first time. The evaluation will answer two questions:

1) qualitative analysis: what are the drivers' experiences of using the app and coaching?
2) quantitative analysis: does the driving behaviour change as a result of using the app and coaching?

### 4.3.1 Design of WP4 pilot test

The effects of the intervention will be tested in a 'before-after without control' field trial with 10 (or more) drivers from Litra in Norway/Sweden[5]. These are different drivers than those who tested the initial version of the app and coaching scheme. Because of the limited number of drivers, and participation being voluntary, random assignment of drivers to an intervention and a control group is not possible.

---

[5] We are aiming to also include drivers from van Waveren, the Netherlands in this pilot test.

Therefore, the results will be methodologically very weak, i.e. validity of the findings will be low.

The absence of a control group is a problem, as it means that no comparable data can be collected from non-participating drivers. To counter this problem, attempts will be made to use data from the companies own telematics system for comparisons between drivers. If possible, this data will indicate how similar participants and non-participants are regarding their driving behaviour, and also how they react to the intervention. However, this evaluation will be limited by the quality of the data available from the telematics system, something which cannot be determined at the present time.

### 4.3.2 Baseline measurements and intervention

After a short period of baseline measurement these drivers will start using the app and coaching scheme.

#### 4.3.2.1 Baseline measurements

Baseline measurements from the app will be limited to a short time window of approximately two months.

#### 4.3.2.2 Coaching scheme

The drivers will start working with a coaching scheme as described in section 4.2.3. This includes a two hour training session on how to coach and fortnightly (once every two weeks) peer-to-peer coaching sessions with a partner driver.

#### 4.3.2.3 Safe driving app

The characteristics of the app that will be tested in the pilot test depends on the progress in the development of the app. We are aiming for at least Version 2, in which the app includes information on the complexity of the situation.

### 4.3.3 Variables & instruments

The variables that will be evaluated are the variables which drivers will get feedback on, such as speed, acceleration and harsh braking variables. When the complexity variables have been developed, validated and implemented, these can be used for holding the influence of the environment constant on the dependent variables, which will increase statistical power (see section 4.3.4.1). For validation purposes these variables will be measured with the app, but also with the IVMS already available in the trucks.

In addition to these variables that measure driving behaviour, we will also measure the drivers' experiences with the app and coaching scheme using a questionnaire. This questionnaire will include questions such as: Did you enjoy using the app? Was it easy to remember to turn on the app for every trip? Did you enjoy the coaching sessions? Were they useful? How can the app or coaching be improved? Etc.

### 4.3.4 Analyses

The evaluation of the pilot test is mainly based on the analysis of the variables: speed, acceleration, harsh braking and harsh cornering [6]. These variables will serve as input for the simulation of effects (See Chapter 5 of this deliverable). Because of the limited number of participating drivers, we do not expect to be able to determine statistical significance. The analysis will describe the potential (maximum) effect of training and qualitative assessment of the drivers' experiences when working with the app and coaching scheme.

To be able to estimate how many drivers could be needed in the trials, power analysis was applied to data similar to the type used in the WP4 evaluation: see Appendix A. These analyses are also applicable to WP5. As can be seen in Appendix

---

[6] If the app develops according to plan we can also analyse variables such as: close following, lane deviations, drowsiness/fatigue, and distraction.

A, power varies substantially between datasets and variables. It can be concluded that significant effects can possibly be achieved with the celeration variable, while variables with less favourable statistical properties will not yield statistically reliable results in WP4. The pilot tests will still provide qualitative (rather than quantitative) information about the effect on these variables. In the WP5 field trial we will aim for a sample size that provides enough statistical power.

## 4.4 Planning

*Table 4.1* shows the preliminary planning of the development and evaluation phases for the app and coaching in WP4. The version of the app that will be tested in the pilot tests is dependent on the progress of the development of the app. The results of this analysis should be available by April 2019, to serve as input for the simulation of effects (See Chapter 5 of this deliverable).

| | 2018 | | | | | | | | | 2019 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apr |
| | V0 – baseline measurement | | | | | | | | | | | | |
| Development (2-4 drivers) | | | V1 – Basic app and coaching | | | | | | | | | | |
| | | | | | V2 - With route traffic complexity, optionally with cameras (not real-time) | | | | | | | | |
| | | | | | | | V3 - additional features , optionally with real-time cameras | | | | | | |
| Evaluation (+ 10 new drivers) | | | | | | Baseline measurement | | | | | | | |
| | | | | | | | | | Pilot tests | | | | |
| | | | | | | | | | | Analyses | | | |
| | | | | | | | | | | | | | Results |

*Table 4.1 Preliminary planning of development and evaluation phases*

## 4.5 Conclusions and next steps

This chapter described the development and evaluation of a coaching system (including an app that provides feedback and peer-to-peer coaching) for truck drivers in MeBeSafe. During development, different versions of the app (with increasing sophistication) will be tested by a small group of 'early adopters'. The version of the app and coaching program that seems most promising from the developing phase will subsequently be pilot tested with a somewhat larger group of drivers (> 10) in a more controlled setting. Because of the limited number of participating drivers and absence of a control group we do not expect to derive hard statistically significant conclusions about the effectiveness of the app and coaching scheme, though it will give insights into whether the app and coaching schemes are ready for use in WP5.

The evaluation will provide an indication of the potential of the measures and a qualitative assessment by drivers (usefulness, feasibility). This information can be used in designing the final app and coaching scheme for the large field test in WP5. Indications on effectiveness can be used in power analyses to assess how much participants are needed for the large WP5 field test.

## 4.6 References

Adell, E., Várhelyi, A. & Hjälmdahl, M. (2008). Auditory and haptic systems for in-car speed management – A comparative real life study. In: Transportation Research Part F: Traffic Psychology and Behaviour, vol. 11, nr. 6, p. 445-458.

af Wåhlberg, A.E. (2006). Speed choice versus celeration behavior as traffic accident predictor. In: Journal of Safety Research, vol. 37, nr. 1, p. 43-51.

af Wåhlberg, A.E. (2007). Aggregation of driver celeration behavior data: Effects on stability and accident prediction. In: Safety Science, vol. 45, nr. 4, p. 487-500.

af Wåhlberg, A.E. (2008). Driver celeration behaviour and accidents – an analysis. In: Theoretical Issues in Ergonomics Science, vol. 9, nr. 5, p. 383-403.

af Wåhlberg, A.E. & Poom, L. (2015). An Empirical Test of Nonresponse Bias in Internet Surveys. In: Basic and Applied Social Psychology, vol. 37, nr. 6, p. 336-347.

Albert, G., Musicant, O., Lotan, T., Toledo, T., et al. (2011). Evaluating Changes in the Driving Behavior of Young Drivers a Few Years After Licensure Using In-Vehicle Data Recorders. In: Proceedings of th Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. p. 337-343.

Bell, J.L., Taylor, M.A., Chen, G.-X., Kirk, R.D., et al. (2017). Evaluation of an in-vehicle monitoring system (IVMS) to reduce risky driving behaviors in commercial drivers: Comparison of in-cab warning lights and supervisory coaching with videos of driving behavior. In: Journal of Safety Research, vol. 60, p. 125-136.

Blower, D. & Campbell, K.L. (1998). Fatalities and Injuries in Truck Crashes by Time of Day. UMTRI-98-48.

Blower, D., Campbell, K.L. & Green, P.E. (1993). Accident rates for heavy truck-tractors in Michigan. In: Accident Analysis & Prevention, vol. 25, nr. 3, p. 307-321.

Blower, D., Lyles, R., Campbell, K. & Stamatiadis, P. (1990). The Michigan Heavy Truck Study. Executive Summary. UMTRI Report 90-1-2. University of Michigan, Ann Arbor.

Bolderdijk, J.W., Knockaert, J., Steg, E.M. & Verhoef, E.T. (2011). Effects of Pay-As-You-Drive vehicle insurance on young drivers' speed choice: Results of a Dutch field experiment. In: Accident Analysis & Prevention, vol. 43, nr. 3, p. 1181-1186.

Boodlal, L. & Chiang, K.-H. (2014). Study of the Impact of a Telematics System on Safe and Fuel-efficient Driving in Trucks. Federal Motor Carrier Safety Administration report FMCSA-13-020.

Braver, E.R., Zador, P.L., Thum, D., Mitter, E.L., et al. (1997). Tractor-trailer crashes in indiana: A case-control study of the role of truck configuration. In: Accident Analysis & Prevention, vol. 29, nr. 1, p. 79-96.

Campbell, K.L., Blower, D.F., Gattis, R.G. & Wolfe, A.C. (1988). Analysis of Accident Rates of Heavy-Duty Vehicles. University of Michigan, Ann Arbor.

Cantin, V., Lavallière, M., Simoneau, M. & Teasdale, N. (2009). Mental workload when driving in a simulator: Effects of age and driving complexity. In: Accident Analysis & Prevention, vol. 41, nr. 4, p. 763-771.

Carney, C., McGehee, D.V., Lee, J.D., Reyes, M.L., et al. (2010). Using an Event-Triggered Video Intervention System to Expand the Supervised Learning of Newly Licensed Adolescent Drivers. In: American Journal of Public Health, vol. 100, nr. 6, p. 1101-1106.

Costigan, C.L. & Cox, M.J. (2001). Fathers' participation in family research: Is there a self-selection bias? In: Journal of Family Psychology, vol. 15, nr. 4, p. 706-720.

De Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H., et al. (2009). The Adaptation Test: The development of a method to measure speed adaptation to traffic complexity. In: Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. Big Sky Montana, USA.

Dingus, T.A., Guo, F., Lee, S., Antin, J.F., et al. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. In: National Academy of Sciences of the United States of America PNAS.

Donmez, B., Boyle, L.N. & Lee, J.D. (2007). Safety implications of providing real-time feedback to distracted drivers. In: Accident Analysis & Prevention, vol. 39, nr. 3, p. 581-590.

Duarte, G.O., Gonçalves, G.A. & Farias, T.L. (2013). Vehicle monitoring for driver training in bus companies – Application in two case studies in Portugal. In: Transportation Research Part D: Transport and Environment, vol. 18, p. 103-109.

FMCSA (2006). Report to Congress on the Large Truck Causation Study. Federal Motor Carrier Safety Administration, Washington DC.

Graf, V.D. & Archuleta, K. (1985). Truck Accidents by Classification. FHWA/CA/TE-85. California Department of Transportation, Sacramento.

Hanowski, R.J., Perez, M.A. & Dingus, T.A. (2005). Driver distraction in long-haul truck drivers. In: Transportation Research Part F: Traffic Psychology and Behaviour, vol. 8, nr. 6, p. 441-458.

Hickman, J.S. & Geller, E.S. (2005). Self-Management to Increase Safe Driving Among Short-Haul Truck Drivers. In: Journal of Organizational Behavior Management, vol. 23, nr. 4, p. 1-20.

Hickman, J.S. & Hanowski, R.J. (2011). Use of a video monitoring approach to reduce at-risk driving behaviors in commercial vehicle operations. In: Transportation Research Part F: Traffic Psychology and Behaviour, vol. 14, nr. 3, p. 189-198.

Hilburn, B. (2004). Cognitive Complexity in Air Traffic Control: A Literature Review. Center for Human Performance Research.

Jahn, G., Oehme, A., Krems, J.F. & Gelau, C. (2005). Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study. In: Transportation Research Part F: Traffic Psychology and Behaviour, vol. 8, nr. 3, p. 255-275.

Jovanis, P.P., Chang, H.-L. & Zabaneh, I. (1989). Comparison of accident rates for two truck configurations. In: Transportation Research Record, 1249, p. 18-29.

Jovanis, P.P. & Delleur, J. (1983). Exposure-based analysis of motor vehicle accidents. In: Transportation Research Record, 910, p. 1-7.

Larson, L.D., Schnelle, J.F., Kirchner, J., R., Carr, A.F., et al. (1980). Reduction of police vehicles accidents through mechanically aided supervision. In: Journal of Applied Behaviour Analysis, vol. 13, p. 571-581.

Lotan, T. & Toledo, T. (2006). An In-Vehicle Data Recorder for Evaluation of Driving Behavior and Safety. In: TRB Annual Meeting 2006.

Lotan, T., Toledo, T. & Prato, C.G. (2009). Modelling the behavior of novice young drivers using data from in-vehicle data recorders. In: Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. p. 491-498.

McGehee, D.V., Raby, M., Carney, C., Lee, J.D., et al. (2007). Extending parental mentoring using an event-triggered video intervention in rural teen drivers. In: J Safety Res, vol. 38, nr. 2, p. 215-227.

Mogford, R.H., Guttman, J.A., Morrow, S.L. & Kopardekar, P. (1995). The Complexity Construct in Air Traffic Control: A Review and Synthesis of the Literature. DOT/FAA/CT-TN95/22. US Department of Transportation.

Musicant, O. & Lampel, L. (2010). When Technology Tells Novice Drivers How to Drive. In: Transportation Research Record: Journal of the Transportation Research Board, vol. 2182, p. 8-15.

Musicant, O., Lotan, T. & Toledo, T. (2007). Safety correlation and implications of in-vehicle data recorder on driver behavior.

Nilsen, R.M., Vollset, S.E., Gjessing, H.K., Skjaerven, R., et al. (2009). Self-selection and bias in a large prospective pregnancy cohort in Norway. In: Paediatr Perinat Epidemiol, vol. 23, nr. 6, p. 597-608.

Olsen, R.L., Hanowski, R.J., Hickman, J.S. & Bocanegra, J.L. (2009). Driver distraction in commercial vechicle operations. FMCSA-RRR-09-042.

Rudin-Brown, C.M., Edquist, J. & Lenné, M.G. (2014). Effects of driving experience and sensation-seeking on drivers' adaptation to road environment complexity. In: Safety Science, vol. 62, p. 121-129.

Sagberg, F., Selpi, Piccinini, G.F.B. & Engström, J. (2015). A Review of Research on Driving Styles and Road Safety. In: Human Factors, vol. 57, nr. 7, p. 1248-1275.

Simons-Morton, B.G., Bingham, C.R., Ouimet, M.C., Pradhan, A.K., et al. (2013). The Effect on Teenage Risky Driving of Feedback From a Safety Monitoring System: A Randomized Controlled Trial. In: Journal of Adolescent Health, vol. 53, nr. 1, p. 21-26.

Soleymanian, M., Weiberg, C. & Zhu, T. (2016). The Value of Usage-Based Insurance beyond Better Targeting: Better Driving. Unpublished report.

SWOV (2016). Speed and speed management. SWOV Fact sheet, November 2016. SWOV, The Hague.

# 5 Simulation-based effectiveness research methodology

## 5.1 Introduction

Simulation-based effectiveness analysis has been applied in the past for different technologies, such as Advanced Driver Assistance Systems (ADAS) (e.g. Van Noort et al., 2015) as well as for first evaluations of automated driving systems (e.g. Fahrenkrog et al., 2017). The general idea is that computer simulations of technologies and interventions are performed to reduce actual real-world road testing for assessing the effectiveness of such measures. Note that the goal is not to replace actual real-world road testing completely (which is infeasible), but to reduce it, and to investigate in a controlled (i.e. computer simulated) setting road and traffic conditions which may be difficult or dangerous to investigate in real-world road testing (e.g. accident and near-accident situations, the possible impact of ADAS in those situations, etc.). The objective in the context of MeBeSafe is to explore to which extent the method can be applied for new traffic safety relevant measures such as nudging and coaching, as well as what updates are required to analyse these measures.

In WP4 the focus is on adapting simulation-based effectiveness analysis for coaching. This requires updates of the existing models that are already applied today in simulation tools – in particular with respect to driver modelling. In the end, the simulation-based effectiveness analysis will be used to estimate the impact of the coaching measures in terms of traffic safety. Here, we must take into account the fact that the real-world study with the coaching app is conducted with a limited number of test drivers. Due to this it is unlikely (and not to be hoped) that a statistically relevant number of accidents will be detected. The simulation should allow us to overcome this issue and scale up the results to a larger (simulated) population, comparing simulated accidents for one population to simulated accidents for another population. The coaching with respect to ACC is not the focus for the simulation based assessment.

## 5.2 Development methodology

In general, the question regarding the impact of a certain technology in terms of traffic safety is relevant for different stakeholders (politics, car manufacturers and suppliers, insurers as well as researchers). The classical approach is to determine the impact by means of analysing accident statistics, comparing accident prevalence with and without certain technology (e.g. Farmer, 2004, Unselt, 2004). However, this requires that the technology has already reached a certain penetration level as well as that no effects of any other technology interfere with the effects of the technology under assessment. In order to investigate the safety impact already before the market introduction or at a low penetration rate, other methods need to be applied.

Basically, there are four different approaches that can be applied in such prospective effectiveness analysis: determine the field of application by a high-level analysis of accident data (e.g. Kocherscheidt, 2004), studies in controlled environments (e.g. Breuer, 2009), Field Operation Tests (e.g. Malta et al., 2012), and simulations (e.g. Helmer, 2014). Among these approaches the simulation approach is the one which allows investigation in detail of many different driving situations at reasonable cost, although conformity with the real world needs to be assured (i.e. the extent to which the simulation is sufficiently realistic). The other approaches are either limited in the level of detail of the analysis or require (much) larger resources. Therefore, the simulation based approach is selected for the analysis in MeBeSafe WP4.

Within the simulation based effectiveness analysis, basically three different approaches are known (Alvarez, 2017):

- Re-simulation of real world traffic situations (accidents or safety critical driving situations);

  - Modified simulation of real world traffic situations (accidents or safety critical driving situations);
  - Simulation of synthetic cases based on relevant characteristics of real world traffic situations.

For the simulations that are planned in WP4 the last approach, "Simulation of synthetic cases based on relevant characteristics of real world traffic situations" – also called stochastic traffic simulation approach - is chosen, since it offers the opportunity to extend the simulation in time and space arbitrarily. This aspect is of particular importance, since coaching effects influence the driving behaviour not only in a particular driving situation, but throughout the whole drive. For the other two simulation based approaches, this type of extension of the simulation is not possible, due to their nature of being necessarily linked to specific, limited duration real world traffic situations.

In the traffic based simulation approach, relevant parameters are modelled by means of statistical probability distributions, which are derived from different traffic data sources. Sampling methods, for example Monte Carlo simulations, are used to vary the characteristics of these parameters that cover, among others, the characteristic of the simulation agents (combination of driver and vehicle) as well as traffic and environmental variables (Helmer, 2014). Since the traffic based simulation approach does not consider explicitly recorded or reconstructed trajectories of the simulated vehicle (derived from real-world measurements or estimates), the movement of the vehicle needs to be determined depending on the given driving situations and the surrounding vehicles. The task to determine the movement of the agents is fulfilled by the driver behaviour model in conjunction with the vehicle model.

For this purpose the integrated Stochastic Cognitive Model (SCM) for highway driving has been developed within BMW. A core aspect of the SCM driver behavioural model is the application of stochastic methods in order to represent the behaviour of different drivers. The SCM consists of five different sub-models that are briefly described in the following, see Figure 5.1 (Wang et al., 2017).

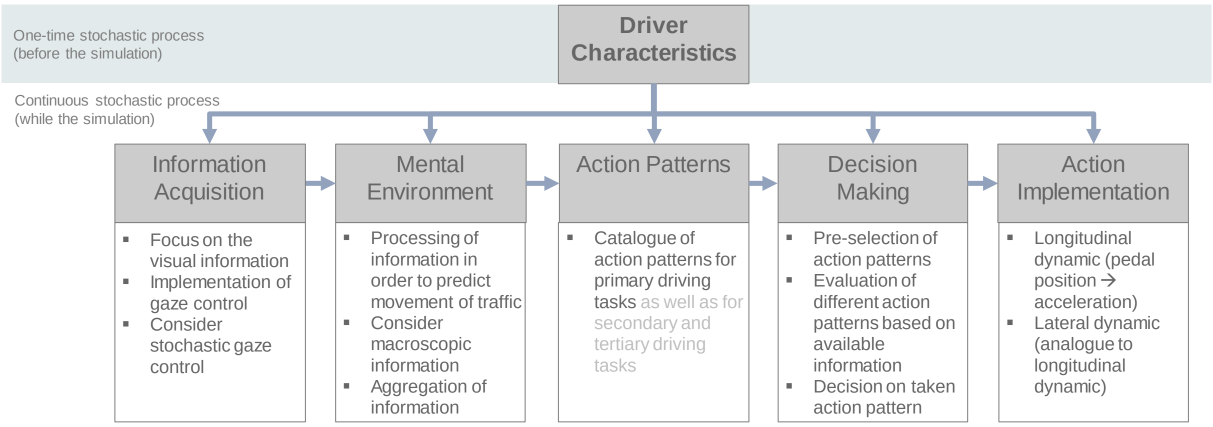

*Figure 5.1: Structure of the SCM-driver behaviour model of BMW.*

*Information acquisition*: This sub-model considers in principle auditory, haptic and visual perception of the driver. In particular the information acquisition sub-model is focused on visual perception, which considers the peripheral and foveal field of view of the driver as well as the gaze distribution.

*Mental environment*: This sub-model describes recognition of situation patterns. This sub-model considers the current information of the information acquisition sub-models, as well as information from memory (data from previous time steps). All gathered information is aggregated to describe the microscopic traffic properties and to extract features of the environment that are needed in the decision making model.

*Decision making*: In this sub-model the current situation is assessed according to the information derived in the previous step. Based on the outcome of the assessment a decision is taken about the next action. Statistical variations are also considered for the selection of the action to be taken.

*Action Patterns*: This sub-model classifies the action to be taken -based on an action pattern catalogue- into primary (acceleration, deceleration, steering and constant driving), secondary (indicator use, light activation, use of the horn etc.) and tertiary driving actions (e.g. telephone or navigation use).

*Action Implementation*: Finally, the information of the previous sub-models is used in order to determine the pedal position – accelerator as well as braking pedal – and steering wheel angle, which together result in the longitudinal and lateral acceleration of the vehicle. By this the movement of the vehicle for the next time step can be determined.

## 5.3 Evaluation methodology

The evaluation methodology in the simulation-based effectiveness analysis for the coaching measures in WP4 is given in Figure 5.2.

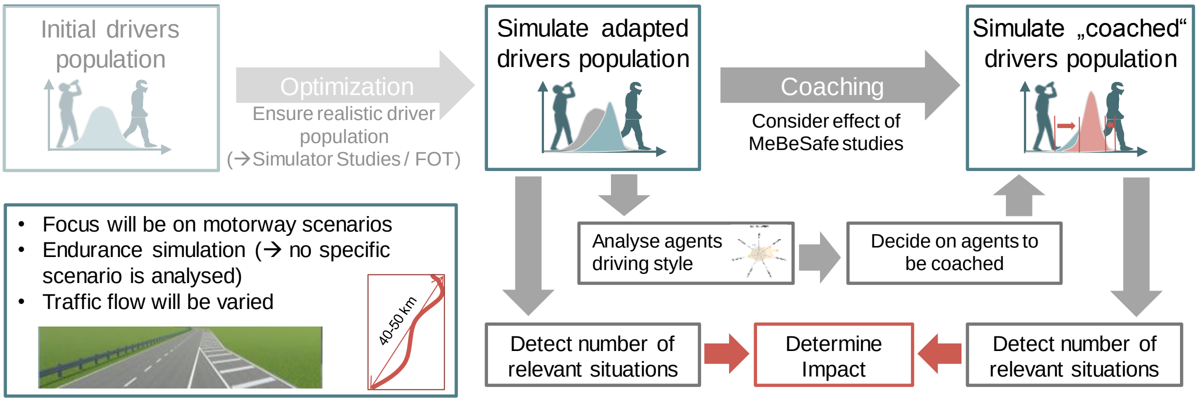

*Figure 5.2: Evaluation methodology in the simulation-based effectiveness analysis for the coaching measures.*

The starting point is an initial agent population that consist of several drivers and vehicles, which are chosen randomly based on the underlying distributions for the agents' characteristics. Up to now, it has not been decided whether only truck drivers, only passenger drivers, or both, are considered in the driver type to be modelled. Depending on the analysis of feasibility of these options, the driver population will need to be adapted accordingly. Once the distributions are defined, a first set of simulations is conducted.

The simulated scenario will be a motorway section that will be adapted, to the extent that that is possible, to the region in which the real-world study with the coaching app takes place. Parameters that are taken into account are for instance the number of lanes, speed limits or frequency of motorway entrances and exits.

The first set of simulations serves as the baseline for the evaluation, and represents the driver behaviour before the coaching. The simulation runs are analysed with respect to the number of relevant situations, which can be accidents as well as safety critical events, which arise over many hours of simulated driving. Furthermore, for each agent its driving behaviour is rated by means of the traffic safety wheel[7] (see Wesseling et al., 2018). Based on the profiling by the traffic safety wheel, agents to be coached will be identified. For the coached agents the driver characteristics will be changed as indicated by the results of the coaching app study.

In the last step a second set of simulations with the coached agent population will be conducted. Also for this set of simulations relevant driving situations will be identified. The impact of coaching in the broader sense will be determined by investigating the number of relevant driving situations as well as their severity, the hypothesis being that coaching may lead to somewhat safer driving behavior and therefore, in the long run (many hours of simulated driving), fewer safety critical events or accidents in the simulations.

## 5.4 Conclusions and next steps

In this chapter the approach taken for the scaling up of results in order to determine the safety impact of coaching measures as considered by MeBeSafe has been described. For this purpose the existing simulation tool – in particular the driver behaviour model –needs to be enhanced.

---

[7] For the traffic safety wheel slight adaptation might be required due to the difference between real world data and data generated in the simulation.

In the next steps the existing driver behaviour model will be adapted. Once the results of the coaching app study are available, they will considered in the simulation model and the necessary simulations will be conducted as described above. In the last step the results of the simulation will be summarized and reported.

## 5.5 References

Alvarez, S., Page, Y., Sander, U., Fahrenkrog, F., Helmer, T., Jung, O., Düring, M. Döring, S., Op den Camp, O. (2017). Prospective Effectiveness assessment of ADAS and active safety systems via virtual simulation: a review of the current practices, 25th Enhance Safety of Vehicles Conference, Detroit.

Breuer, J. (2009), Bewertungsverfahren von Fahrerassistenzsystemen, Handbuch Fahrerassistenzsystem, Vieweg+Teubner, 1. Auflage, 55 -68.

Fahrenkrog, F., Wang. L., Rösener, C., Sauerbier, J., Breunig, S. (2017). Impact analysis for supervised automated driving applications, AdaptIVe Deliverable D7.3.

Farmer, C.M. (2004). Effect of Electronic Stability Control on Automobile Crash Risk Traffic Injury Prevention 5, 317–325.

Helmer, T. (2014). “Development of a Methodology for the Evaluation of Active Safety using the Example of Preventive Pedestrian Protection,” Dissertation, TU Berlin.

Kocherscheidt, H. (2004). Wege zur Effizienzbetrachtung von Fahrerassistenz-systemen, 1. Tagung Aktive Sicherheit durch Fahrerassistenz, München.

Malta, L., Ljung Aust, M., Faber, F., Metz, B., Saint Pierre, G., Benmimoun, M., Schäfer, R. (2012). Final results: Impacts on traffic safety, euroFOT Deliverable D6.4.

Unselt, T., Breuer, J., Eckstien, L., Frank, P. (2004) Avoidance of „loss of control accidents“ through the benefit of ESP, FISITA 2004 World Automotive Congress, Barcelona.

Van Noort, M., Taoufik, B., Fahrenkrog, F., Dobberstein, J. (2015),SIMPATO - The Safety Impact Assessment Tool of interactIVe, IEEE Intelligent Transportation Systems Magazine, Volume: 7, Issue: 1, 80 – 90.

Wang, L., Fahrenkrog, F., Vogt, T., Jung, O., Kates, R. (2017). Prospective safety assessment of highly automated driving functions using stochastic traffic simulation, 25th Enhance Safety of Vehicles Conference, Detroit.

Wesseling, S., de Craen, S., Jansen, R., Bakker, B., Neuhuber, N., Moertl, P., Adams, P., Baker, A., van Schagen, I., van Nes, N., Mons, C., Lillington, T. (2018). Driver & situation profiling, MeBeSafe Deliverable D4.1.

# 6 Concluding remarks

This document described the research (and coaching) methodologies employed in the various tasks of Work Package 4. Each of the research methodologies has been defined, and where necessary aligned.

Next, we will continue the development of the coaching schemes and apps and the simulation framework. When that has been completed, we will proceed to the small-scale evaluation planned within WP4. The results of that small-scale evaluation (real-world pilot and simulations) will be used to realise final changes and finetuning for the coaching schemes and app versions going into the large-scale field tests of WP5.

A point of concern is that within WP4, we will not be able to do pilot testing with many drivers, meaning that we cannot use statistically reliable results to guide decisions about the final coaching schemes and apps in WP5. However, we will be able to measure effects and assess them on a more qualitative level, such that the pilots and simulations will give insights into whether the apps and coaching schemes are ready for use in WP5.

# Appendix A

The variables included in the evaluation of HGV driver coaching will have very different statistical properties, and to ensure that sample sizes are sufficient, and the results correctly interpreted, power analysis was undertaken on some available data.

The first question which needs to be asked is how large an effect can be expected from the intervention under the circumstances known to be at hand in the participating companies. Given that participants will probably work for companies which have an interest in safety, and it is known that they already have telematics systems in place in order to enhance safe behaviour, the effect of the coaching system is suspected to be limited.

It is also suspected that drivers participating voluntarily in the intervention will be biased towards the best drivers in the company (Costigan & Cox, 2001; Nilsen et al., 2009), although such differences need not always be large (af Wåhlberg & Poom, 2015). In sum, a bias toward a selection of the best drivers in our intervention group could result in effect sizes for a normal population being under-estimated.

It can then be asked what a standard effect size for coaching and feedback to drivers could be? No meta-analysis has investigated this, and the literature is very fragmented, with different researchers using very different types of populations, feedback, coaching and evaluation methods, all of which probably influence the effect sizes. Examples of studies similar to WP4 include Bolderdijk et al. (2011) where speeding[8] was reduced by nine percent, Hickman and Geller (2005) where extreme braking was reduced (two effect sizes; d=0.35, 1.6), Musicant and Lampel (2010) where 'events' where reduced by 59 percent, Musicant, Lotan and Toledo

---

[8] Speeding is a variable where it is easy to achieve large effects with very small actual changes in speed. See for example Lai, Jamson and Carsten (2012), where speeding was reduced 56 percent with a four percent decrease in mean speed. The same problem exists for harsh braking.

(2007) with a 44 percent reduction in collisions, and Carney et al. (2010) where 'coachable events' were reduced by 61 percent. These effects may seem impressive, but they were achieved mainly with teenagers, where the potential is large. Also, some of the dependent variables were of doubtful validity, and the methodology not always of high quality. See further Hickman and Hanowski (2011), Larson et al., (1980), McGehee et al. (2007), Simons -Morton et al. (2013) and Lotan and Toledo (2006).

Table 1 shows power calculations on some of the variables which will be included in the evaluation. Here, it has been assumed that the effect for the highly experienced drivers of Gasnor/Litra will be small. The data used is from different sources, and the values may not be applicable to Norwegian truck drivers, but it should be apparent that different variables have very different statistical properties, and that significant effects cannot be expected for all of them. It can be noted that the standard deviations for the celeration variables in the table are much smaller than could be expected in the WP4 trial, as they were gathered on standard routes.

| Source of data | Variable | Mean (before) | Std | N/Power | N/Power |
|---|---|---|---|---|---|
| van Waveren (truck drivers) | Harsh brake distance | 3.42 | 1.55 | 50/20% | 500/80% |
| Uppsalabuss (bus drivers) | Celeration | 0.466 | 0.062 | 20/50% | 100/98% |
| Lindkvist (truck drivers) | Celeration | 0.258 | 0.059 | 50/46% | 150/86% |
| Wassink (truck drivers) | Harsh brake distance | 5.84 | 6.82 | 100/11% | 500/25% |
| Wassink (truck drivers) | Brake distance | 31.61 | 17.08 | 100/24% | 500/68% |

*Table 1: The estimated power of the evaluation at two different sample sizes, given an expected effect of five percent improvement/reduction. Assumed correlation between measurements .50, p<.05.*