Delivery Report for

MeBeSafe

Measures for behaving safely in traffic

Deliverable Title	Report on effective feedback
Deliverable	D4.5
WP	WP4 Driver coaching
Task	Task 4.4 Data back-end – evaluation of coaching schemes

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

Disclaimer

The opinions expressed in this document reflect only the author's view and reflects in no way the European Commission's opinions. The European Commission is not responsible for any use that may be made of the information it contains.

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	D4.5
Deliverable Leader (according to GA)	Volvo Cars
WP Leader	Saskia **de Craen**, Shell
Deliverable Leader(s)/ (Editor(s))	Saskia **de Craen**, Shell
Dissemination Level (Confidentiality)	Public
Nature	Report
Status	Final
Due Date	M28 (Aug 2019)
Main Author(s)	Saskia **de Craen**, Shell / SWOV
Other Contributing Author(s) *alphabetical order by lastname	Felix **Fahrenkrog**, BMW Teri **Lillington**, Shell Mikael **Ljung Aust**, Volvo Cars Charles **Smeets**, Shell Anders af **Wåhlberg**, Cranfield University
Reviewer(s)	Olaf **Op den Camp**, TNO
Formal review and upload by Coordinator	Institute for Automotive Engineering (ika) – RWTH Aachen University

Please refer to this deliverable as follows:

De Craen, S. et al. (2019). Report on effective feedback (Deliverable 4.5). Retrieved from MeBeSafe website: https://www.mebesafe.eu/results/.

Abstract
This deliverable describes the results of the pilot test with the coaching scheme and DriveMate app for Heavy Good Vehicle (HGV) drivers and the app (to increase ACC use) for Volvo drivers. The coaching scheme for HGV drivers consists of an online (app based) and offline (face-to-face coaching) part. Because of the unfortunate delay in the development of the app, the pilot test was very limited in scope, and no face-to-face coaching was initiated. The analysis of the preliminary data collected with the DriveMate app does seem to indicate that, with the exception of some errors, the system is generally working as planned concerning the data gathered. The DriveMate app (as part of the coaching scheme for HGV drivers) needs to be improved considerably before it can be used in the field trial. With the current (V1) version we expect to achieve only a small effect of online and offline coaching, which is not expected to show up in the field trial. The further development of the app is dependent on a pending amendment request. For the pilot test with Volvo drivers, a collaboration between the MeBeSafe pilot test and another (in-house) Volvo project called In-Car Test Drive was set up. An Adaptive Cruise Control (ACC) activation coaching function was developed, and the feedback and experiences from customers involved in the In-Car Test Drive pilots were analysed. The results show that the app based coaching was highly successful in terms of coaching drivers to use ACC. First time usage was accomplished for a number of individuals who would otherwise never have tried to activate ACC. However, it also became very clear that the app itself did not provide a sufficiently robust, natural and trustworthy interaction for drivers with limited interest in new technology and in activating functions like ACC (in other words, the intended target group for ACC coaching in MeBeSafe). As a much more sophisticated app design would be required to overcome those difficulities, it was

decided that the best way forward would be to employ a Wizard of Oz-approach in the field trial. This makes it possible to understand to which degree the target group of non ACC users are coachable into ACC usage without spending a prohibitively large sum of money on further app development first.

Version	Date	Comment
1	15 April 2019	First outline send to co authors
2	09 July 2019	Draft for WP4 partners
3	15 July 2019	Draft for final check partners
4	19 July 2019	Reviewed document with comments
5	23 August 2019	Reworked version for reviewer
6	29 August 2019	Final version with okay from reviewer
7	29 August 2019	1^{st} formal review ika

table of document history

Table of Contents

List of Figures

List of Tables

Acronyms

ACC	Adaptive Cruise Control
Coaching	The social/educational interaction between two or more individuals with the aim of improving the coachee's performance in some area of endeavour
DriveMate	The name given to the MeBeSafe truck driver coaching app
HGV	Heavy Goods Vehicle
IVMS	In Vehicle Monitoring System
Offline coaching	The term used in the MeBeSafe proposal for human interaction for improved driver behaviour
Onboarding	The prepatory information about coaching techniques given to the drivers in the DriveMate app before the actual coaching sessions are started
Online coaching	The term used in the MeBeSafe proposal for the actions of the DriveMate app and Volvo's in-car driver support
PA	Pilot Assist
Safety topic	A traffic safety theme (like fatigue in driving) for discussion in the truck driver coaching session, suggested by the app to the drivers. Also, the phrase is used for the information contained in the app which is used by the drivers as a basis for discussion
V0	The first version of the Drivemate app delivered to the truck drivers, where no coaching sessions are undertaken. Measurements are made and data gathered, but no information or feedback is given. On-boarding sessions are delivered approximately every two days, eighteen times. Onboarding is in this case short texts about coaching techniques
V1	The first version of the app where coaching is introduced

V2	A more advanced version of the app, with control for road complexity in the feedback and more coaching functions, and possibly with the inclusion of video in various forms

1 Executive Summary

This deliverable describes the results of the pilot test with the coaching scheme and app for Heavy Goods Vehicle (HGV) drivers and the app (to increase ACC use) for Volvo drivers.

1.1 Coaching scheme for HGV drivers

Unfortunately, we have encountered some delays in the development of the DriveMate app (for HGV drivers).[1] As a result we only had a first version of the app available for pilot testing; which also included some technical flaws (see *D4.4. – App to induce behaviour change*). This delay in the app development impacted the pilot test in WP4 considerably, and the results of the pilot are therefore not a full evaluation of the effect of the implemented system, but rather a proof of concept and a technical trial for the app.

1.1.1 Pilot test

The coaching scheme for HGV drivers consists of an online (app based) and offline (face-to-face coaching) part. The face-to-face coaching is supported and initiated by the DriveMate app. Because of the delay in development, the app was administered to only 13 drivers in our pilot with a Norwegian truck company. Data was collected for a relatively short period of time, and no face-to-face coaching was initiated.[2] The analysis of this first data from MeBeSafe WP4 does seem to indicate that, with the exception of some errors, the system is generally working as planned concerning the data gathered.

[1] This was reported to the PO in November 2018. Because of a reorganisation, Shell – resonsible for the development of the app – no longer has the internal capability for programming software for apps such as the DriveMate app. The programming of the app needs to be out-sourced. For this an amendment is neccesary.

[2] The app is supposed to prepare drivers for coaching, and alert them when a coaching session is due. Because of the technical issues with the app the drivers never received such an alert.

However, the DriveMate app needs to be improved considerably before it can be used in the field trial. With the current (V1) version we expect to achieve only a small effect of online and offline coaching, which is not expected to show up in the field trial. The development of the app is dependent on the pending amendment request.

1.1.2 Simulation

The simulation to estimate the large-scale impact of the coaching scheme on road safety, was also impacted by the delay in the development of the DriveMate app. Because of the delay, no estimated effect of coaching was available as input for the simulation. Instead, different scenarios based on expert opinion were used in order to quantify the effect of coaching. The results of this simulation show a safety benefit in case coaching affects the desired time gap and the desired velocity. For the other parameters (mean velocity, time driving with short distances, time driving in critical situations, frequency of hard braking manoeuvres) the simulations also show benefits. Furthermore, from this analysis, it can be concluded that it is in general possible to determine the safety impact of coaching measures by means of the simulation approach. However, during the work also different limitations were encountered.

1.1.3 Business case

The DriveMate App and coaching scheme have the potential to reduce the loss of human lives. However, as well as reducing the number of road safety incidents and fatalities, the DriveMate app has several business prospects. The most likely business case is using the DriveMate app as a replacement for hardwired In Vehicle Monitoring Systems (IVMS). The DriveMate App can additionally nudge drivers to reduce fuel consumption and potentially reduce maintenance costs, through smoother driving (with less harsh braking and harsh acceleration) .

1.2 Pilot test with Volvo drivers

For the pilot test with Volvo drivers, a collaboration between the MeBeSafe pilot test and another (in-house) Volvo project called In-Car Test Drive was set up. The In-Car Test Drive project aims to help potential new car buyers explore vehicle features while taking a test drive in a dealer setting by means of an in-vehicle coaching app. In addition to other features, an ACC activation coaching function was developed, and the feedback and experiences from customers involved in the In-Car Test Drive pilots when using this function were analysed. The downside of this setup was that it is not possible to single out the intended target population for ACC coaching beforehand, since the pilots relied on drop-in customers willing to participate in the study. On the positive side, the pilots covered many more participants, in more markets (Sweden, England, US) and with a more sophisticated app than would have been possible for MeBeSafe alone.

The pilot results show that in terms of coaching drivers to use ACC, the in-vehicle app based coaching was highly successful, in the sense that a first time use situation was accomplished for a number of individuals who would otherwise never have tried to activate the function. However, it also became very clear that the app itself did not provide a sufficiently robust and natural interaction for all users. Drivers with an interest in technology in general, and a genuine interest in trying ACC out in particular, had no problems following the app instructions. On the other hand, drivers with limited interest in new technology and in activating functions like ACC (in other words, the intended target group for ACC coaching in MeBeSafe) had great difficulties understand what the app was asking them to do. Analysis of those interaction difficulties showed that a much more sophisticated app design would be required to overcome those difficulties, including highly refined speech processing and dialogue design, allowing for open ended questions and answers. Several also reported that activation felt scary, even with a test leader in the vehicle, and had a number of follow up questions on what the function could and could not do. It is therefore not certain that they actually would have activated ACC if the test leader had not been present in the vehicle.

Based on these results, it was concluded that the best way forward for the ACC coaching objective in the field trial is to employ a Wizard of Oz-approach. In this approach, the test participant is lead to believe that they are interacting with a computer based function of some type, while in reality an experimenter (the "wizard") is simulating the behaviour of the application. This will make it possible to have a coaching dialogue that does not fall through due to technical limitations in speech and dialogue processing, and hence makes it possible to understand to which degree the target group of non ACC users are coachable into ACC usage without spending a prohibitive sum of money on app development first.

2 Contribution by each Partner

This deliverable was compiled by Saskia de Craen, Shell / SWOV from contributions by Anders af Wåhlberg, Cranfield University (*Chapter 4: Pilot test with coaching scheme for HGV drivers*), Felix Fahrenkrog, BMW Group (*Chapter 5: Results of simulation of the coaching scheme for HGV drivers*), Charles Smeets and Teri Lillington, Shell (*Chapter 6: Business case of the coaching scheme for HGV drivers*) and Mikael Ljung Aust, Volvo Cars (*Chapter 7:* Pilot test with Volvo drivers).

All other partners in WP4 contributed by giving their feedback on the development of the coaching scheme and app; and by reviewing this deliverable. All partners fulfilled their tasks in satisfactory time and quality.

3 Introduction

Following the description in the GA, this Deliverable is meant to "outline the detailed protocols on exactly what feedback works most effectively when given to drives in what specific situation / context / frequency". However, this WP4 will also deliver two other reports (*D4.3 – Final coaching scheme*; and *D4.4. – App to induce behaviour change*). We decided to focus these two deliverables on the *description* of the coaching scheme and supporting app (i.e. the detailed protocols); and focus the current deliverable on the *results* from the pilot tests.

Unfortunately we have encountered some delays in the development of the DriveMate app (for HGV drivers).[3] As a result we only had a first version of the app available for pilot testing; which also included some technical flaws (see *D4.4. – App to induce behaviour change*). This delay in app development impacted the pilot test in WP4 considerably.

The DriveMate app for HGV drivers was pilot tested with a Norwegian truck company, to study the robustness and functionality of the system (see *Chapter 4*). Due to the delay and technical issues in the implemented version, the app was administered to only 13 drivers. Data was collected for a relatively short period of time; and no face-to-face coaching was initiated.[4] The results reported in *Chapter 4* are therefore not a full evaluation of the effect of the implemented system, but rather a proof of concept and a technical trial for the app.

The pilot test of WP4, but also the large scale field trial in WP5, are conducted with a limited number of test drivers. It is rather unlikely that any crash will occur during the

[3] This was reported to the PO in November 2018. Because of a reorganisation, Shell – resonsible for the development of the app – no longer has the internal capability for programming software for apps such as the DriveMate app. The programming of the app needs to be out-sourced. For this an amendment is neccesary.

[4] The app is supposed to prepare drivers for coaching, and alert them when a coaching session is due. Because of the technical issues with the app the drivers never received such an alert.

field trial. To assess which effect, in terms of traffic safety, can be expected from the application of coaching HGV drivers in traffic, BMW conducted a simulation to scale up the results to a larger (simulated) population (*Chapter 5*). This simulation-based effectiveness analysis will be used to estimate the impact of the coaching measures on road safety. Originally, we aimed to use the results from the pilot test (i.e. the estimated effect of coaching) as input for the simulation. However, because of the delay in the app development, the first pilot test could not meet the requirements for the results to be used reliably as input for the simulation. Therefore, the coaching assumptions have been made based on expert judgement, how the single agent characteristic would change under the influence of coaching.

To increase the chances of succesful implementation of the HGV coaching in the industry, a business case was developed and described in *Chapter 6*.

In addition to the online and offline coaching scheme for HGV drivers, WP4 also developed a online (app based) coaching scheme to increase ACC use amongst Volvo drivers. The pilot tests (*Chapter 7*), were carried out in collaboration between MeBeSafe and another (in-house) Volvo project called In-Car Test Drive. An ACC activation coaching function was developed, and the feedback and experiences from customers involved in the In-Car Test Drive pilots were analysed. In particular, their response and behaviour when being prompted to activate ACC were of interest.

4 Pilot test with coaching scheme for HGV drivers

4.1 Introduction

In MeBeSafe WP4, a peer-to-peer coaching scheme for truck drivers, delivered and supported by a behaviour-measuring app, was developed (See *MeBeSafe Deliverable 4.3 - Final coaching scheme* (af Wåhlberg et al., 2019). This system was implemented in a pilot trial in a Norwegian truck company, to test the robustness and functionality of the system, which is reported in the present deliverable.

Due to delays in the development of the app, and technical issues in the implemented version, the pilot test started much later than planned. The available amount of data for this deliverable was therefore small, both in terms of number of drivers and the amount of data per driver. This deliverable is therefore not a full evaluation of the effect of the implemented system, but rather a proof of concept and a technical trial for the app. The coaching component was not started within this trial period. The participating drivers received onboarding sessions on their phones, where drivers are instructed by the app on how to do coaching. But no face-to-face coaching sessions were organized. The responses of the drivers to the onboarding sessions were not measured in this trial, although preparations for this have been made in the form of a survey to be administered in the field trials.

4.2 Method

4.2.1 General setup of the pilot trial

The Shell contracted haulier, who employs the drivers participating in the pilot test, delivers liquid gas by trucks and boats to customers over the whole area of Norway. Road conditions are variable due to the quickly shifting landscape and weather, and therefore the driving is very challenging. Roads are winding up and down mountains, where the weather can change from fog to snow storms in half an hour's drive. For the app developers, the many tunnels (where GPS connection is usually lost) and travel on ferries, were challenges of a kind which is uncommon in other countries. The driving undertaken by the drivers can also differ a lot between days, as the company has deliveries within for example the city of Bergen, which means several hauls can be undertaken in a day, while others are long distance, and a single haul might take several days.

4.2.2 The DriveMate app

The app developed within WP4 runs on Android phones, and is described in detail in *MeBeSafe Deliverable 4.4 – App to reduce behavioural change* (Varlamov et al., 2019). To ensure compatibility of the phones used with the app, drivers were issued with new phones from the project. For the current deliverable, the important part of the setup was that the phone records data when the drivers start the app and the trip recording function. This data is used both for calculating trip summary values which are shown to the drivers after each trip and for research. The phone does not connect to the CAN-bus of the vehicle, but instead relies only on its internal sensors. For the present purpose, only the variables derived from GPS position and time are used, in addition to the data entered manually by the drivers after each trip.

Some consequences of this setup for the data can be noted; a trip can have any length, because a driver can end a trip whenever he has stopped, and start a new one. If the driver stops and then starts again without ending the trip, the trip is automatically resumed. Furthermore, as the driver needs to start the phone, the app and the trip recording, it is possible that some trips are not recorded. The data will therefore not be totally comprehensive, which is a consequence of the principle of the drivers being in control of the data acquisition, as described in *MeBeSafe Deliverable 4.3 – Final coaching scheme* (af Wåhlberg et al., 2019).

The version of the app issued to the drivers was V1, which is a basic version with trip recording and coaching functions. As described in Deliverable 4.3 and 4.4, the full functionality of the planned V2 would include much more advanced functions. In the

WP5 field trial, the drivers will first be issued with V0, which does not yield any trip statistics, but only record trips and send onboarding information (see next section). The data from this version make up the baseline, to which the data from after the intervention has started will be compared.

4.2.3Coaching training in the app

The DriveMate app issues the drivers with two different kinds of instructions about coaching. First, the drivers are given a PowerPoint presentation which includes the purpose and approach of the coaching scheme and instructions on how to coach. Second, there are eighteen documents presented in the DriveMate app, which contain the onboarding information. These are coaching instructions, i.e. a short course on how to do peer to peer coaching, including techniques from cognitive-behavioural therapy. These instructions are presented one by one, with at least a two day period in between. This means that before actual coaching sessions are started, there is a period of at least five weeks where drivers are recording trips but not getting any coaching. This is the pre-intervention period to which later data is to be compared. It can be noted that this method does not control for seasonal variations, for which data of at least a full year is needed.

4.2.4Drivers and timeline

Four drivers from the haulier were introduced to the DriveMate app and coaching in December 2018 by two members of the WP4 team. In March 2019, a further nine drivers from the haulier were issued with phones and instructions on how to use the app by their supervisor. Meanwhile, the four original drivers had to start new app accounts (identities), due to a technical problem which led to their data not being transferred into the research database. Data for the present deliverable was therefore gathered from March to June 2019.

4.2.5 Variables

The variables extracted for this deliverable included the three summary trip information parameters on driver behaviour (harsh braking, harsh acceleration and smoothness) all of which had a range of zero to one hundred percent, with the latter being good. Furthermore, three variables were manually recorded by the drivers after each trip on five step scales; state of the road (dry to snowy), amount of traffic (hardly any to congestion) and load (none to full). These variables are currently used to annotate the trips, and are shown with the trip data to the drivers. They are intended as possibly being included in more advanced algorithms for driver feedback, if it can be shown that they do yield useful information. Also, the total distance and the time taken for the trips were recorded. Finally, from the raw data the mean speed and celeration of the trip were calculated (the average of the absolute acceleration values when the vehicle is moving; af Wåhlberg, 2007a; 2008).

4.3 Results

4.3.1 Data

Data had been gathered from nine drivers from the haulier in Norway, while four drivers did not seem to have recorded any data. They were all males, between forty and sixty years of age, with extensive experience of truck driving. The number of trips recorded differed strongly between the drivers (from 7 to 34 and 100). For the current purpose, a maximum of twenty trips were analysed per driver.

4.3.2 Descriptive results

Average values and standard deviations of those can be seen in Table 4.1. The values would seem to be reasonable for truck driving in Norway (although it should be remembered that the harsh braking and acceleration cut-off values are arbitrary). The celeration values are about one third of those for bus drivers in a city environment (af Wåhlberg, 2007a), and the trips mainly more than 100 kms. It can be noted that mean speed was calculated in the same way as celeration; as the average when the vehicle is actually moving, not as the distance divided by the time for getting from A to B.

The variable duration of the drive turned out not to be reliable, as some values (deleted from Table 4.1) exceeded what was possible. As this variable is not intended for any evaluation analysis, such an error is not crucial, but should be investigated in V2 development, as it might indicate a problem which influences other measurements and calculations.

Driver	1	2	3	4	5	6	7	8	9
Number of trips (outliers deleted)	20	13	19	7	20	7	9	17	20
Duration (hours)	12.26/6.43**	11.11/9.21	1.11/0.51	22.32/27.27	5.20/4.50	8.45/4.71*	116.74/169.58**	8.01/6.53	2.90/2.40*
Distance (km)	326.6/164.1	417.5/331.3	58.3/31.0	667.8/301.0	303.0/273.1	345.4/175.8	429.9/298.8	267.4/193.2	87.2/58.7
Traffic conditions(5-1)	3.50/0.61	3.92/0.28	2.58/0.84	3.57/0.79	4.35/0.81	3.17/0.80	2.50/0.53	3.82/0.64	3.05/1.05
Load conditions (5-1)	2.55/1.91	2.77/1.92	1.89/1.15	3.86/1.57	2.60/2.01	2.67/1.00	3.75/1.28	3.00/1.97	1.80/1.24
Road conditions (5-1)	4.80/0.41	4.08/0.86	4.58/0.51	4.14/0.38	5.00/0	4.33/0.50	3.12/146	4.88/0.33	4.15/0.81
Smooth score (%)	99.76/0.52	99.99/0.03	99.76/1.04	99.79/0.20	99.77/0.52	99.63/0.80	99.83/0.29	99.80/0.75	99.80/0.52
Harsh acceleration score (%)	98.45/3.37	9936/0.67	98.88/2.49	98.39/2.47	98.61/3.58	93.96/13.40	96.10/4.70	99.64/0.59	96.23/6.38
Harsh braking score (%)	89.26/12.10	95.06/7.89	92.27/9.42	83.51/8.14	86.77/15.63	84.90/19.30	84.50/9.87	86.39/17.23	92.14/8.12
Mean speed (m/s)	16.57/2.23	19.84/1.53	15.39/184	19.35/1.81	17.69/2.58	13.40/6.0	16.05/1.72	16.45/2.09	15.52/1.87
Celeration (m/s/s)	0.157/0.033	0.135/0.020	0.202/0.026	0.153/0.020	0.181/0.050	0.200/0.010	0.170/0.03	0.162/0.029	0.202/0.019
* one outlier deleted, ** two outliers deleted									

Table 4.1. Descriptive statistics for the sample of drivers. Shown are the number of trips, and mean values with standard deviations for the variables collected, per driver (extracted 2019-06-10). Conditions from manual reports by drivers, scores calculated by the app and presented to the drivers, mean speed and celeration calculated from raw data from the app, based upon GPS positions. Conditions were reverse scored on a 5-1 scale.

4.3.3 Driver behaviour; correlations between variables

In

Table 4.2 the correlations between different features of trips, calculated over all the drivers' trips can be seen. It can be noted that the load conditions have an association with the acceleration score, probably because the load will impose a physical limit to the possible acceleration. Furthermore, both load and traffic conditions are associated with mean speed and celeration. Road conditions, on the other hand, seem to influence only the braking score, with ice and snow actually increasing harsh braking. This is in contrast to (unpublished) results for bus drivers, where icy conditions reduced the level of celeration despite a higher amount of passengers. However, it is possible that the present results are not reliable, as it can be seen from the descriptive statistics that the mean values for the conditions variables are very high and variation small.

Comparisons of the smooth score and celeration yield some interesting observations. Apparently, the former is to a large degree determined by the same underlying factor as the harsh braking score, with some similarity to celeration. However, although both these variables could be expected to be very similar, it can be seen that they tend to have association patterns with other variables which are very different. The smooth score was developed by the app team, while the celeration variable is a scientific concept (af Wåhlberg, 2008). The reasons for differences on these two variables will be investigated when more data is available.

Variable	Traffic conditions	Load conditions	Road conditions	Smooth score	Harsh acceleration score	Harsh braking score	Mean speed
Load conditions	.131	-					
Road conditions	.177*	-.092	-				
Smooth score	-.083	-.058	.069	-			
Harsh acceleration score	.165	-.216*	.249**	.125	-		
Harsh braking score	.105	-.034	.004	.447***	.048	-	
Mean speed	.374***	.245**	-.018	.090	.068	.089	-
Celeration	-.332***	-.278**	.056	-.281**	-.173*	-.357***	-.671***
* p<.05, ** p<.01, *** p<.001							

Table 4.2. The correlations between variables. Calculated over all trips for all drivers. N=131.

4.3.4 Development over time

As the measurements started in March 2019, it could be expected that the values for all the driver behaviour variables would improve, due to the seasonal change in weather and light (af Wåhlberg, 2007b). This was found to be true for seven out of the nine drivers, with the remaining two having few trips.

4.3.5 Stability of behaviour over time

A central concept in driver behaviour measurements is the continuity of differences in behaviour of drivers. It has previously been found that drivers behaviour, conceptualised as celeration, is very similar in different time periods (af Wåhlberg, 2007b). It is not influenced by crash involvement, but a general decline in values (probably due to experience) can be seen over time (af Wåhlberg, 2012).

4.4 Discussion

4.4.1 Conclusions

The present analysis of the first data from MeBeSafe WP4 seems to indicate that the system is in general working as planned concerning the data gathered, with the exception of some errors (discussed in the next section). Although the available data in terms of number of drivers and trips was too small to run certain tests with acceptable statistical power, the analysis so far does not give any counter-indications to a conclusion of acceptable quality of the data.

4.4.2 Limitations and errors of the system

The self-reports about driving conditions yielded some significant associations with other variables, but these were probably under-estimations of the actual ones, because in Norway the drivers will encounter very differing conditions during a single trip. These differences cannot be captured by a single self-report scale. When the app is used in other countries, with more homogenous driving environments, the self-reported data will probably become more representative of the actual driving undertaken.

The calculation of duration of the trip was found to be unreliable, as instances of zero duration were coupled with distances travelled of several kilometers and events had been recorded. Similarly, durations were in some instances very large. Here, two possible cases can be identified. First, a driver might forget to end the trip and let the app run. If the phone runs out of power, the trip is continued when the phone is started again. This can yield trips of hundreds of hours. The second case must be due to an error in the app or time signals, as some values of thousands of hours were found in the data.

5 Results of simulation of the coaching scheme for HGV drivers

5.1 Scope

The scope of the simulations is to analyse, which effect in terms of traffic safety can be expected from the application of coaching in traffic. For this purpose a simulation-based assessment approach is taken.

The 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 the technology under assessment are conducted in relevant traffic scenarios to reduce actual real-world road testing. The simulation complement real-world (road) testing by investigating road and traffic conditions, which may be difficult or dangerous to investigate in real-world road testing (e.g. crash and near-crash 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 measures such as nudging and coaching, as well as which updates are required to analyse these measures.

In WP4 the focus is on adapting the 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. Consequently, it is unlikely that any of the vehicles will be encountered in a crash during the field trial. The simulation should provide results regarding the number of crashes that can be prevented in case of large scale implementation of the coaching app. In simulation, results of the field trial can be scale to a larger (simulated) population, comparing simulated crashes for one population (considering no coached driver) to simulated crashes for second population (considering a coached driver). Considering the crash rate of trucks in Germany is 1.8451 E-07 crashes/km (DESTATIS18, KBA19), meaning that in 1 million driven km statistically 5.4 crashes are expected to occur. Depending of the effect of coaching and the required level significance for the analyse obviously a multiple million km of driving need to be simulated to statistically prove a positive safety effect by the reduction of the number of crashes as a result of coaching. This implies that required driven distance for this analysis is several million kilometres, which is out of scope for the testing in this project. The simulation gives also the opportunity to consider a more risky driver bevaiour of the simulated populations, which results in a higher rate of critical situation and crashes compared to the real world. By this the required kilometre to make first statements on the safety effect of the coaching measures can be reduced.

The coaching with respect to ACC is not the focus for the simulation-based assessment.

5.2 Methodology

In general, the question regarding the impact in terms of traffic safety of a certain technology is relevant for different stakeholders (politics, car/truck manufacturers and suppliers, insurers as well as researchers). The classical approach is to determine the impact by means of analysing crash statistics, comparing crash prevalence with and without certain technology (e.g. Farmer, 2004; Unselt et al., 2004). However, this requires that the technology has already reached a certain market penetration rate 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 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 crash 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 costs, 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 et al., 2017):

- Re-simulation of real-world traffic situations (crashes or safety critical driving situations);
- Modified simulation of real-world traffic situations (crashes 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.

The evaluation methodology in the simulation-based effectiveness analysis for the coaching measures in WP4 is given in Figure 5.1.

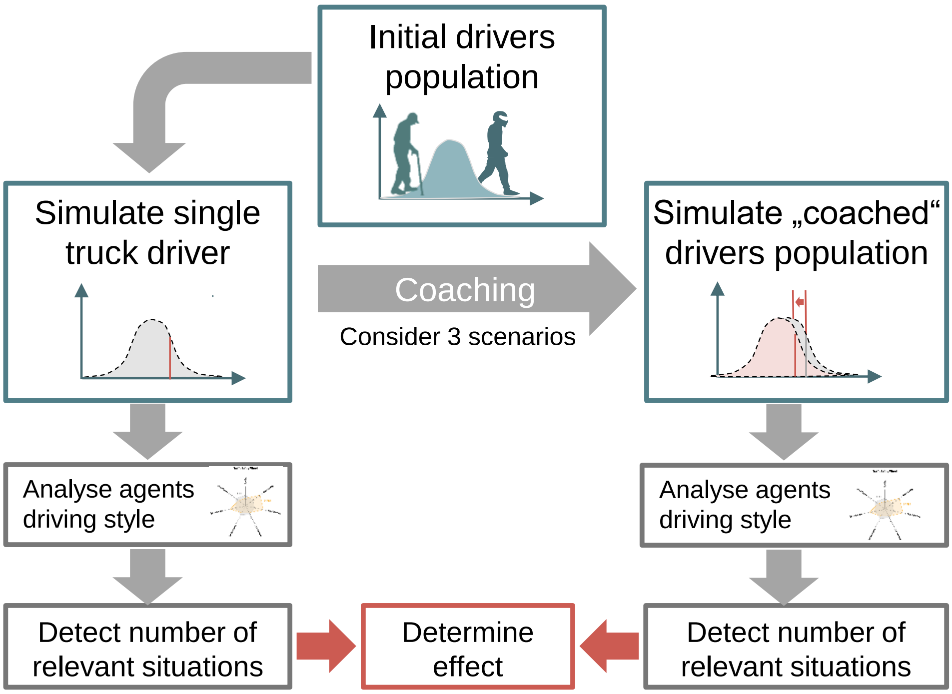

Figure 5.1. Evaluation methodology in the simulation-based effectiveness analysis for the coaching measures.

To determine the effect of coaching, the general characteristics of the agents are of particular interest. These characteristics cover in an explicit or implicit manner aspects, which describe how the traffic agent behaves in normal traffic. Examples for such agent characteristics are the number of km/h by which the driver is willing to exceed a speed limit, the preferred time headway or the maximum applied lateral / longitudinal acceleration. The characteristics are changed for a particular agent during the simulations in order to simulate the effect of coaching measures.

The original plan to use the first study results from the pilot test in order to feed the simulation, could not be used, since the required data have not been available at the point of time when the simulation needed to be conducted. Therefore, the coaching assumptions have been made based on expert judgement, how the single agent characteristic would change under the influence of coaching. In order to cover a broader range for each analysed agent characteristic three treatment scenarios are simulated. In the treatment scenarios the analysed agent characteristic is changed by 2.5%, 5% and 10% to the safer side. This means that in the first scenario the speed above the speed limit is reduced by 2.5%, whereas the time headway is increased by 2.5%. Next to the treatment scenario, of course the baseline scenario needs also be simulated, in which the driver characteristics are not affected. Thus, overall four simulations need to be conducted per simulation run.

For the assessment of the traffic safety potential of coaching, first a population of truck drivers is defined, which should be coached. The population consist of n = 10 drivers. For each driver the different agent characteristics are randomly chosen based on the original distribution of each parameter. Due to the stochastics, it is rather unlikely any driver has the same character parameters as another driver. One of the coached truck drivers is always presented in the simulation runs.

There are several simulation runs conducted for each driver in different traffic scenarios. For each simulation run, the relevant traffic safety performance indicators are calculated over the entire driver population of this run as well as for the coached truck driver.

The traffic safety performance indicators are:

- Frequency of crashes,
- Frequency of situations with deceleration above 3 m/s^2,
- Time during the simulation with a THW less than 0.9 s,
- Time during the simulation with a TTC less than 3 s,
- Time during the simulation with a deceleration of more than 3 m/s^2.

By comparing the traffic safety performance indicators over the conducted simulation runs for the four scenarios the effect of coaching is determined.

The simulation itself is set up in the following manner. First, different relevant traffic scenarios are defined. The traffic scenario describes the road (number of lanes, speed limit and length of the simulated road section) and the surrounding traffic (traffic density, fraction of trucks) of each simulation run. The traffic scenarios cover longitudinal traffic on motorway, rural and urban roads. The analysis does not cover crossing traffic by vehicles or vulnerable road users. This limitation must be taken into account for the analysis. The simulated traffic scenarios are reported in more detail in the following sub-chapter. For each traffic scenario per "coached" truck driver n = 250 simulation runs are conducted.

In the second step the simulation run is set up considering a stochastic variation of the traffic. The number of the traffic agents in the simulation run is defined by the traffic density, the length of the road section and the expected driven speed. Once the number of traffic agents is known, the number of the trucks and passenger cars can also be calculated. For each agent a starting position on the track respectively time, starting lane and starting velocity is randomly chosen. In addition, the parameters of the driver characteristic for each agent are also randomly chosen. Here, different distributions are used for passenger car and truck agents. For the truck agents the distribution is the same as for the coached truck drivers. However, in contrast to the coached truck drivers the parameters do not stay constant during the simulation.

Thus, the non-coached traffic differs in each run. Once the simulation run is set up, the simulation is started. Each simulation run is simulated for 100 seconds. It is important to note that the starting conditions of the four coaching scenarios (baseline, three treatment scenarios) are identical in each simulation run.

5.3 Tools & Models

The agent-based simulation for analysing the safety potential of coaching have been implemented and conducted in MATLAB 2018b. All necessary models (vehicle, road / environment and driver behaviour model) have been also implemented in MATLAB 2018b.

For the driver behaviour model an adapted version of the IDM driver behaviour model of Treiber, Hennecke and Helbing (2000) is used. The IDM driver behaviour model is a time-continuous car-following model. The adaptations are mainly related to the perception of the surrounding traffic. In contrast to the original model, the agents perceive only the information about the surrounding traffic in the area, at which they are looking at. Furthermore, for the parameters within the model have also been adapted in order to achieve a more realistic crash frequency. Three agent characteristics of the driver behaviour model have been identified in order to simulate the effect of coaching. These characteristics are:

- **Desired time headway:** the time gap the driver tries to keep to the predecessor in car following situations;
- **Desired velocity:** the velocity the agent would drive in case of free traffic;
- **Compliance with the given speed limit:** a factor, which describes how well the agent follows the speed limit in case of free traffic.

The three characteristics are changed according to the three treatment scenarios that have been described in the previous chapter.

Next to the simulation itself also the simulated traffic scenarios need to be defined. As stated in the previous chapter the focus for the simulation study is on the longitudinal traffic behaviour. Therefore, the road for all traffic scenarios is a 10 km long road stretch, for which the number of lanes, the speed limit as well as the traffic density is varied. These parameters have been defined that they represent motorways, urban roads and rural roads. The traffic scenarios as described in

Table *5.1* have been simulated in order to assess the safety potential of coaching.

Scenario Parameter					Scenario ID		
Number of Lanes [-]	Speed Limit [km/h]	Traffic density [veh/h]	Pro-portion Truck [-]	Targeted Road Type [-]	Desired time headway	Desired velocity	Com-pliance w speed limit
2	50	1200	17%	Urban	1	9	15
2	130	1600	13%	Motorway	(3)	(10)	N/A
1	50	600	17%	Urban	5	11	16
3	130	2400	13%	Motorway	(6)	(12)	N/A
1	80	800	13%	Rural	7	13	17
2	100	1600	13%	Motorway	(8)	(14)	N/A

Table 5.1 Overview on the simulated traffic scenarios.

Overall, 17 different traffic scenarios have been defined for assessing the effect in case the behaviour characteristics of a driver changes due to the coaching. For the characteristics "compliance w. speed limit" the motorway scenarios are not simulated, since the speed limit of the road is above the speed limitation of trucks. Therefore, for these no effect is presumed.

5.4 Results

The different traffic scenarios as defined in *Chapter 5.3* have been simulated for 10 coached drivers. For each driver, the impact of coaching one of the behaviour characteristics in terms of frequency of crashes, the mean velocity, the time driving with a THW of less than 0.9 s and the time driving with a TTC of less than 3 s has been analysed. The data of the stochastic generated traffic has also been evaluated. However, for the analysis it is focused on the coached drivers.

For the motorway scenarios (traffic scenarios 3, 6, 8, 10, 12 and 14) the first analysis showed no difference in terms of crashes related to coaching measures. Therefore, it has been decided to focus on the urban and rural scenarios (traffic scenarios 1, 5, 7, 9, 11, 13, 15, 16 and 17). The results are presented in the following per coached driving characteristic. For the results, it must be taken into account that the analysis focuses on longitudinal traffic conflicts. Conflicts with crossing are not taken into account.

Overall, approx. 2'711'000 vehicle km have been simulated in the urban and rural traffic scenarios (coached vehicle approx. 17'000 veh. km; surrounding traffic approx. 2'694'000 veh. km).

5.4.1 Desired time gap

The first coached behaviour characteristic is the desired time headway, which describes the time headway that the driver tries to keep during normal car following. The three analysed coaching scenarios are an increase of the desired time headway by 2.5%, 5% and 10%.

For the mean velocity in the traffic scenarios no difference between the baseline and all three treatment scenarios can be determined (see Figure 5.2). This result has been expected, since an increase of the time headway during the simulation has only minor influences on the driven velocity, since in car following situations, in which characteristic becomes relevant, the velocity is mainly defined by the speed of the predecessor.

For the time driving with a THW below 0.9 s a coaching, affected by desired time headway behaviour, has a large effect (see Figure 5.3). This is in fact not surprising. It is rather the amount of the effect which the simulation shows. This implies that even a minor improvement could lead to a significant reduction of the time driving with very close distances.

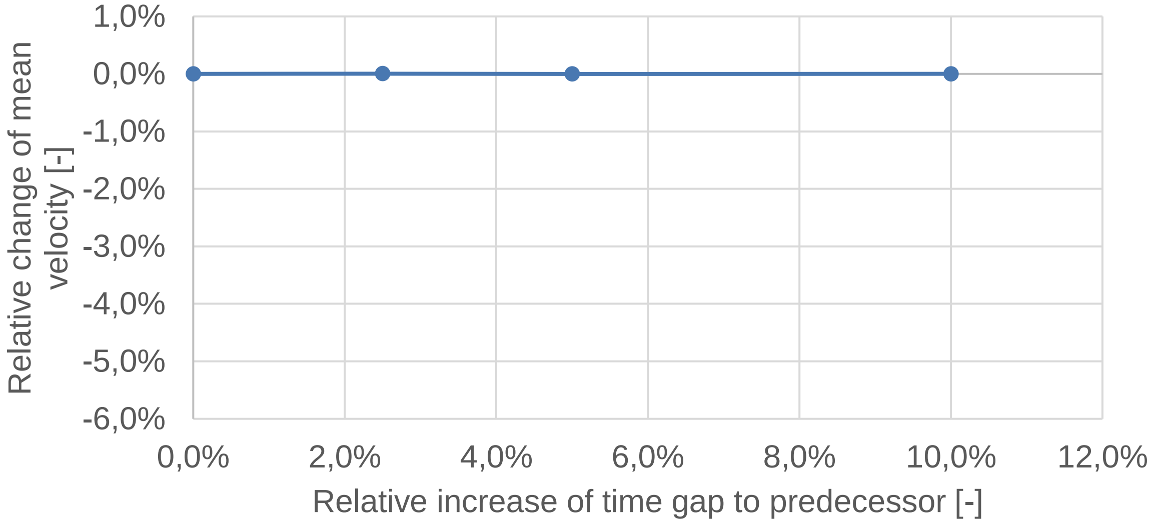

Figure 5.2. Relative change of the mean velocity due to a changed desired time headway behaviour.

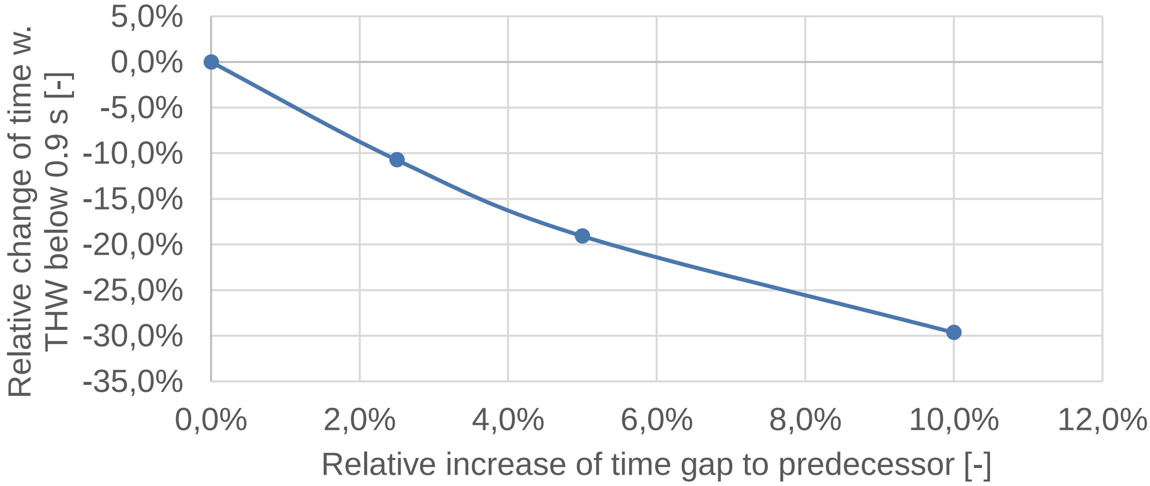

Figure 5.3. Relative change of time driving with a THW below 0.9 s due to a changed desired time headway behaviour.

A similar tendency is observed for the time driving with a TTC below 3 s, see Figure 5.4. Although the relative change is not equally high as for the driving time with a low

THW. The time to collision is a better indicator for detecting safety critical driving situations, since it describes the time remaining until a collision, if the movement of the traffic participants stays constant.

Therefore, driving situations with low TTC occur less frequently then driving situations, in which a vehicle follows the predecessor closely. This has also an influence of the results. In the graph in Figure 5.4 a slight disturbance due to the applied stochastic can be observed. A similar effect is determined for the analysis of frequency of situation with high decelerations (see Figure 5.5)

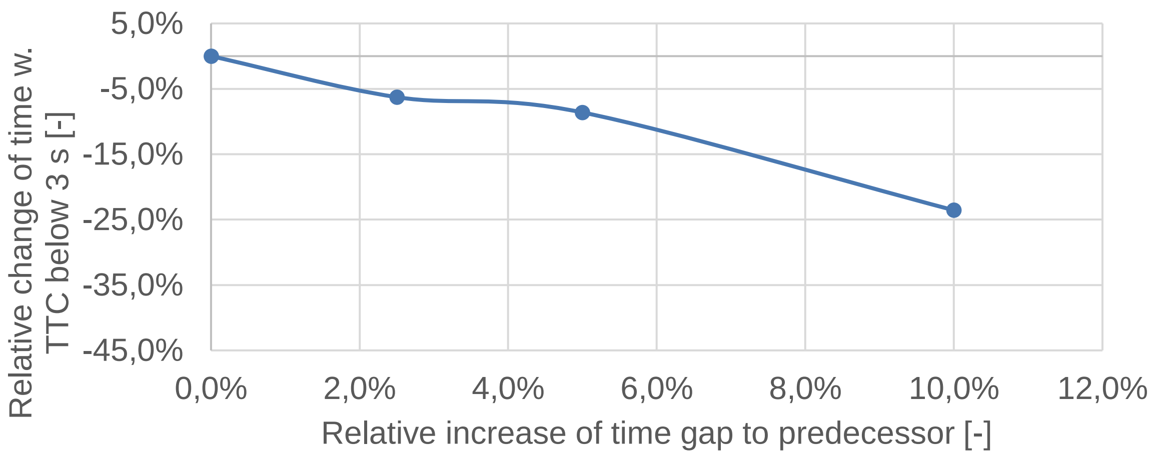

Figure 5.4. Relative change of driving time w. TTC below 3 s. due to a changed desired time headway behaviour.

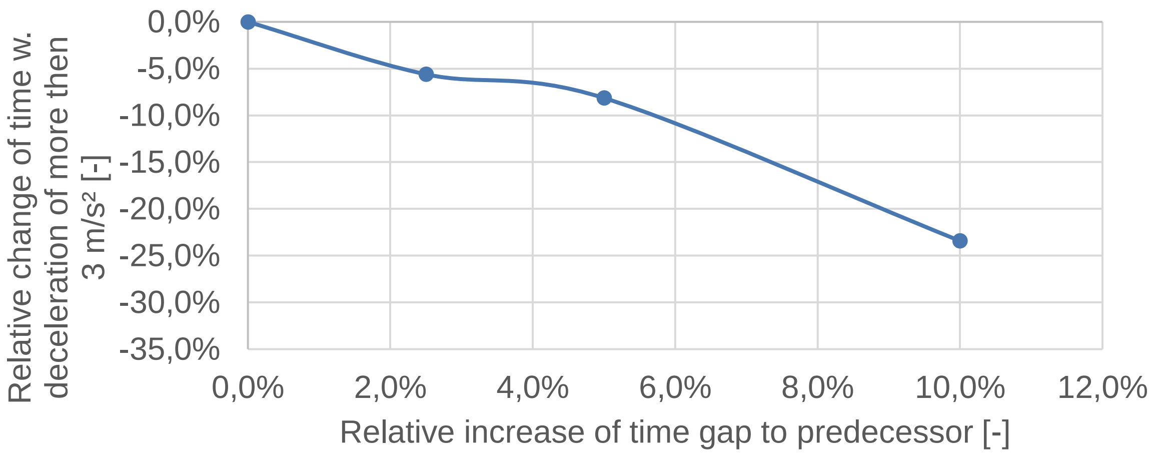

Figure 5.5. Relative change of frequency of situations w. deceleration aboive 3 m/s² due to a changed desired time headway behaviour.

This effect becomes even more relevant, once the number of relevant situations decreases further. This happens if the crashes that are detected in the simulation are analysed (see Figure 5.6). When computing the linear regression it becomes obvious that the results show a high variation within themselves due to applied stochastics. The natural approach to minimize this variation would be to increase the number of simulation runs. However, this has not been possible due to limited resources.

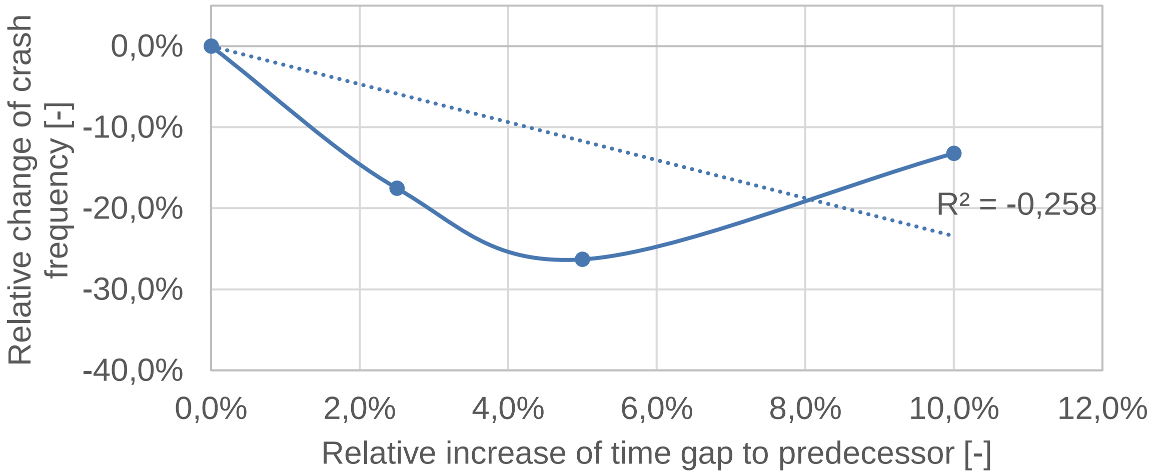

Figure 5.6. Relative change of crash frequency due to a changed desired time headway behaviour.

Overall, the simulation results show that if the time gap to the predecessor is increased due to coaching the number crash frequency decreases.

5.4.2Desired velocity

In case the desired velocity is affected by coaching, the mean velocity is decreased. Since the desired velocity is only of relevance in free driving situations, the mean velocity does not decrease in the same amount as the characteristic is changed, see Figure 5.7.

Regarding the distance behaviour, a similar effect due to a reduced desired velocity is determined as for an increased desired head way, see Figure 5.8. The time amount, in which the vehicle is in a close following situation, decreases significantly. The simulations show a reduction up to 29% for a by 10% reduced desired velocity compared to baseline.

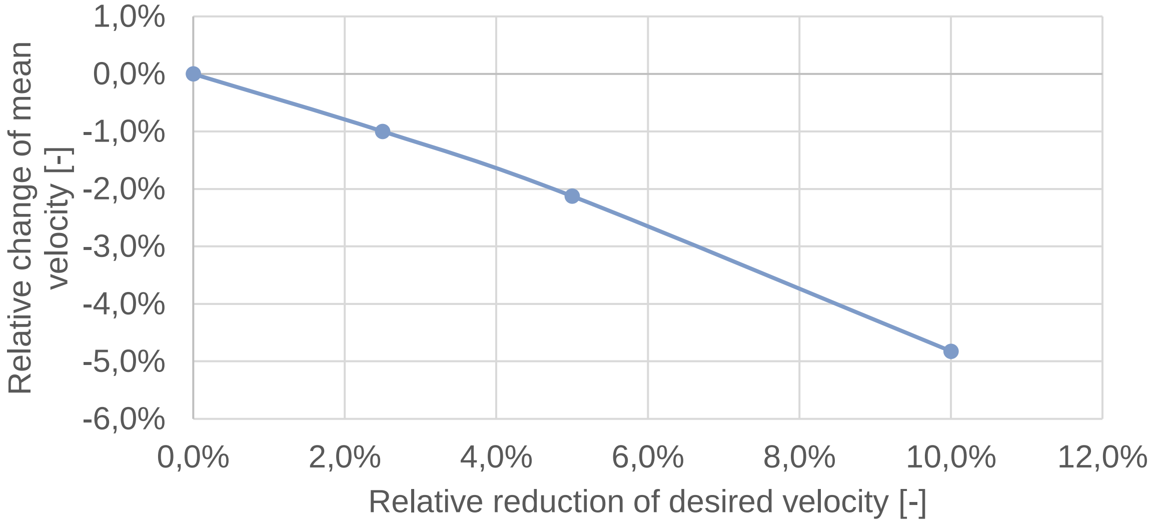

Figure 5.7. Relative change of the mean velocity due to a changed desired velocity behaviour.

Also the time driving in critical situations (TTC < 3 s) is reduced in case the desired velocity is reduced due to coaching, see Figure 5.9. The determined effect is even larger for the desired time gap. One possible explanation is that a reduction in the velocity leads to smaller difference velocity, which makes it easier to react in time and leads in general to larger TTC values.

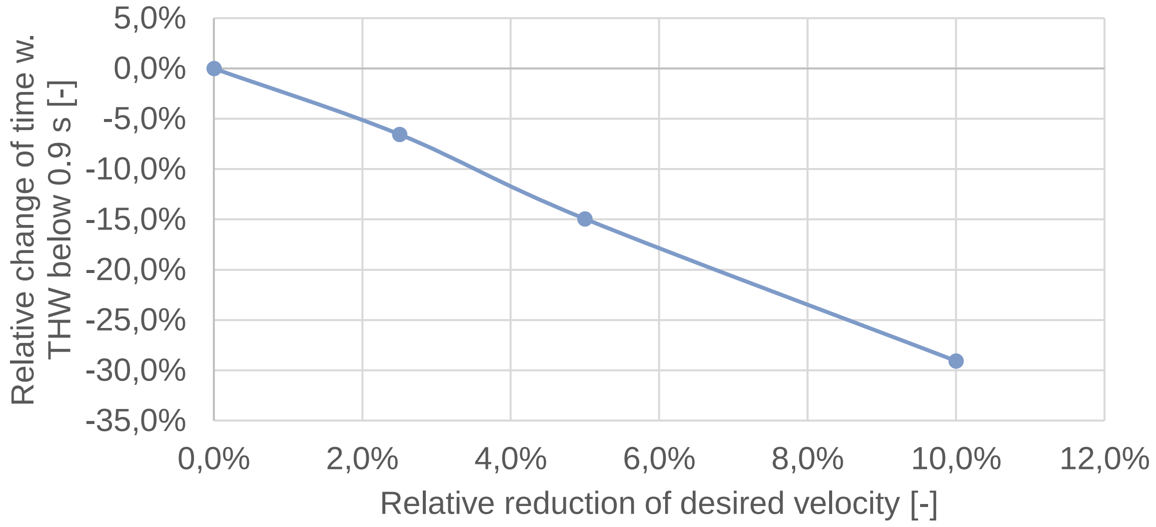

Figure 5.8. Relative change of time w. THW below 0.9 s due to a changed desired velocity behaviour.

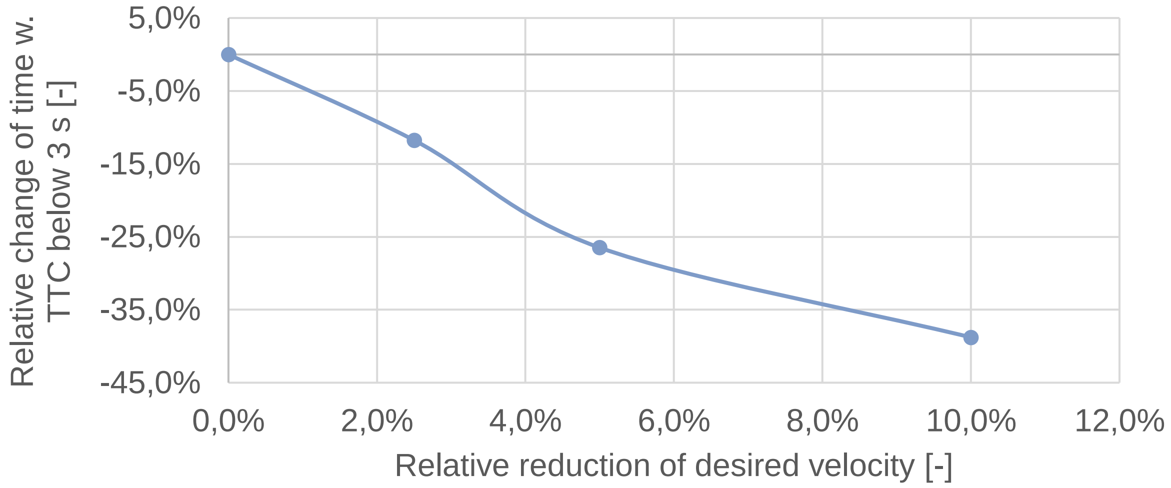

Figure 5.9. Relative change of time w. TTC below 3 s due to a changed desired velocity behaviour.

As for the desired time gap a similar result has been obtained for the frequency of brakings with a deceleration of more than 3 m/s^2 (see Figure 5.10).

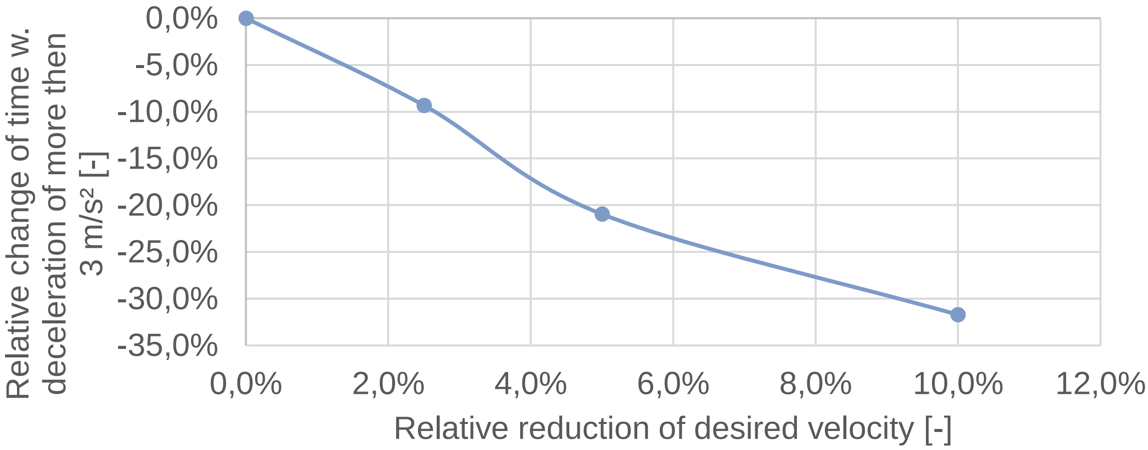

Figure 5.10. Relative change of frequency of situations w. deceleration above 3 m/s² due to a changed desired velocity behaviour.

The positive indication of the critical situations is confirmed for the crash frequency: the crash frequency is reduced in case the desired velocity is reduced, see Figure 5.11. The results fit quite well with the linear regression. However, this might be related to coincidence, since the analysis of time gap showed that the number of simulations is rather low for generalizing the result. Compared with the previous results for an increased time gap the effect is slightly less.

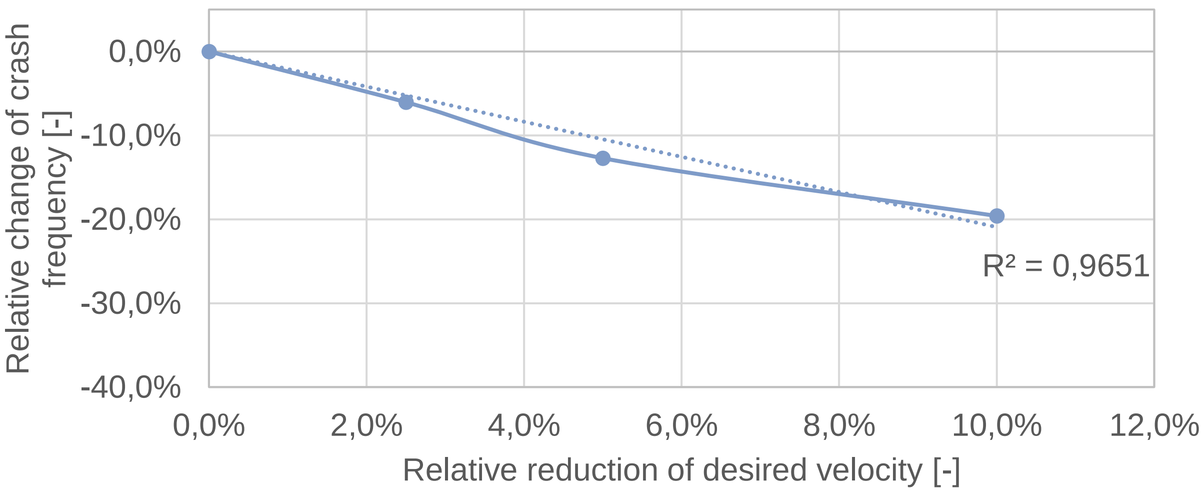

Figure 5.11. Relative change of crash frequency due to a changed desired velocity behaviour.

5.4.3 Compliance with the given speed limit

The last characteristic that is analysed is the compliance with the given speed limit, respectively the number of drivers that are willing to exceed the given speed limit. Regarding the mean speed limit, an increase in the compliance with the speed limit leads only to minor differences (see Figure 5.12). The maximum relative change is 0.2% compared to the baseline.

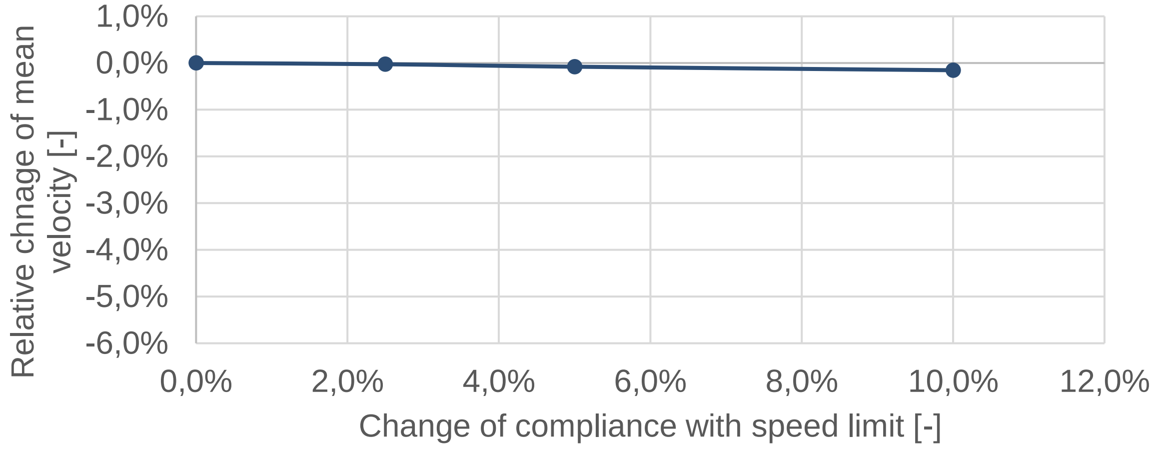

Figure 5.12. Relative change of the mean velocity due to a changed compliance with the given speed limit behaviour.

With respect to the car following behaviour, a decrease of driving time at close distances is determined for an increase of the compliance with the speed limit (see Figure 5.13). There is a more or less linear relation between the compliance with the speed limit and the time driving at close distance. The relative change is less compared to the previous two characteristics. For the comparison with the desired velocity this seems reasonable. Both of these characteristics affect the driving speed. However, the compliance with the speed limit becomes only relevant if the vehicle is driving at a speed close to the speed limit, whereas the desired velocity is relevant throughout the entire drive.

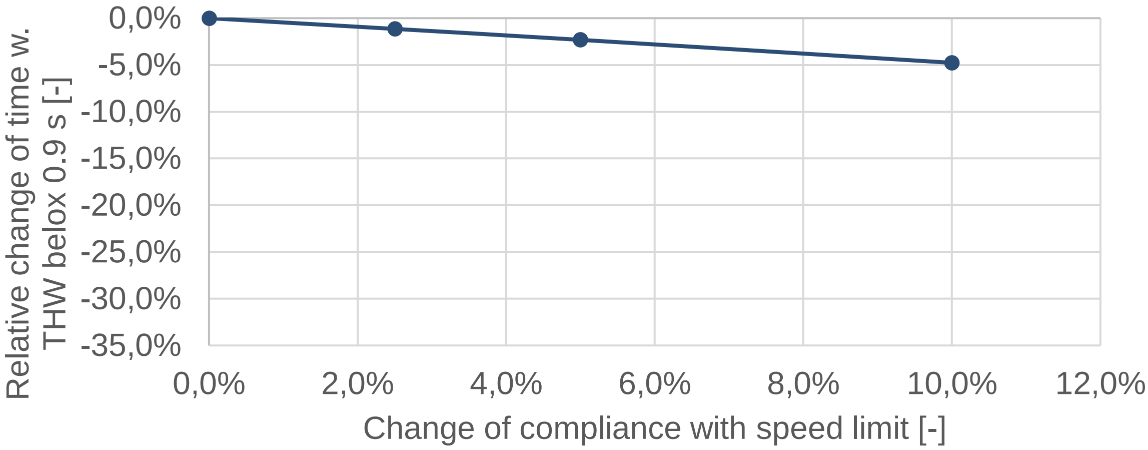

Figure 5.13. Relative change of time driving with a THW below 0.9 s due to a changed compliance with the given speed limit behaviour.

The time driving in situations with lower TTC (< 3 sec.) is also decreased in case the compliance with the given speed limit is increased due to coaching (Figure 5.14).

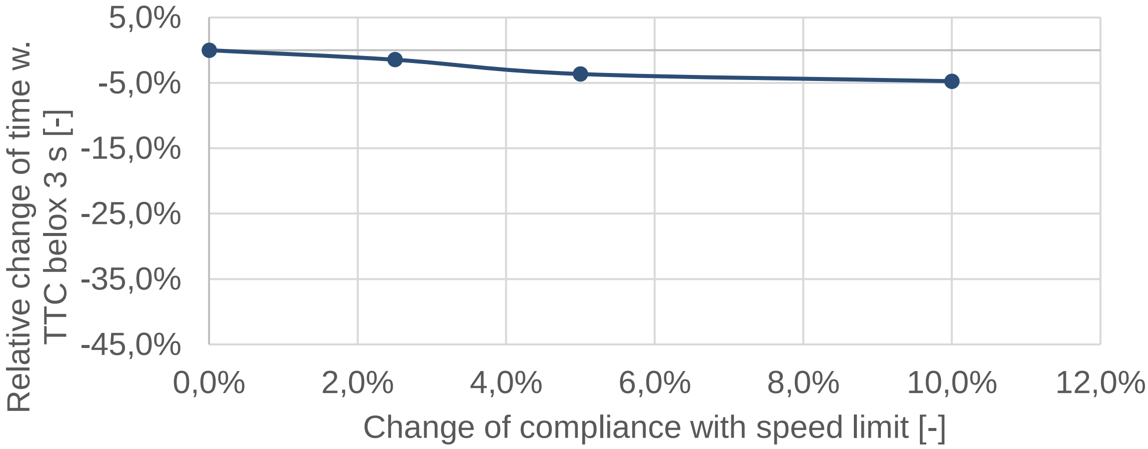

Figure 5.14. Relative change of driving time w. TTC below 3 s due to a changed compliance with the given speed limit behaviour.

At maximum, the time driving with a TTC below 3 s is reduced by 5% compared to the baseline. Again, the effect for this characteristic is smaller than the effects that are determined for the previous characteristics.

The frequency of braking manoeuvres with a deceleration of more than 3 m/s² is also reduced in case the compliance with the speed limit is increased (see Figure 5.15).

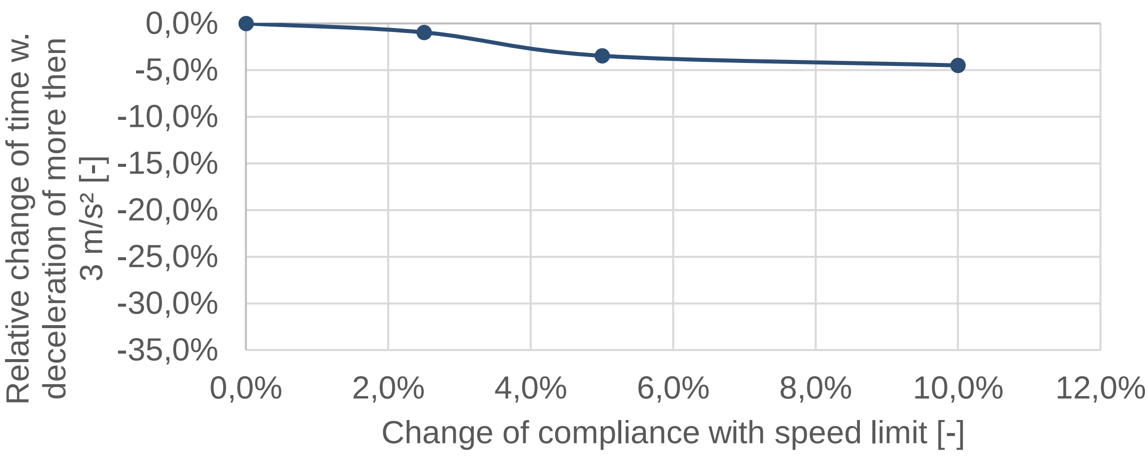

Figure 5.15. Relative change of frequency of situations w. deceleration aboive 3 m/s² due to a changed compliance with the given speed limit behaviour.

For the frequency of crashes, the simulation results are not as clear as for the previous characteristics (see Figure 5.16).

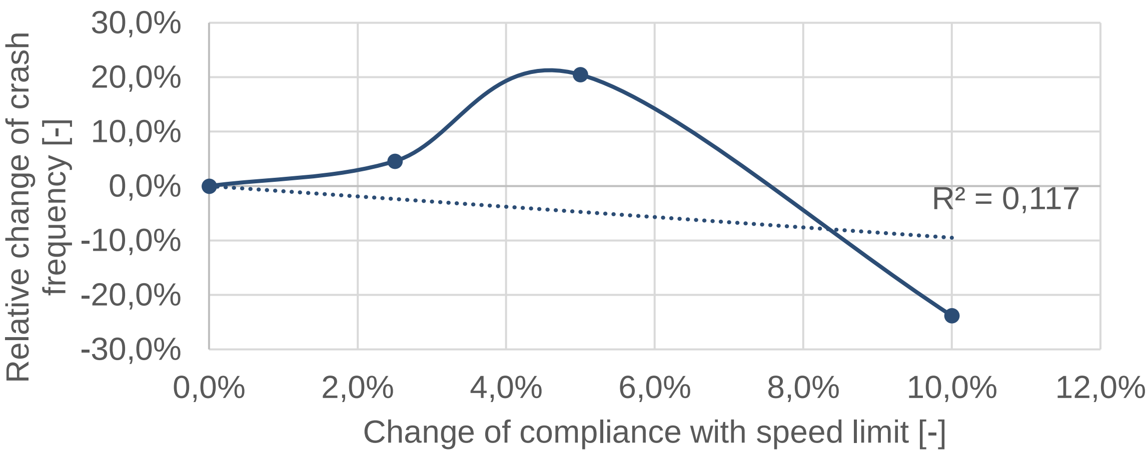

Figure 5.16. Relative change of crash frequency due to a changed compliance with the given speed limit behaviour.

The linear regression shows overall a decrease of the crash frequency for an increasing compliance with the speed limit. However, for two of the three treatment simulations the crash frequency is above the baseline. Thus, there is a very high variation in the results due to the stochastic simulation. Therefore, no clear statement is possible with respect to the crash reduction potential by means of these simulations in case coaching leads to increased compliance with speed the limit.

A more detailed analysis with a higher number of simulation runs is required in order to analyse the effects in more detail. Due to limited resources an increase of the simulation runs has not been possible in this work package and it must further be considered that the simulation of the required amount of driven kilometre – this will be above millions of kilometres – is out of the project's scope. The purpose is rather to give first indictations on the expected effects. This issue is discussed in more detail in chapter 5.5.

5.4.4Analysis of all simulated traffic scenarios

In the previous section the analysis has been conducted for each characteristic independently. Of course the frequency of crashes can also be analysed over all conducted simulation runs. For this purpose the mean crash frequency for the baseline and the three treatment conditions is calculated. Afterwards the relative change for each condition versus the baseline condition is calculated. For the third treatment condition (10% change as a result of coaching) a reduction of the crash risk of 15% for the coached driver is determined (see Figure 5.17). However, the issue of a relative small number of simulations is also present for this analysis, since the second treatment condition leads to lower reduction of the crash risk than the first one. This seems rather unlikely.

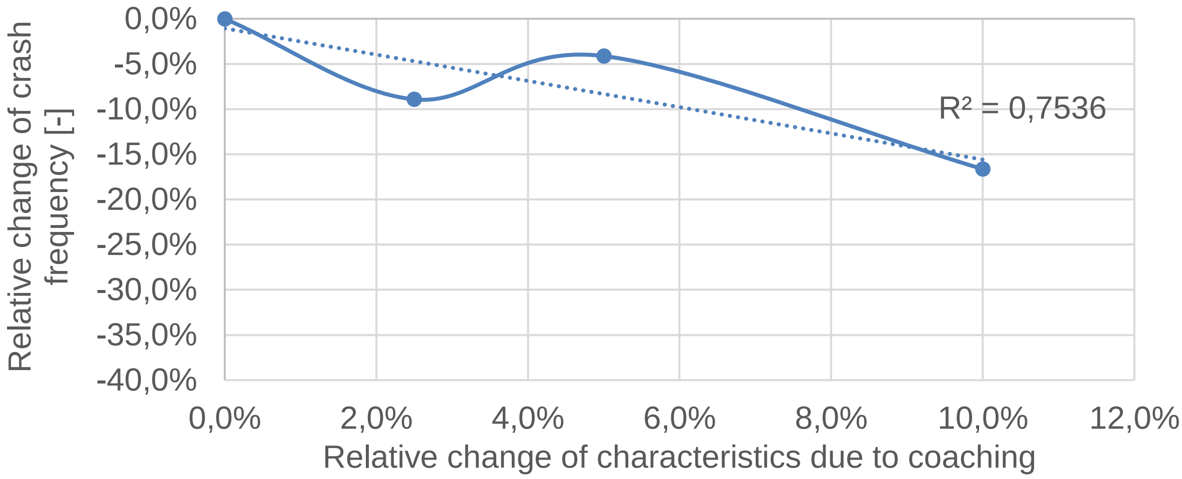

Figure 5.17. Overall analysis of the relative change of crash frequency of the coached agents due to relative change of characteristics.

As stated in the beginning of this chapter, the simulation has been focused up to now on the coached drivers. But, the crash frequency can also be analysed over all agents (see Figure 5.18).

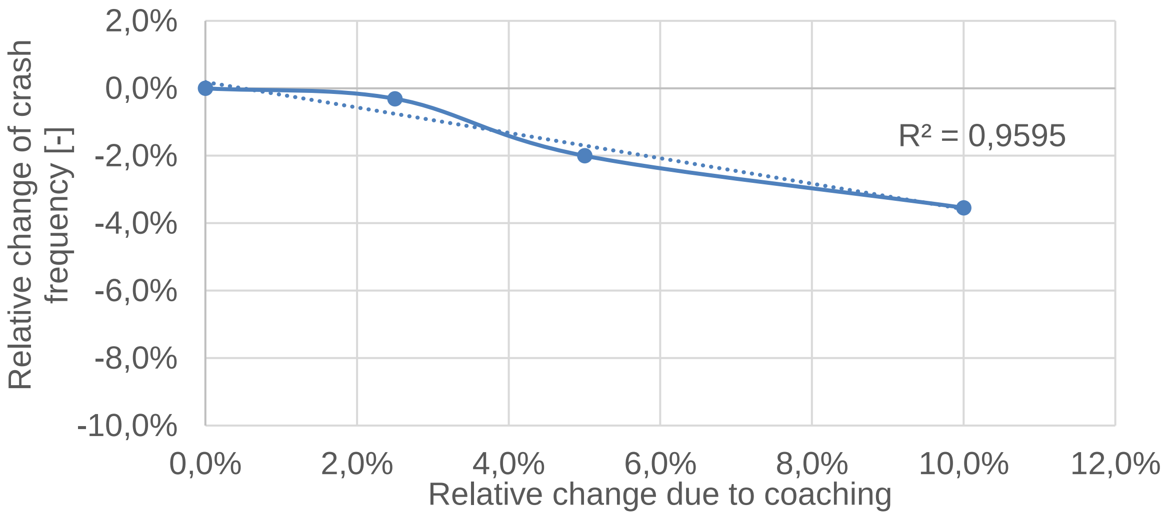

Figure 5.18. Overall analysis of the relative change of crash frequency of the all agents due to relative change of characteristics.

Of course, the determined effect is lower compared to the analysis only conducted for the coached drivers, since only one driver is "coached" in each simulation run. This does not necessarily mean that the surrounding traffic does not also benefit from a changed driving style of the coached driver.

Overall, the analysis shows a reduction of the crash risk of close to 3.5% over all traffic agents in case the coaching leads to a change of driver behaviour by 10% to a safer driving style.

5.5 Discussion

The main objective of the applied simulation was to investigate the safety potential of the measure "coaching". Moreover, it was important to investigate to which extent the simulation-based prospective safety effectiveness assessment can be applied for coaching measures.

The results show a safety benefit in case coaching affects the desired time gap and the desired velocity. For the other parameters (mean velocity, time driving with short distances, time driving in lower TTC situations, frequency of hard braking manoeuvres) the simulations also show benefits. For the results it must be taken into account that driver parameters in the simulation have been defined based on expert opinion in the best possible way, since real data were not available. When comparing the crash rate of the relevant drivers in the baseline simulation (9.37359 E-03 crashes/km) with the crash data in Germany (1.8451 E-07 crashes/km, DESTATIS18, KBA19), it becomes obvious that the crash frequency in the simulation is much higher. On the one hand, this is necessary to get to a certain number of crashes with reasonable simulation effort. On the other hand, this means that the "coached" drivers in the simulation have been driving in a more risky manner. Therefore, measured improvement due to coaching might be higher than is possible in reality, where the drivers drive safer.

In other scenarios no clear effect can be identified. This does not necessarily mean that there is no effect. It must rather be considered that the simulation applies a number of stochastic variables. Therefore, a certain number of simulation runs is required to reach stable results. Increasing the number of kilometres to multi-million kilometres is out-of-scope of the project. The simulation time of 250 runs with a simulation time of 100 s on the used computer[5] equals for one driver to approx. 10 h. For the simulation in the project of course more than one computer has been applied. However, even if the computation time of the simulation is improved, the simulation time stays a limiting factor.

Next to the simulation time other limiting factors for the application of the simulation to investigate the effects of the measure coaching have been identified:

- First, the method covers mainly the effects of the changed driver behaviour due to coaching. Since the pilot test did not deliver estimates of behaviour change, for the simulation a scenario based approach with three treatment conditions has been chosen, in which the expected changes due to coaching have been pre-defined. A more sophisticated approach would be to develop a model that links the coaching measure to the change of the driving behaviour. The development of such a model has not been possible in this project, since no relevant data has been available at the time of setting up the simulation. Nevertheless, the assessment approach would benefit from such a model.
- The second issue is a limit in the number of simulation runs. The number of simulation runs depends on the variation within the simulation runs, but also on the effect of the applied solution or measure. This means that for measures, which provide minor changes compared to the baseline, the number of simulation runs / trials must be increased to reach significant effects. Whether coaching achieves a strong or minor effects depends heavily on the starting point. If the coached driver is already a good driver in terms of traffic

[5] HP Z4 G4 Workstation (Intel Xenon W-2125 CPU 4.00 GHz, 32 GB RAM)

safety, the room for improvement is limited – meaning that the possible effect is also expected to be limited. The issues are even more relevant in this project, which focus for coaching on truck drivers. Truck drivers are professional drivers and can be expected to be very experienced drivers. This assumption is confirmed by a comparison of the crash frequency in Germany based on crash data (DESTATIS, 2018) and the driven mileage (KBA, 2019). Considering the police reports, the crash frequency of trucks above 3'500 kg is 32.8% lower than the crash frequency of passenger car drivers. Thus, for truck drivers it appears to be even harder to show an improvement.

- The third limitation is that the simulation in the applied form is limited to longitudinal scenarios, although different environments have been simulated. For instance, crashes occuring with crossing traffic or in turning manoeuvres are not covered by the simulation of this project.

- A further limitation is that the simulation approach as it is applied in this project does not cover effects over time. If the effect of the coaching decreases after the coaching session over time and the driver falls back to old habits, this will lead to a reduction of the potential safety effect. This effect is not covered by the simulation approach. By means of an additional model, which addresses this particular effect, this limitation can be solved. However, creating such a model requires further research and data about coaching and its effects. This is outside the scope of this project.

It can be concluded that it is in general possible to determine the safety impact of coaching measures by means of the simulation approach. However, for the simulation approach several limitations exist (e.g. number of required simulation runs, finding the right parameters for the baseline). These limitations need to be addressed in order to get to a more solid analysis. Of course these limitations have an effect on the presented results and must always be considered when referring to the results.

6 Business case of the coaching scheme for HGV drivers

According to the World Health Organization there are approximately 1.25 million deaths each year from road traffic incidents (WHO, 2018). Road safety is a global concern and interventions are necessary. The DriveMate Application has the potential to reduce the loss of human lives. However, next to reducing the number of road safety incidents and deaths, the DriveMate application has several business prospects. The most likely business case is described below: The DriveMate app as replacement for hardwired In-Vehicle Monitoring Systems (IVMS).

6.1 In-Vehicle Monitoring System potential business case

In-Vehicle Monitoring Systems (IVMS) are often the standard within many logistics and transport companies. The IVMS , also known as telematics systems are devices that accumulate and integrate GPS-based information about vehicle location and data from the vehicle's own electronic management systems. IVMS vary in terms of their complexity and costs. Among IVMS there are simple and advanced GPS, video and hazard warnings with telematics and emergency response systems. The DriveMate Application has the potential to replace current systems for the driver coaching aspect and deliver the same level of performance but at significantly lower cost.

Current IVMS costs include a variety of expenses. Firstly, the IVMS hardware and software should be installed in vehicles. Secondly, there are continuous costs of subscriptions for the IVMS services and data sharing. The costs of data sharing vary greatly depending on the geographical location as in some remote areas no 3G/4G network is available, so IVMS is required to use satellite connection to transmit the data, what makes the process significantly more expensive. Thirdly subscription costs are dependent on the selection of services required from the providing company. Finally, the interpretation of the obtained information and presenting the data is costly.

The European market (European Union and EFTA) size for IVMS (potential number of Vehicles that can use the DriveMate Application) in 2017 was 13.048.717 (ACEA, 2017). Based on the current applications in the market an early estimate of annual license fee would be a price range of 20 to 60 dollars per vehicle. With a target market share of 0.5% and a steady increase to 2% over a period of three years, an annual NPV of 5.220.000 dollars is expected (based on a license fee of 40 dollars). The proposed licensed fee is significantly below market standards. The predicted NPV is based on the scenario that the profit margin does not increase with an increased sale.

From a client perspective the DriveMate IVMS solution will reduce IVMS costs and also improve road safety in a novel way and it allows increasing productivity and efficiency by reducing costs and making use of the latest technology. With the DriveMate App we can also nudge driving behavior to reduce the fuel consumption through smoother driving (with less harsh braking and harsh acceleration), and potentially allow bringing down the maintenance costs. Deployment of DriveMate as an IVMS includes peer-to-peer coaching of drivers (and their supervisors) on how to interpret and act upon received data which is normally also a cost component. An easy and cost-effective deployment of the DriveMate Application in combination with above unique selling points will result in a swift uptake by the market.

6.2 Human life case

Next to the above business case the DriveMate application has the potential to save lives. After developing the first version of the app we still have a strong confidence in earlier analysis that the application will save lives.

Despite Heavy goods vehicles (HGVs) making up only 2% of the EU vehicle fleet (Grant Agreement - MeBeSafe) they are disproportionately involved in fatal traffic crashes (largely due to their very high annual mileage compared to passenger cars). A UK study concluded that HGVs make up only 10% of motorway traffic but are involved in 52% of fatalities (Grant Agreement - MeBeSafe). A similar disproportionate ratio is valid for inter-urban traffic. This does not mean HGVs are causing these crashes—in many cases the behavior of other drivers is the primary causation factor but the high mass and large size of the HGV increases the severity of consequences. We do however expect that measures directed at strengthening the risk-avoiding behaviour of HGV drivers will have an (also disproportionate) positive effect on all traffic users.

If the DriveMate coaching measures for HGV drivers have the desired effect (reducing harsh driving events by 40%; Grant Agreement), this can be argued to reduce all types of collisions: Harsh braking is the default response for HGV drivers (who cannot easily swerve) to acute hazards spotted very late—the source of 58% of all crashes involving HGVs (Volvo Trucks, 2013). In a study by Volvo Trucks and LYTX effective coaching has shown to lead to 4% reduction in crash costs for a fleet of HGVs (Grant Agreement - MeBeSafe). If we assume that the effect of our coaching measure is along the same lines, we can argue that our coaching to reduce harsh driving events reduces HGV-related crashes by 40% x 58% x 4% = 1%. With a baseline estimate of 3,000 road fatalities in the EU involving HGVs by 2025 this would lead to 30 lives saved.

However, this estimate may be too modest. According to LYTX (Grant Agreement - MeBeSafe) 20% of HGV drivers account for 80% of all harsh braking events. If these receive coaching and reduce the occurrence of harsh braking events by 40%, the total number of harsh braking events would drop by 20% x 40% = 8%. As per above we can equate harsh braking with late detection of hazards, leading to 58% of all HGV-related crashes. We can then argue the effect of our coaching to save 8% x 58% x 3,000 = 140 lives. For each fatality there are on average (for all types of vehicles) 8 serious injuries—we have used this number in our calculation though it can be argued that crashes involving HGVs usually have more severe outcomes.

It is clear that the business opportunities for DriveMate in the industry are enormous. Next to those opportunities DriveMate can impact lives. In sum, if the DriveMate application is used by all HGV drivers, this has the potential to save 30-140 lives and avoid 240-1100 serious injuries annually in the EU. Nothing is more valuable than human life and health.

7 Pilot test with Volvo drivers

As has been further detailed in D4.3 Part 2 (af Wåhlberg et al., 2019), when it comes to ACC usage, the overall picture of ACC usage levels indicate that ACC users can be grouped into three types; the intensive users, the modest users and the non-users, where the last group does not use ACC at all. Furthermore, a clear difference in mind-set can be identified between users and non-users. Both the intensive and modest users were well aware of how ACC operates and comfortable with using it while driving. Drivers in the non-user group on the other hand appear to be afraid of activating ACC, because they do not trust it to be capable of actually regulating speed and the distance to lead vehicles.

From that analysis, it has been determined that ACC oriented coaching might not have its largest impact on drivers who are already using ACC, but rather on drivers who do not use ACC at all. In principle, since nudging toward increased ACC usage only can be applied on drivers who are already function users, non-users must first become users before nudging can be applied. It has been decided that coaching will be applied primarily towards non-users, with the goal of turning them into users, and hence become available subjects for nudging efforts.

7.1 Design

Given this direction for coaching described above, clear synergies can be created by a collaboration between Volvo's MeBeSafe pilot testing of what conditions and processes need to be applied to create sufficiently effective feedback to turn non-users of ACC into users with another (in-house) Volvo project called In-Car Test Drive. The goal of In-Car Test Drive is to address the fact that currently available methods for potential customers to learn about car features may not always be able to create awareness of the full offer. Some car features may not be understood, or pass unnoticed in e.g. a test drive context. Thus, new methods for discovering and learning about car features in a real driving context need to be explored.

As can be seen, the goal of In-Car test drive case is closely related to the MeBeSafe goal of turning non-users of ACC into users. For both, the goal is to create awareness of particular features, and to provide a context where these can be explored in a controlled manner. The difference lies in two scope dimensions. While in-Car Test Drive covers a broad range of vehicle functions suitable for a prospective customer, the MeBeSafe ACC coaching focuses on a single function. Also, the mind-set of the test persons is different. For In-Car test drive, any person walking into a dealership for a test drive is a target customer, whereas for MeBeSafe, only drivers who are afraid of delegating vehicle control to a vehicle function are in scope.

However, in both cases the MeBeSafe use case can be described as a subset of the In-Car Test Drive use case. Hence it has been decided to complete MeBeSafe testing by analysing the feedback and experiences from customers involved in the In-Car Test Drive pilots. In particular, their response and behaviour when being prompted to activate the ACC and Pilot Assist (PA) functions is of interest. Pilot Assist is an Advanced Driver Support function that combines ACC capacity (i.e. it will maintain speed and/or time headway to a lead vehicle) with lane keeping through steering support. It therefore will definitely not be activated by drivers who are afraid of activating ACC.

All in all, the In-Car Test Drive ecosystem contains an in-In-Car App and an Admin tool, both integrated with the back-end systems consisting of VIN-lookup, dialogue platform, speech recognition and the Volvo Connectivity cloud. The In-Car App is named the *Interactive Quick-guide*. It sits in the centre screen display and presents information mainly by voice, to avoid potential visual distraction elements. This allows for information and coaching in a real driving context, moving beyond theoretical presentations and instructions.

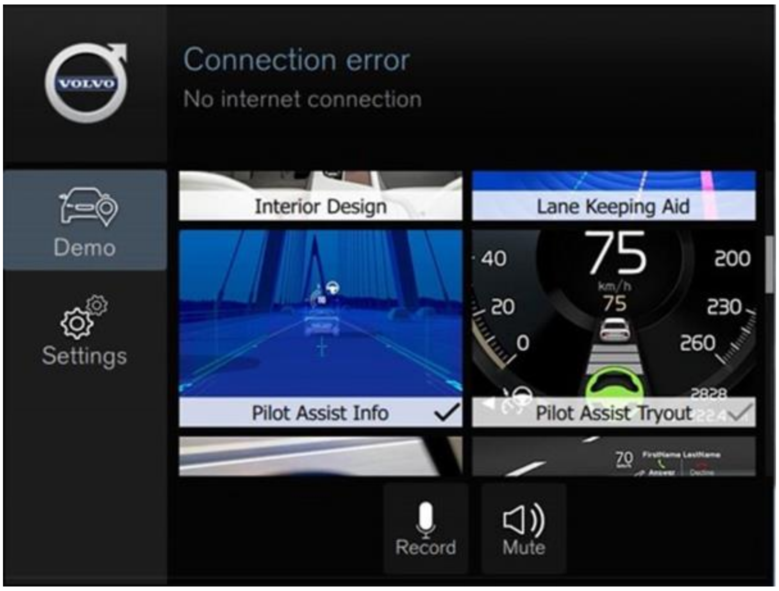

Figure 7.1. Screen shot from In-Car Test Drive app, as presented in centre stack display during test drives with potential customers. The latter are able to choose any topic they liked from those available at any time they liked.

When clicking the "Pilot Assist Info" icon, the following text is read to the driver by the voice interface:

> *"One of the most interesting features of this car is called Pilot Assist. This is a Driver Assistance function that helps to keep the car in its lane and at a safe distance from the car in front. While keeping your hands on the steering wheel, Pilot Assist gives steering support that helps you feel more relaxed during the drive"*

Clicking the "Pilot Assist Tryout" icon prompted the following voice based dialogue (transcript from one participant interaction):

> ***Interactive Quick-guide:*** *"Pilot Assist can be selected by using the buttons on the left side of the steering wheel. Would you like to learn how to activate it?"*

> ***Driver:*** *"Yes, please"*

Interactive Quick-guide: *"Press the middle button to activate Adaptive Cruise Control (ACC). This will keep your car at a set speed and safe distance to the car in front. Are you ready for the next step?"*

Driver: *"Yes"*

Interactive Quick-guide: "Now, press the right arrow to switch to Pilot Assist and keep your hands on the steering wheel. If Pilot Assist is active, the driver display will show a green steering wheel. The car helps you keep the car in the correct lane. Did it work?"

Driver: *"Yes, it worked"*

Interactive Quick-guide: *"Great! You can adjust the speed by pressing the plus and minus button. Remember that Pilot Assist only works with your hands on the steering wheel"*

In total, three pilots were performed, one in Sweden (30 test persons), one in the US (10 test persons) and one in England (6 test persons). The drives were either scripted (i.e. following a pre-determined route) or free (test persons choose route themselves). A test leader was always present in the car to answer any additional questions and sort out potential hiccups. After the drive, test subjects were interviewed and also asked to fill in a questionnaire.

7.2 Results

The results from the pilot tests were interesting in many aspects. First of all, the test persons in general ranked the usefulness of the *Interactive Quick-guide* as high (avg 4.4 on a 1 to 5 scale). Several remarked that they learned about car functions they would not otherwise have known about. Also, many participants wished that such a feature would be available at all times in the car, to be able to learn about functions also after bringing the car home from the dealer.

In terms of the specific MeBeSafe Pilot question, the user interaction results were good. The interview material shows that the group of test participants as a whole spanned the full whole spectrum from intensive to non-users regarding ACC/PA. Most importantly here, when looking specifically at the users who could be characterized as non-users with low interest in new car technology, several explicitly stated that they would never have tried to use the ACC or PA function if they had not been prompted by the *Interactive Quick-guide*. Thus, the app based coaching was highly successful, in the sense that a first time use situation was accomplished for a number of individuals who would otherwise never have tried to activate the function.

On the other hand, several persons in that group reported that testing ACC and PA felt scary, even with a test leader in the vehicle. They also had a number of follow up questions on, and a desire to discuss, what the function could and could not do (possibly best characterized as a type of mental familiarisation procedure). It is therefore not certain that they actually would have followed through with ACC and PA activation if the test leader had not been present in the vehicle.

Test drivers also had numerous interaction problems with the *Interactive Quick-guide*. Some of this was purely technical and mainly due to the relatively limited capacity of the speech interaction engine used in the pilot. It does not have the same capacity for speech recognition as e.g. Apple or Google's counterparts, so there were lot of stops and start overs due to erratic speech recognition. Also it does not have the same capacity for natural language interaction in comparison to the Apple or Google counterparts. This meant that attempts to ask open type questions in the same way people are becoming accustomed to interact with e.g. Siri and Alexa went unanswered. For example, in these pilots, one participant asked the car "Hey, what does safe distance mean?" and got no reply.

Interestingly, some interaction problems were not technical but rather seems to follow the user groups. While those interested in car technology had no problems activating ACC and PA, many of the prior non-users had large problems with activation. In several instances, the test leader had to help out and give instructions in order to complete the activation. One interpretation is that non-users with limited interest in advanced car technology do not have any deeper motivation to try to understand what the system is asking them to do. Since their interest and attention is limited, they also do not fully grasp the instructions, despite these being both very clear and sequenced in several steps (see dialogue transcript above).

7.3 Conclusions and way forward

These pilots have several important implications for the field trial. First, while the app in its current format likely is sufficient for the original In-Car Test Drive use case, it will not provide a sufficiently robust and natural interaction for those users who have a limited interest both in working with the app to understand what is meant, as well as in activating functions like ACC and PA. To reach that level, a natural speech dialogue engine with a wider degree of background knowledge available, along with recognition and response performance closer to Siri/Alexa, would have to be implemented. Though not actually difficult per se, such development would require resources beyond those currently available in MeBeSafe (the budget for the first implementation of the *Interactive Quick Guide* amounted to roughly 800,000 euro; a second implementation would likely not be cheaper).

Second, for the results of the field trial to be clear, it is important to avoid possible confounding factors. The core MeBeSafe research question is whether non-users of ACC can be turned into users through coaching. In that perspective, it would be unfortunate if technical activation difficulties were to interfere with the effects of coaching.

Third, while the coaching toward ACC and PA usage was highly successful in these pilot tests, it cannot be ruled out that the presence of a test leader in the vehicle had a large influence. In other words, even if the App had been perfect, some drivers who now did activate ACC and PA may have refrained from doing so if there had been no test leader in the vehicle.

Given these conclusions, it becomes quite clear that the best way forward is to employ a Wizard of Oz approach in the field trial. Wizard of Oz testing is commonly used to understand interaction patterns for functionality which is not yet fully developed. The test participant is lead to believe that they are interacting with a computer based function of some type (such as a self-driving car), while in reality an experimenter (the "wizard") is simulating the behaviour of the application (in the case of self-driving cars, a hidden back seat driver is controlling the vehicle). Sometimes this is done with the participant's a-priori knowledge and sometimes it is a low-level deceit employed to manage expectations and encourage natural behaviours.

In this particular case, test participants that belong to the non-user group, will be coached toward ACC activation through a remote "wizard", i.e. the voice dialogue will be controlled from somewhere outside the car, with inward and outward cameras positioned in the vehicle such that the "wizard" clearly can see what the participants are doing and send computer voice prompts that talks the driver through the activation. To understand the importance of having another person present when doing this, half the group will be coached alone while the other half will have a test leader present in the vehicle.

8 Deviations from Workplan

Because the development of the HGV app had to be stopped before a robust version of the app was finished, we were not able to sufficiently pilot test the effect of the app and consequently the face-to-face coaching (which is based on the app) on safe driving. The reported results are rather a proof of concept and a technical trial for the app than a full pilot.

Consequently, there were no estimates of coaching effects to feed the simulation from BMW. Instead, expert opinion of the effect of coaching was used which, can be argued, is less accurate than estimated effects from a sufficiently large pilot study.

There are small deviations from the Description of Work (DoW) in the sense that the DoW on occasion describes that Volvo will develop an app for drowsiness alert. This is an error in the General Agreement (GA). The Volvo app developed in this project is aimed to coach Volvo drivers to use their ACC more often. The drowsiness alert (not an app) that is developed by Volvo Cars in MeBeSafe will be described in the deliverables of WP2.

As discussed above, the field trial of the ACC coaching app will take the form of a Wizard of Oz trial rather than using the app itself, since too much resources would be required to improve the app to a level where it is suitable for the target group.

9 Conclusion

9.1 Coaching scheme for HGV drivers

The coaching scheme for HGV drivers consists of an online (app based) and offline (face-to-face coaching) part. The face-to-face coaching is supported and initiated by the DriveMate app. Unfortunately we have encountered some delays in the development of the DriveMate app.[6] As a result we only had a first version of the app available for pilot testing; which also included some technical flaws. This delay in app development impacted the pilot test in WP4 considerably.

Because of the delay in development, the app was administered only to 13 drivers in our pilot with a Norwegian truck company. Data was collected for a relatively short period of time, and no face-to-face coaching was initiated[7]. The results reported in this deliverable are therefore not a full evaluation of the effect of the implemented system, but rather a proof of concept and a technical trial for the app. The analysis of this first data from MeBeSafe WP4 does seem to indicate that, with the exception of some errors, the system is in general working as planned concerning the data gathered.

The simulation to estimate the large-scale impact of the coaching scheme on road safety, was also impacted by the delay in the development of the DriveMate app. Because of the delay, no estimated effect of coaching was available as input for the simulation. Instead, different scenarios defined based on expert opinion were used in order to quantify the effect of coaching. The results of this simulation show a safety

[6] This was reported to the PO in November 2018. Because of a reorganisation, Shell – resonsible for the development of the app – no longer has the internal capability for programming software for apps such as the DriveMate app. The programming of the app needs to be out-sourced. For this an amendment is neccesary.

[7] The app is supposed to prepare drivers for coaching, and alert them when a coaching session is due. Because of the technical issues with the app the drivers never received such an alert.

benefit in case coaching affects the desired time gap and the desired velocity. For the other parameters (mean velocity, time driving with short distances, time driving in lower TTC situations, frequency of hard braking manoeuvres) the simulations also show benefits. Furthermore, from this analysis, it can be concluded that it is in general possible to determine the safety impact of coaching measures by means of the simulation approach. However, during the work also different limitations were encountered.

9.2 Pilot test with Volvo drivers

The pilot tests with Volvo drivers were highly successful, in the sense that a first time use situation was accomplished for a number of individuals who would otherwise never have tried to activate ACC. On the other hand, it also became clear that a much more sophisticated app in terms of speech recognition and dialogue capabilities would be required to meet the needs of the intended target population. Also, several persons classified as belonging to the target population reported that testing ACC and PA felt scary, even with a test leader in the vehicle. It is therefore not certain that they actually would have followed through with activation if the test leader not been present in the vehicle.

9.3 Implications for WP5 field tests

These pilots have several important implications for the field trial. The DriveMate app (as part of the coaching scheme for HGV drivers) needs to be improved considerably before it can be used in the field trial. With the current (V1) version we expect to achieve only a small effect of online and offline coaching, which will be extremely hard to demonstrate in the field trial. The development of the app is dependent on a pending amendment request.

For the ACC coaching objective, it was concluded that the best way forward is to employ a Wizard of Oz-approach in the field trial. In this approach, the test participant is lead to believe that they are interacting with a computer based function of some type, while in reality an experimenter (the "wizard") is simulating the behaviour of the application. This allows for a coaching dialogue that does not fall through due to technical limitations in speech and dialogue processing, and hence makes it possible to understand to which degree the target group of non ACC users are coachable into ACC usage, without spending a prohibitive sum of money on additional app development first.

References

ACEA (2017). *ACEA Report Vehicles in use - Europe 2017*. European Automobile Manufacturers Association, ACEA, Brussels. https://www.acea.be/uploads/statistic_documents/ACEA_Report_Vehicles_in_use-Europe_2017.pdf.

af Wåhlberg, A., de Craen, S., Ljung Aust, M. & Neuhuber, N. (2019). *Final coaching scheme (Deliverable 4.3)*. Retrieved from MeBeSafe website: https://www.mebesafe.eu/results/.

af Wåhlberg, A.E. (2007a). *Aggregation of driver celeration behavior data: Effects on stability and accident prediction.* In: Safety Science, vol. 45, nr. 4, p. 487-500. https://doi.org/410.1016/j.ssci.2006.1007.1008.

af Wåhlberg, A.E. (2007b). *Long-term effects of training in economical driving: Fuel consumption, accidents, driver acceleration behavior and technical feedback.* In: International Journal of Industrial Ergonomics, vol. 37, nr. 4, p. 333-343. https://doi.org/310.1016/j.ergon.2006.1012.1003.

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. https://doi.org/310.1080/14639220701596722.

af Wåhlberg, A.E. (2012). *Changes in driver celeration behaviour over time: Do drivers learn from collisions?* In: Transportation Research Part F: Traffic Psychology and Behaviour, vol. 15, nr. 5, p. 471-479. https://doi.org/410.1016/j.trf.2012.1004.1002.

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

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

DESTATIS (2018). *Verkehr – Verkehrsunfälle 2017.* Statistisches Bundesamt (Destatis).

Fahrenkrog, F., L., W., Rösener, C., Sauerbier, J., et al. (2017). *Impact analysis for supervised automated driving applications*. AdaptIVe Deliverable

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

Grant Agreement - MeBeSafe *Grant Agreement number: 723430 — MeBeSafe.* H2020-MG-2016-2017/H2020-MG-2016-Two-Stages.

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

KBA (2019). *Erneut mehr Gesamtkilometer bei geringerer Jahresfahrleistung je Fahrzeug.* Kraftfahrt-Bundesamt (KBA). Retrieved from https://www.kba.de/DE/Statistik/Kraftverkehr/VerkehrKilometer/verkehr_in_kilometern_node.html in June 2019.

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

Malta, L., Ljung Aust, M., Faber, F., Metz, B., et al. (2012). *Final results: Impacts on traffic safety*. euroFOT Deliverable D6.4.

Treiber, M., Hennecke, A. & Helbing, D. (2000). *Congested traffic states in empirical observations and microscopic simulations.* In: Physical Review E, vol. 62, nr. 2, p. 1805–1824.

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

Van Noort, M., Taoufik, B., Fahrenkrog, F. & Dobberstein, J. (2015). *SIMPATO - The Safety Impact Assessment Tool of interactIVe.* In: IEEE Intelligent Transportation Systems Magazine, vol. 7, nr. 1, p. 80-90.

Varlamov, O., Bakker, B., Neuhuber, N., Wahlberg, A.a., et al. (2019). *App to induce behavioral change (Deliverable 4.4).* Retrieved from MeBeSafe website: https://www.mebesafe.eu/results/.

Volvo Trucks (2013). *European Accident Research and Safety Report*. Volvo, Gothenburg. http://www.volvogroup.com/content/dam/volvo/volvo-group/markets/global/en-en/about-us/traffic-safety/ART-report-2013.pdf.

WHO (2018). *Global status report on road safety 2018: summary*. (WHO/NMH/NVI/18.20). License: CC BY-NC-SA 3.0 IGO. World Health Organization, WHO, Geneva.

Annexes