Delivery Report for

MeBeSafe

Measures for behaving safely in traffic

Deliverable Title	Trial Design
Deliverable	D5.1
WP	WP5 Field evaluation
Task	Task 5.1 Trial design

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

Disclaimer

Copyright

Grant Agreement No.	723430
Project Start Date	01/05/2017
Project End Date	31/10/2020
Duration of the Project	42 months
Deliverable Number	D5.1
Deliverable Leader (according to GA)	Volvo Cars
WP Leader	Mikael **Ljung Aust**, Volvo Cars
Deliverable Leader(s)/ (Editor(s))	Mikael **Ljung Aust**, Volvo Cars
Dissemination Level (Confidentiality)	Public
Nature	Report
Status	Final
Due Date	M27 (July 2019)
Main Author(s)	Mikael **Ljung Aust**, Volvo Cars
Other Contributing Author(s) *alphabetical order by last name	Saskia **de Craen**, SWOV Anna-Lena **Köhler**, ika/ RWTH Aachen University Olaf **op den Camp**, TNO Pontus **Wallgren**, SAFER/ Chalmers University
Reviewer(s)	Mikael **Ljung Aust**, Volvo Cars
Formal review and upload by Coordinator	Institute for Automotive Engineering (ika) – RWTH Aachen University

Please refer to this deliverable as follows:

Ljung Aust, M., et al. (2019). Trial Design (Deliverable 5.1). Retrieved from MeBeSafe website: https://www.mebesafe.eu/results/.

Abstract
The main objective of WP5 is to run a set of field trials with naïve users (i.e. not experts involved in the development of the measures) for all nudging and coaching measures developed in WP2-4. This deliverable describes the designs of those field trials, on a per project objective basis.

Version	Date	Comment
1	15 June 2019	Draft
2	01 July 2019	Draft for English check
3	05 July 2019	Draft for internal review
4	15 July 2019	Edited version after internal review
5	29 July 2019	1st formal review ika
6	30 July 2019	Final corrections
7	31 July 2019	2nd formal review ika

table of document history

Table of Contents

List of Figures

List of Tables

Acronyms

1 Executive Summary

The main objective of MeBeSafe WP5 is to run a set of field trials with naïve users (i.e. not experts involved in the development of the measures) for all nudging and coaching measures developed in WP2-4. This deliverable describes how these field trials have been designed for each MeBeSafe objective. The descriptions cover trial design and duration, test population sizes, participant recruitment strategy, data collection parameters, interview guides, site properties for site based nudges, data analysis plans and finally which criteria should be met for each measure to be considered successful.

2 Contribution by each Partner

This deliverable was written by Mikael Ljung Aust, with contributions by Olaf op den Camp, Anna-Lena Köhler, Saskia de Craen and Pontus Wallgren. All other partners in WP5 contributed by giving their feedback on the development of the trial designs and reviewing this deliverable. All partners have fulfilled their tasks in time and with satisfactory quality.

3 Introduction

The objective of Task 5.1 is to define the trial protocols for each field trial that will be run in WP5. Defining a trial protocol includes the following items (see Grant Agreement):

- Calculating the minimum trial duration required to be able to assess each measure, given how frequently it will be administered
- Calculating the population size required for assessing each measure
- Defining a recruitment strategy that with reasonable effort gives access to test participants without unduly biasing the sample
- Define the data collection parameters, i.e. what data needs to be collected and with which resolution and accuracy, etc. to successfully complete the analysis following data collection.
- Develop interview guides and/or questionnaires as appropriate
- For site based nudging, define required properties of suitable sites where the site based nudges can be implemented
- Define a data analysis plan, i.e. specify which post-processing of the collected data is required, and how to calculate performance indicators that can be used to determine the effectiveness of each measure
- Establish predetermined thresholds for whether the evaluated measure is effective in terms of influencing driver behaviour in the intended way, i.e. determine what the necessary level of impact on behaviour is for the measure to be judged successful

The exact details for each of these points differ slightly for each of the Objectives which MeBeSafe has set out to address. Hence, the trial protocols will be described on a per objective basis in the following.

4 Study plan for O1: Driver alertness feedback (Volvo Cars)

4.1 Trial design and duration

The study design of the field trial is a between group design where the nudge is the independent variable introduced, and the averaged response to the nudge over repeated exposures is what will be studied.

Empirical data pulled from Volvo driver show that Driver Alert warnings are not administered very frequently. There is large between driver variability, but on average, these alerts are issues on average once every 1000 km, or about once a month if you're an average driver in Sweden. To secure a sufficient number of exposures to the nudge, the field trial therefore needs to run for at least 6 months to capture any population change due to the nudge.

4.2 Population size

Previous studies of response to the Driver Alert suggestion of taking a break show that compliance is very low in the population. Hence any positive responses to the study will result in medium to large effect sizes, in the sense that any drivers stopping in response to the nudge gives a significant difference from baseline behaviour. With a 6 month field trial and assuming one exposure per month, this gives us 6 exposures per driver in the treatment group. A conservative power analysis suggests that relatively small groups would suffice (i.e. n=20 or so). However, to have some statistical margins available, we aim to include 100 drivers in total (50 in treatment, 50 in baseline).

4.3 Participant recruitment

Drivers will be recruited from within the Volvo Cars company car fleet, with special attention on avoiding selection of drivers who are directly involved with development of Advance Driver Assistance Systems.

4.4 Data collection parameters

For every Driver Alert warning issues, a data segment including the alert and the following ten minutes will be collected. The segment will include a number of vehicle variables, the most important of which are speed, position, state of the ignition and also state of the Driver Alert algorithm.

4.5 Data analysis plan

The Key performance Metric for the field trial is whether drivers actually stop the car (defined as speed going to zero and staying there for more than one minute, or ignition goes to an off state) within the 10 minutes following the Driver Alert.

Instances of stopping will be controlled for position, in order to remove instances where the driver was able to reach the intended destination within those 10 minutes, and thus where the influence of the nudge is not possible to determine.

Finally, the state of the Driver Alert algorithm will be analysed, to determine if the nudge may have an effect on drivers' level of alertness even if they do not stop the car.

4.6 Criteria for success

If drivers receiving the nudge actually stop in at least one of three instances where they get the Alert, the nudge will be determined as successful.

5 Study plan for O2: Usage of safety ADAS to prevent close following (Volvo Cars)

5.1 Trial design and duration

The study design of the field trial is a within group design where the nudge is the independent variable introduced, and the level of Adaptive Cruise Control (ACC) usage over time is what will be studied. Since ACC is a support system to normal driving; changes in the level ACC usage will show up fairly quickly (e.g. it will represent a proportion of the total driving that the test participant is doing, which normally is several hours per week). However, to secure sufficient exposure to the nudge, and to try to capture any plateaus or other changes in ACC usage in response to the nudge, the field trial will run for (at least) 4 months.

5.2 Population size

Since this is a within group design with a high number of data points per participants (e.g. average ACC usage per week over a 4 month period), a group of 15 drivers or more would be sufficient to run the study.

However, empirical data from current Volvo drivers show that the average usage of ACC varies greatly between users; spanning from complete non-users (even though they have the function in the car) to very high level users (i.e. ACC is on 40-50 % of their total driving time). Both ends of this scale represent a challenge to the field trial. The high level users are already using ACC as much as it can reasonably be used; hence nudging will not change their usage level, and they are not in need of nudging to keep up the usage either. Drivers with this usage characteristics therefore need to be removed from the data set in order to not bias the study of the nudge effectiveness.

On the other end of the scale, interviews conducted with non-users suggest that something more than a nudge is required to change them into ACC users. They have an inherent scepticism toward this type of support systems, and cannot bring themselves to try to activate them, because they do not feel sufficiently in control to do so. For this reason (for further details on this, see D4.3), drivers who shows a zero usage characteristic will be selected for the ACC coaching program (see study plan for O4 below).

Given the likely need to exclude some drivers at both ends of the scale, the population size for this trial will need to be larger than 15; i.e. more in the 30-35 drivers range to have a sufficient margin for exclusion of super users as well as non-users.

5.3 Participant recruitment

Drivers will be recruited from within the Volvo Cars company car fleet, with special attention on avoiding selection of drivers who are directly involved with development of Advance Driver Assistance Systems (ADAS).

5.4 Data collection parameters

For each driver and drive, the state of ACC (i.e. whether it is active or not), along with a number of vehicle variables such as speed, time headway and whether other ADAS are active or not are will be collected at min 10 Hz.

5.5 Data analysis plan

The basic data analysis consists of analysing ACC usage ratio over various time segment, e.g. per trip, per week, per 500 km driven, etc. Two ACC usage ratios and one average time headway index will be calculated for each time segment, both over total driving duration and over total driving distance. For all drivers, the per time unit series of ACC usage ratios and average time headway will then be analysed for significant changes from the start to the end of the nudge exposure. Drivers who early on show zero usage will also be referred to the ACC coaching study, as discussed above.

5.6 Criteria for success

If the ACC usage among drivers who show low ACC usage at onset of the field trial gas increased with 40% at the end of the field trial (i.e. excluding high users and non-users from the analysis as discussed above), the nudge will be deemed a success. Similarity, if time headway shows a 40 % increase or more over time in the same group while ACC is active, the nudge will be deemed a success.

6 Study plan for O3: Attention to potential hazards in intersections (TNO)

The effectiveness of different nudging HMI's in terms of being able to direct driver attention toward a specific area have already been assessed in a driving simulator study, the complete results of which are reported in D2.3.

A key takeback from the simulator study is that at least one HMI design was successful in redirecting driver attention toward potential hazards. The objective of the field trial therefore becomes to validate the simulator study outcomes for drivers in real traffic on public roads. In particular, the field trial will focus on assessing drivers' responses to in-vehicle warnings from the developed HMI that will be triggered by the static part of the developed hazard model, i.e. the part of the attention redirection system which aims to direct driver attention towards the location of a possibly approaching cyclist that has yet to become visible, due to view blocking obstructions.

6.1 Trial design and duration

The field trial will be run as a within-subjects study. A fixed route through the inner city of Eindhoven will be generated where each subject passes a total of 30 busy, non-signalized, non-priority intersections. The route will take 30 to 45 minutes to complete.

The first independent variable in the study is whether the nudging HMI is triggered or not (per intersection). The second independent variable is the level of pre-drive instructions given to the subjects. Half of the subjects will receive clear instructions on the presence of the nudging HMI and its intended function, while the other half will remain naïve, not receiving information about the nudging HMI and its intended function until at the debriefing after the drive.

Three HMI triggers will be used: true positives (a nudge is displayed in an intersection with view-blocking obstructions), false positives (a nudge is displayed without the presence of an intersection), and false negatives (no nudge is displayed, even though the intersection has view-blocking obstructions). The HMI triggers will be randomised over the intersection exposures, but with a bias toward placing the false positives at the end of the drive, to avoid negative "cry wolf"-effects on the true positive events.

6.2 Population size

Since each study participant receive every stimulus, the field trial effects will be possible to analyse using Analysis of Variance tests (ANOVA). A medium effect is expected in the study; hence an effect size of 0.5 is selected. This results in a good power estimate (over 0.9) for 16 participants or more. To have a bit of margin, a population of 20 subjects will therefore be the target.

6.3 Participant recruitment

Drivers for the field trial will be recruited from TNO employees, or employees at the Automotive Campus in Helmond. For running a first pilot in September, 6 people will be required. For completing the complete FOT for one vehicle, 14 to 18 additional people will be needed. Participants will be recruited out of a pool of people that are not involved in the nudging HMI development.

6.4 Data collection parameters

Data collection will be done in the vehicle in which the nudging HMI is integrated. The following time dependent signals will be recorded from the vehicle with a frequency of 10 Hz or more:

- Vehicle location (GPS)
- Vehicle speed and acceleration, brake and gas pedal position, angular steering position from the vehicle CAN-bus

- The time evolution of the HMI triggers
- The gaze direction of the driver analysed from a context camera directed at the driver's face
- The object level data of the road users detected by the vehicle's sensor system: type, position and speed of the most important objects relative to the ego-vehicle
- The video images of the various forward-looking context cameras.

All data will be synchronously recorded, to avoid the need for retrospective synchronisation.

6.5 Instructions and questionnaires

Both instructions and questionnaires will be developed. The instructions will inform the instructed subject group on the experiment and the HMI that they will be subjected to.

All subjects will receive a debrief after the drive, and experiences from the drive will be collected by means of a questionnaire.

6.6 Required site properties

Though for O3 we study an in-vehicle nudge and not a site-based nudge, the sites for testing this nudge have several requirements to be suitable for testing. A route through the inner city of Eindhoven is chosen, as here many non-priority intersections are found, several with a typical view-blocking obstruction from houses only separated from the road by a narrow sidewalk. Many cyclists make use of the roads (not the sidewalk), as not many separate cyclist lanes are present in this part of the city.

6.7 Data analysis plan

It will be analysed how the gaze direction of the driver is influenced by the HMI with directional information on where the hazard potentially will appear. The gaze direction distribution over time (probability density function of the gaze direction) will be determined for the last 6 seconds of an approach to each intersection. The width of the distribution, or excursions of the distribution in the direction of a potentially indicated hazard will be studied.

It will be related to the presence or absence of an HMI trigger, the presence of other road users, and the presence of view-blocking obstructions. The possible response of the driver in brake and gas pedal will be investigated as well.

6.8 Criteria for success

In case the gaze direction distribution over the last 6 seconds before an intersection is wider when the driver is being nudged by the HMI, or shows convincing evidence in another way that the driver indeed pays attention in the direction indicated by the HMI, the HMI nudge will be considered successful.

7 Study plan for O4: Behavioural change through online private driver coaching (Volvo Cars)

Interviews with drivers who do not use ACC in their ACC equipped cars show that they are afraid of activating ACC, because they do not trust it to be capable of actually regulating speed and the distance to lead vehicles, and are worried something might go wrong if they were to try (see D4.3 for further details).

It is therefore clear that coaching of the drivers being exposed to the nudge under O2 has the largest potential not on drivers who are already using ACC to some extent, but rather on drivers who do not use ACC at all, because they are afraid of delegating vehicle control. Non-users must become users, and this is where coaching can be particularly helpful and effective.

7.1 Trial design and duration

For drivers who currently display zero ACC usage, trial duration can be very short, as any usage would constitute a large change in behaviour compared to what they were doing previously. Also, it is possible that it may take more than one coaching session to motivate them to become ACC users. Given that, the field trial will last for a minimum of 3 months, with one coaching session at the start of each month (given that the zero usage pattern prevails after previous coaching session).

7.2 Population size

This population will consist of the parts of the O2 that fulfil the non-usage criteria. While it is not possible to foresee exactly how many drivers that will be, it is estimated that at least 10 drivers from that group will fulfil this criteria, and that will be sufficient to analyse the effect of coaching.

7.3 Participant recruitment

Drivers will be recruited from within the Volvo Cars company car fleet, with special attention on avoiding selection of drivers who are directly involved with development of Advance Driver Assistance Systems.

7.4 Data collection parameters

For each driver and drive, the state of ACC (i.e. whether it is active or not), along with a number of vehicle variables such as speed, time headway and whether other ADAS are active or not are will be collected at min 10 Hz.

7.5 Data analysis plan

The basic data analysis consists of analysing ACC usage ratio over various time segment, e.g. per trip, per week, per 500 km driven, etc. Two ACC usage ratios will be calculated for each time segment, both over total driving duration and over total driving distance. For all drivers, the per time unit series of ACC usage ratios will then be analysed for any non-zero segments, indicating that the driver actually has started using ACC after being coached.

7.6 Criteria for success

If 50% or more of the drivers in the non-user group activates ACC one or more times after being coached, the nudge will be determined as successful.

8 Study plan for O5: HGV driver behavioural change through coaching (Shell)

8.1 Trial design and duration

There is no generally accepted minimal length of coaching interventions, neither concerning time frame, nor number of sessions. In the general coaching literature, there is a wide span of these two parameters between studies (four to fifty-two weeks and one to ten sessions in the meta-analysis by Jones, Woods & Guillame, 2016). But more interesting, no increased effect over time/sessions has been found (Jones, Woods & Guillame, 2016). For driving-related coaching, the situation is similar (e.g. Bell, Taylor, Chen, Kirk & Leatherman, 2016; Stanton, Walker, Young, Kazi & Salmon, 2007), although the interventions do seem to be somewhat longer.

So the literature is inconclusive on the minimum amount of coaching necessary for behavioural change. We will therefore aim for the maximum length of the trial period possible, given the anticipated end of the MeBeSafe project and the start of WP5 (see timeline – Annex 1). We will include a control group (which starts 2 months later than the experimental group); to control for experiment bias. After a baseline measurement of two months, the experimental group will receive 5 months of online and face-to-face coaching. The control group will receive 3 months of online and face-to-face coaching.

8.2 Population size

Deliverable 4.2 (Appendix A) included power estimations for different variables in this field test. The conclusion is that we need about 100 (for celeration variable) and 500 (for harsh braking) drivers. We conclude that it is practically feasible to include 100 drivers in the study (corresponding with the GA: "multiple HGV fleets from different countries"). So we are aiming for ~100 drivers from different countries:

- ~ 50 Litra – Norway

- ~ 20 Bertschi – UK
- ~ 35 ABS Bonifer - Hungary

8.3 Participant recruitment

Shell will utilize its network of hauliers. These transport companies will recruit and supervise drivers for participation in the field test (see Annex 2 for the road map for participation in the field test).

8.4 Data collection parameters

Data will be collected from the phones used by the drivers, including GPS position, time and events. The information captured by the Drivemate app is detailed (1Hz) GPS position data for each driver, continuously recorded while the vehicle is moving. From this and time, speed and acceleration are calculated. These two variables are the main sources of data which will be used in post-processing. The measurements are planned to continue for a period of about five months. This means that there will be thousands of files with millions of data points to process into files which can be analysed.

8.5 Interview guides and questionnaires

As part of the coaching material a survey on driver competences has been developed (see deliverable 4.3 Section 4.5.13). In addition, a questionnaire for the participating drivers has been developed, which focuses on user experience and reactions. This survey will be distributed to the drivers at the end of the trial.

8.6 Data analysis plan

The research question is whether a change in driver behaviour can be seen when coaching is started. This means that the data need to be categorized into pre- and post-coaching sections. However, as it is not possible to know exactly when an effect might happen, or if it might decline over time, the data need to remain disaggregated over reasonable time periods. The plan is therefore the aggregate data per week, i.e. overall numbers per driver will be calculated, which represent the average behaviour of the driver each week.

The variables calculated will be *celeration* (the average of all absolute acceleration values when the vehicle is moving; af Wåhlberg, 2008); this can be regarded as a measure of smoothness of driving. And the number of times acceleration and deceleration exceeds certain thresholds (*harsh braking* and *acceleration*). However, as 'harsh braking' has no accepted definition, the exact values for this type of variable have not been established. The general plan is to use several different cut-off values, to capture different levels of severity.

These three indicators (celeration, harsh braking and acceleration) have been chosen on different grounds. Celeration has a proven correlation with crash involvement (af Wåhlberg, 2006) and is therefore a valid indicator of traffic safety. Harsh braking has rarely been associated with crashes, but is specified in the grant agreement as the main indicator of the project (similar to harsh acceleration), and is therefore included as an indicator. Performance indicators will therefore be calculated as percent change in behaviour over time on the three variables described.

8.7 Criteria for success

There is no accepted level for effectiveness in behaviour change. Rather, research in social science is usually occupied with the question of whether an effect exists or not. However, for applied problems, it can always be asked whether the resources used to implement something could be better used in some different way to achieve the same kind of result.

The end result of MeBeSafe is traffic safety, and the number of traffic crashes which could be prevented by the app and coaching system developed in WP4 could therefore be evaluated as to this number in relation to the costs of implementing the system. If the system yields more of a reduction per resource put in than other systems or interventions, it can be considered to be successful.

However, such comparisons are difficult to make, because similar systems have rarely been evaluated as to their effects on crashes, and the costs have apparently not been published at all. Any reduction of harsh braking and celeration values is probably beneficial for safety, but to estimate the efficiency of the system, a calculation is needed which shows the costs of the system as compared to the savings for crashes. Such a result can be compared to those for other systems, so that the most efficient system can be chosen by hauliers.

To be able to calculate the efficiency, three numbers are needed; the cost of implementing and running the MeBeSafe coaching system, the average cost of truck crashes, and the estimated effect on crashes given some different levels of behaviour change. All of these numbers will be estimated with a rather large degree of uncertainty.

The costs for the system include phones/tablets, time used for drivers and managers. A phone, cradle, power adapter, and one year of data transfer would be about €300. Each driver would use a maximum of one hour per week for coaching, reading and checking trip data, plus three hours for the initial start-up. A supervisor would use ten hours to set up the scheme and phone, hold the introductory meeting and a few further report meetings with the drivers in the first year. The total cost will be calculated for one year. As the initial costs are one-off, the long-time costs will be lower. With a cost of €40/hour for drivers and €50/hour for a manager, the total cost per driver in the first year would be 300+(55*40)+(10*50)=€3000. If there is no license fee for DriveMate, the cost per driver would thus be a maximum of €3000 per year.

The costs of truck crashes can be estimated from a number of published papers on these from various countries. However, these costs are usually given as totals, i.e.

the costs for the hauliers are not given separately. However, they can be estimated to be about half of the total.

Study	Country	Average cost for all trucks	Average cost for big trucks
Zaloshnja & Miller, 2004	US	52 000 (total cost) Updated by FMCSA 2008: 131 000	78 000 (total)
Al-Masaeid, Al-Mashakbeh & Qudah, 1999	Jordan	2 600 (cost of damage only, in fatal crashes)	-
Farah, 2017	Nigeria	56 000 (cost of vehicle damage only)	

Table 1: Summary of estimated total costs per truck crash in some different countries. Estimates given in euros, not adjusted for changes in monetary value over time

From Table 1, it can be seen that estimated costs for crashes differ very strongly between sources, and that the sum put into any calculation is heavily dependent upon exactly what kind of cost, and to whom, is of interest.

Turning to the association between changes in behaviour and crashes, the only available estimate for the variables used in this project is for celeration (af Wåhlberg, 2008). From data used to the validate the concept (af Wåhlberg, 2007), it can be estimated that a ten percent reduction in celeration level would decrease the number of crashes by about twenty-five percent. This means that if truck drivers have on average one crash in five years (costing the company €60 000), a ten percent reduction in celeration by the use of DriveMate would save roughly as much money as the investment in the coaching would cost.

This estimate is of course very sensitive to the exact numbers used; the gain would be larger in companies with a higher mean crash rate and higher costs per crash. Similarly, hauliers with lower wages would have more to gain. Also, this calculation has not taken into account the gain for society. If this is included, the break-even point of DriveMate would be five percent.

9 Study plan for O6 and O7: Safe speed/trajectory on inter-urban roads (ika)

9.1 Trial duration an design

This field trial has two complementary parts, where a larger, quantitative, site based study will be enhanced by qualitative data collected by means of questionnaires from neighbouring households and by recruited drivers. In the quantitative study, the speed of the vehicles entering the motorway exit is measured and, dependent on whether that speed exceeds a certain threshold, nudging is applied (for details, please see subsequent Deliverables D3.2 and D3.3). The vehicle trajectory will be measured as well. The key performance indicator is whether the vehicle slows down sufficiently in response to the nudge. This is relevant for both, safe speed and guiding drivers along a preferred trajectory.

One qualitative survey will be conducted to collect information about the general perception of the light system installed at the motorway exit and acceptance of the nudge by car drivers that frequently use that exit.

However, these two data sets can not be linked to each other, since the qantitative data collection only uses thermal imaging cameras and cannot read licence plate numbers (for data privacy reasons). Thus, in order to get direct feedback from drivers right after they have experienced the system, an additional on-site survey will take place. Also, a number of separately recruited participants will drive through the motorway exit with and without being nudged, and will then be surveyed after their drive in order to gain further insights into their perception of the field test location and the nudge. Here, a direct link between quantitative data collection and qualitative questionnaire will be possible.

The first testing phase will start as soon as the calibration phase is completed in order to ensure that the system is working as intended. In this testing phase, the most promising nudging interventions as identified in WP3.1 and described in both, D3.2 and

D3.3, are tested in a period of one full week each, in order to gain insights whether the nudging intervention has a different effect on different days of the week for e.g. traffic density reasons or at different times of day for e.g. differences in available daylight and the resulting contrast with the light sources. After this, an iterative process will ensure that the specifications that work best in the field are identified under real traffic conditions. The qualitative and quantitative data collections will then start at the same time.

9.2 Population size

As the measure and the test will target every driver passing the exit and exceeding the speed limit thresholds for the defined section (see D3.2 and D3.3), the number of drivers passing the exit was measured in a baseline measurement by ISAC before the location was chosen for the field test. During the baseline measurement, the average daily traffic volume through the exit (on working days) was 4,700 vehicles per day. Since the road is mainly used by commuters, the traffic volume on weekends is expected to be significantly lower.

For a reliable statistical estimation, a power analysis using G*Power for F-tests (ANOVA for fixed effects, main effects and interactions) with a small to medium effect size at the worst and a confidence interval of 95% revealed a total sample size necessary of 196 participants for expected power 0.80. As stated above, the field test will be much over this threshold, so statistical power is expected to be sufficient.

The qualitative data collection is an additional means to get further feedback on the nudging system and not the main dependent variable. In the study with recruited drivers, based on the number of between-factors, a group of minimum 20 drivers per between-variation will be recruited for controlled trials (G*Power estimation for a small to medium effect with estimated power of 0.80 with 3 measurements).

To calculate the ideal sample size for the qualitative questionnaire that will be distributed online, the population size should be determined. It is expected that most drivers driving through the curve do this daily or at least multiple times a week, therefore the population size is set on the number of vehicles in the curve throughout one day (N= 4,700). With a confidence level of 99% and a confidence interval of 5%, the ideal sample size would be 625 respondents.

9.3 Participant recruitment

For the overall quantitative measurement, no distinct recruitment of drivers is necessary, as the field test will be conducted with all drivers who pass the site (situated FOT as stated in the Grant Agreement).

For an additional study with recruited participants in order to directly link quantitative and qualitative data, participants can be recruited from the Netherlands and Germany. As stated in the Grant Agreement, private drivers will be recruited from a wide age range and will have a minimum driving experience of 2 years of recent driving experience with a minimum of 5000 km driven per year.

For on-site surveys, drivers will be asked if they would like to participate in a quick survey after verification that they did encounter the nudging system on a voluntary basis after the nudging intervention. It will be ensured that stopping and surveying drivers happens at a safe location close to the site.

For the qualitative survey to be distributed among residential areas surrounding the test site, the Digi panel from the city of Eindhoven is used, which is a way to contact the citizens of Eindhoven to give feedback regarding relevant issues in the city, as well as the communication channels of the participation project *JouwLichtop040*. Target populations are selected based on postal code. Furthermore, area managers for the area's Tempel, Blixembosch, Heesterakker and en Esp in Eindhoven are included in the recruitment strategy.

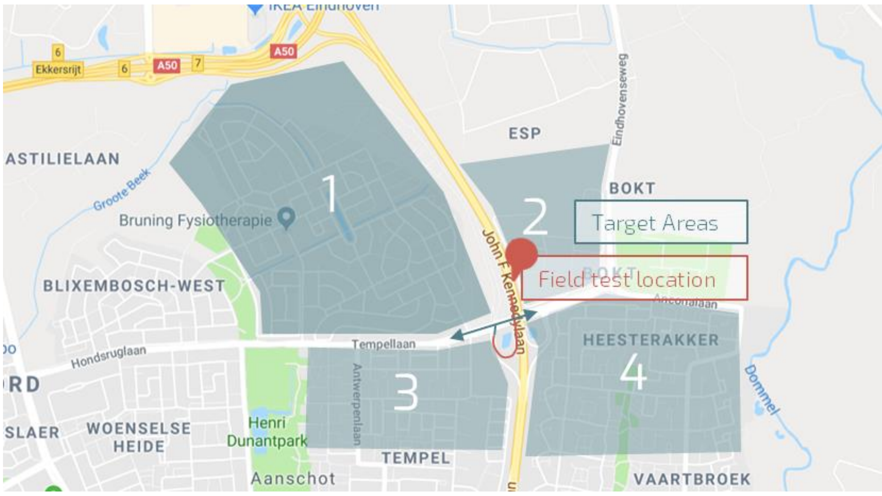

Figure 1: Digi Panel survey areas in Eindhoven around the test site

9.4 Data collection parameters

For the quantitative data collection, the following metrics are collected: timestamp, vehicle ID, length and width of the vehicle, current speed at each timestamp, from which the acceleration/deceleration can be calculated post-hoc, the current position on the street, both longitudinal and lateral, whether the vehicle was nudged, in which nudging scenario it was nudged and the brightness of lights. Weather conditions during the measurements will be added to the data post-hoc.

The dependent variables are speed for O6 and lateral position of the vehicles on the street for O7. The data recording rate is 30 frames per second and position and speed are updated every 0.2 seconds.

9.5 Interview guides and questionnaires

The onsite survey will be designed to be completed by drivers that are asked right after they took the motorway exit and triggered the nudging system. The online questionnaire used for the qualitative survey is designed to be completed by participants who live in or around Eindhoven and take the motorway exit regularly.

The qualitative survey for recruited drivers aims at assessing the subjective impression of the field test and directly linking these to the quantitative data. Both, instructions and questionnaires will be developed within WP5.

9.6 Properties of selected site

The testing location is the motorway exit J.F. Kennedylaan – Tempellaan (GPS coordinates: 51°29'04.6"N 5°29'21.2"E) in Eindhoven, the Netherlands. The motorway exit itself was selected as a suitable location, as it fulfilled the given requirements, compared with other exits, the best. The requirements included the possibility to improve traffic safety with a nudge, low traffic complexity, good circumstances for installation and maintenance, and accessibility and potential impact. For details of the location selection process, please see D3.2.

9.7 Data analysis plan

The purpose of the data analysis of quantitative data is to demonstrate that the nudging intervention is effective to achieve the objectives, in particular reducing speed (O6) and guiding drivers along a preferred trajectory (O7). For objective O6, the speed distribution at different cross sections along the motorway exit is of interest. The measurement system will collect the speed of drivers at least every 0.2 sec., from which longitudinal acceleration and deceleration can be calculated. For objective O7, the lateral position on the road is focused on. Whether a trajectory can be considered safe is highly dependent on the speed driven in a certain situation (for details see D3.2). In order to address this, drivers are clustered into different speed profiles and the ideal, safe trajectory will be defined.

Quantitative and qualitative data will be analyzed in an appropriate way in order to ensure meeting scientific standards. Quantitative data such as speed and lateral position on the street will be analyzed with appropriate measures of inferential statistics such as analysis of variance (ANOVA) or, where applicable, multifactorial

ANOVAs, in order to be able to draw reliable and statistically proven conclusions. This will be done according to the hypotheses on expected behavior, which will be adapted to the hypotheses within task 3.2 (see D3.2).

For qualitative questionnaire data, the different types of questions are analyzed depending on the format of the question. For demographic data, descriptive statistics are applied. For questions from validated questionnaire tools, the required analyses as stated in the corresponding manuals will be performed. Responses from open questions will be clustered and consulted for further understanding the results of the statistical analysis of quantitative data. Where applicable, comparisons over time will be considered.

9.8 Criteria for success

The objective O6 of reducing speed by 10% will be considered successful if drivers who exceed the posted speed with more than one standard deviation when entering the motorway exit reduce their speed by 10% in response to the nudge. In a first baseline measurement with a radar system in January 2018, the mean measured speed at the beginning of the exit was at M = 82 km/h with a standard deviation of *SD* = 12 km/h. The drivers over this threshold are the ones that will be targeted by a speed reducing nudging intervention.

For objective O7, the aim is to nudge drivers along a preferred trajectory in a certain section of the road will be considered successful if drivers who deviate from that trajectory with more than one standard deviation reduce their deviation with 40%, compared to a baseline.

10 Study plan for O8: Cyclists' speed reduction (SAFER/ Chalmers University)

Two different types of nudges will be tested for cyclists. In Sweden a nudge that is intended to affect speed will be tested, while a nudge that is intended to affect trajectory will be tested in the Netherlands (see D3.1 and D3.3 for details on the nudges chosen). The test set-up will be very similar with on-site measurement of speed and trajectory through video analyses combined with short on-site interviews, although quantitative measurements with instrumented bikes only will be performed in Sweden.

10.1Trial design and duration

Previous field tests have shown that the selected nudges do have a significant effect on cyclist speed and trajectory. However, these tests have only captured the cyclists' initial reaction to it. To assert functionality over time, the nudges must be up for an extended period of time. A measurement period of 3 months will allow commuters to be exposed to the nudges long enough to reach a steady state. Data will be collected at select dates through the test period to track change over time. The plan is to have a baseline period of approximately three weeks at the start of the trial and then 10 weeks of treatment, ending with a single week of baseline. Data will be recorded during the whole time period and dates will be selected out of that period for analyses in order to get comparable weather and traffic situations.

10.2 Population size

For speed, the previous field test found clear statistical significance from 93 recruited participants in a between-subject design. As the nudge will be applied in the field and regular cyclists will be measured, the amount of actual cyclists measured will be far, far higher. A baseline period of 3-5 days would give a strong foundation of potentially several thousand cyclists. The amount of cyclists encountering the nudge will be so high that it will most likely be possible to calculate significant differences between baseline and measure for each day.

10.3 Participant recruitment

Cyclists passing the nudge on their actual routes will be included in the on-location objective measurements. The large amount of passing cyclists each day will lead to an un-biased sample.

For the individual on-bike objective measurements, cyclists will be recruited out of the people passing the nudge. Although there might be a risk that only bike infrastructure interested people will agree to participate, care will be taken to get as good spread as possible regarding gender and bicycle type (factors that seems to have the biggest impact on riding style based on previous studies).

For the on-location subjective measurements, a booth will be put up proclaiming "free coffee and buns for cyclists" and questions will be asked. It can be assumed that different types of cyclists are similarly prone to stop for free food and drink. Moreover, this test is more or less a replica of what has already been done in a controlled group.

10.4 Data collection parameters

Speed-trajectory curves will be collected on-location for each cyclist approaching the nudge. Videos will be recorded, and events such as approaching cars, close following or overtaking should be noted as well.

In addition to the on-site measurements, 20 participants who pass the nudge will be recruited to cycle their full commute with their own bike instrumented with a Garmin Virb Ultra 30 recording video as well as speed and location at 10 Hz.

10.5 Interview guides

The cyclists will be asked the following questions (translated to respective language and tested and slightly tweaked to assure that they are interpreted correctly).

1. Have you encountered anything special during your ride today? If so, what?
2. (if not, show image of nudge) Do you recognise this? (from today!)
3. On a scale from 1-5 how positive/negative (rotate to balance) do you find it?
4. Why do you think this thing is there?
5. Do you think it affected you somehow?

10.6 Required site properties

The speed nudge should be implemented before a non-light controlled intersection between bikes and cars. The intersection should not be curved, the ground smooth, and the line-of-sight should ideally be obstructed. Accidents or incidents should have happened in the intersection to justify the nudge. It should be possible to continue the nudge up until the intersection, and no other marks should be present on the ground. Three such locations have been identified.

For the trajectory affecting nudge, it will be implemented at a T-intersection between two bicycle tracks where the flow is fairly high and accidents have happened. A number of such intersections have been identified.

10.7 Data analysis plan

The on-location measurements will be taken up by video camera and analysed by Viscando. Speed-trajectory curves are to be exported for each cyclist. If the cyclist decreases speed, the highest speed before the nudge will be noted, as will the lowest speed. The location for the lowest speed will be noted. In case the speed is increasing, the initial registered speed and the speed right before the nudge will be noted. This will be noted for each cyclist. Averages will be calculated per day, per week, over the entire period, and when/if the effect has been stabilised. Statistic comparisons will be made between the baseline period and the various periods defined above.

$V_{reduction}=(V_{highest}-V_{lowest})/V_{highest}$ if $V_{highest}$ occurs before V_{lowest}

$V_{reduction} = (V_{initial}-V_{final})/V_{initial}$ if V_{lowest} occurs before $V_{highest}$

Each cyclist reducing their speed (1% or more) will be marked, and the share of cyclists reducing the speed will be noted.

$Share_{decreasing}=(N_{decreasing\_speed\_1\%\_or\_more})/N_{total}$

The time when each cyclist is 20 meters away from the intersection and the time when they reach the intersection will be noted. From this, the average time to reach the intersection will be calculated. This measures the period when the cyclist can be seen and see others, and shows the combined effect on speed over the nudging period.

$T_{reach\_intersection}=Timestamp_{at\_intersection} - Timestamp_{20\_m\_before\_intersection}$

The distance from the lowest speed to intersection will be noted for each cyclist, to see where the lowest speed is reached.

$D_{lowest\_speed}=X_{intersection} - X_{lowest\_speed}$

From the subjective measurements, the share of cyclists spontaneously mentioning the nudge when asked if they had seen anything special.

$Share_{mentioning}=(N_{mentioning\_nudge\_spontaneously})/(N_{total\_cyclists\_interviewed})$

The share of cyclists either spontaneously mentioning the nudge or recognising the nudge from a photo (as something they encountered today)

$Share_{noticing}=(N_{mentioning\_nudge\_spontaneously}+N_{recognising\_nudge\_from\_photo})/(N_{total\_cyclists\_interviewed})$

This measurement captures if the nudge caught any of their conscious attention.

The share of cyclists stating a correct reason for the nudge to be put up (i.e. decrease cyclist speed before intersections).

$Share_{understanding}=(N_{providing\_correct\_explanation\_for\_reason})/(N_{total\_cyclists\_interviewed})$

This serves to measure how apparent the meaning of the nudge is.

The share of cyclists stating that they think they were affected by the nudge.

$Share_{believe\_affected}=(N_{believing\_they\_were\_affected\_personally\_by\_nudge})/(N_{total\_cyclists\_interviewed})$

Their rating can however not be compared to their actual individual behaviour, only on an overall level. Then it measures how many cyclists are affected more than baseline, compared to how many think they are affected.

Each cyclist gives the nudge a number on how positive it is from 1-5 or how negative it is from 1-5 (to counteract questioning bias). This is converted to the same positiveness score and an average is taken for all cyclists. This serves to find how much they appreciate the nudge.

$$Rating_{positive} = average(rating)$$

The objective individual measuring provides info on how each cyclist reacts to the same nudge over time. Charts will be drawn for with speed reduction in percentages before the nudge at the y-axis and time at the x-axis (in terms of how many times they have encountered it before). The curves will be described and analysed. If the effect stabilises over time, this will be taken as the final decrease, otherwise the average effect from the last week/two weeks will be used and compared with the decrease the initial time.

$$Decrease_{over\_time} = average(Decrease_{final} / Decrease_{initial})$$

10.8 Criteria for success

$V_{reduction}$ will be found successful if the average speed reduction is 7.5% more than baseline speed reduction. The studies on haptic nudges found that speed bumps can reduce speed 5%, and above this is a total success, as it is much better than anything in use today.

$Share_{decreasing}$ is successful if 20% more cyclists reduce their speed than in baseline scenarios.

$T_{reach\_intersection}$ is successful if the time to reach intersection is 7.5% larger than in a baseline condition.

D_{lowest_speed} is successful if the location is 10% closer to the intersection than in a baseline scenario.

$Share_{mentioning}$ is successful if less than 40% spontaneously mention the nudge.

$Share_{noticing}$ is successful if less than 60% have noticed the nudge.

$Share_{understanding}$ is successful if less than 50% understand the purpose of the nudge.

$Share_{believe\_affected}$ is successful if the share thinking they were affected is less than the share actually decreasing their speed; $Share_{decreasing}$.

$Rating_{positive}$ is successful if the average rating on positiveness is 3 or more on a scale from 1-5.

$Decrease_{over\_time}$ is successful if the final effect is 50% or more of the initial effect.

References

Al-Masaeid, H., Al-Mashakbeh, A. A., & Qudah, A. M. (1999). Economic costs of traffic accidents in Jordan. *Accident Analysis and Prevention*, 31, 347-357.

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

FMCSA (2008). *Current FMCSA Crash Cost Figures*. Federal Motor Carrier Safety Administration

Farah, B. (2017). *The Economic Cost and Policy Implications of Heavy Goods Vehicles Road Traffic Accidents in Nigeria*. Doctoral thesis, University of Huddersfield.

Jones, R. J., Woods, S.A., & Guillaume, Y. R. (2016). The effectiviness of workplace coaching: A meta-analysis of learning and performance outcomes of coaching. *Journal of Occupational and Organizational Psychology*, *89*, 249-277.

Stanton, N. A., Walker, G. H., Young, M. S., Kazi, T., & Salmon, P. M. (2007). Changing drivers' minds: the evaluation of an advanced driver coaching system. *Ergonomics*, *50*, 1209-1234.

af Wåhlberg, A. E. (2006). Speed choice versus celeration behavior as traffic accident predictor. *Journal of Safety Research*, 37, 43-51. https://doi.org/10.1016/j.jsr.2005.10.017

af Wåhlberg, A. E. (2007). Aggregation of driver celeration behavior data: Effects on stability and accident prediction. *Safety Science*, *45*, 487-500. DOI 10.1016/j.ssci.2006.07.008

af Wåhlberg, A. E. (2008). Driver celeration behaviour and accidents – an analysis. *Theoretical Issues in Ergonomics Science, 9*, 383-403. DOI 10.1080/14639220701596722

Zaloshnja, E., & Miller, T. (2004). Costs of truck-involved crashes in the United States. *Accident Analysis and Prevention, 36*, 801-808.