SAFER

VEHICLE AND TRAFFIC SAFETY CENTRE AT CHALMERS

Sweden-Michigan Naturalistic Field Operational Test (SeMiFOT) Phase 1

WP 5

Evaluation of Methodology

- Final Report

SAFER REPORT 2010:06
PROJECT C3 SeMiFOT

SAFER Vehicle and Traffic Safety Centre at Chalmers
PO Box 8077
SE-402 78 Gothenburg, Sweden
Phone: +46 31 764 70 00
Fax: +46 31 764 71 88
Web: www.chalmers.se/safer
E-mail: safer@chalmers.se

WP5

Evaluation of Methodology - Final Report

Authors:
Trent Victor, VTEC; Fredrik Moeschlin, VTEC; Magnus Hjälmdahl, VTI; Christer Ahlström, VTI; Carina Fors, VTI; Erik Steinmetz, SP; Martin Lönn, SAAB; Annie Rydström, Chalmers; Jorge Leon Cano, VCC; Christian Blåberg, Chalmers; Tim Gordon, UMTRI; David LeBlanc, UMTRI; Krithika Srinivasan, UMTRI; Adam Blankespoor, UMTRI; MariAnne Karlsson, Chalmers; Pontus Engelbrektsson, Chalmers; David Sandberg, Chalmers; Emma Tivesten, VCC; Christopher Patten, SRA; Jonas Bärgman, Chalmers; Peter Axén, Volvia; Francios Dion, UMTRI; Ralph Robinson, UMTRI; Melinda Morang, UMTRI;

Duration of WP	Jan 2010 to Dec, 2010	Report status	Review draft
WP Leader	Trent Victor, VTEC	Distribution	External
WP participants	AB Volvo, Autoliv, Länsförsäkringar, SAAB, Lindholmen Science Park, Scania, SP, SRA, Test Site Sweden, UMTRI, VTI, VCC, Volvia	Reviewed	By Hans-Erik Petterson
		Reviewed	By Karsten Heinig, VTEC
		Approved	By
		Signature	

1 Table of Contents

2 Introduction

2.1 Objectives and Organization

This report describes the outcome of the Work Package 5 "Evaluation of Methodology" (WP5) within the project SeMiFOT. The overall aim of this workpackage is to **evaluate the Naturalistic FOT research methodology developed in the SeMiFOT**, as such the evaluation is of the methodology itself rather than an evaluation of systems or functions. A *Naturalistic Field Operational Test* is defined as a study undertaken using unobtrusive observation in a natural setting, typically to evaluate the relationship between (permanent or temporary) driver-, vehicle-, or environment factors with crash risk, driving behavior, and countermeasure effectiveness.

The work package was divided into 16 task objectives, each with a corresponding Task Report. 12 of these tasks were clustered under three headings, and an additional 4 tasks were presented independently. Full task reports for each of the 16 tasks are included as appendixes. The 16 specific task objectives within this WP (in **bold**) are as follows:

I. **Data Preparation** – This objective was added during the project. This objective was divided into four tasks:
   1. **Analysis Calculation Guidelines**
   2. **Data Quality**
   3. **Driver ID**
   4. **Map Data**

II. **Analysis of the Naturalistic FOT method as a Tool for Accident Research** – This objective was divided into three tasks:
   5. **Crash-Relevant Events Analysis of Accident Causation**
   6. **Visual Behavior Analysis**
   7. **Automatic Speed Camera Analysis**

III. **Evaluation of selected functions regarding safety, usage and subjective data** – This objective was divided into five tasks:
   8. **Events-Prevented Analysis of Lane Departure Warning**
   9. **Crash-Relevant Event-based Safety Impact Assessment (CRESIA) Method**
   10. **Visual Behavior Analysis of Adaptive Cruise Control**
   11. **Usage Analysis of Lane Departure Warning**
   12. **Subjective Data Analysis**

13. **The opportunities for FOT/Naturalistic data contribution of safety knowledge in the product and infrastructure development process –** This objective encompasses more than the previous name for this objective – "Countermeasure innovation and development".
14. **Analysis of consumer- and research-driven data collection opportunities for full-scale FOTs**
15. **Benefits of Origin and Destination Information in VII Data Set** – Together with objective 7 below, this was split out of a previous objective and work package called "Analysis of the requirements for testing the impact of Cooperative System- (VII) and intelligent vehicle functions in the next, larger phase"
16. **Analysis of requirements and interest for a next phase - SeMiFOT2** – Together with objective 6 above, this was split out of a previous objective and work package called "Analysis of the requirements for testing the impact of Cooperative System- (VII) and intelligent vehicle functions in the next, larger phase"

Final conclusions regarding the main project goals are made in the conclusions section of SeMiFOT final report. However, to help the reader understand the focus of this WP5, these goals are listed here:

- **Goal 1**: to further develop the Naturalistic FOT method into a powerful tool for a) accident research, and b) evaluation of safety, efficiency, and usage & acceptance, and c) countermeasure innovation and development.
- **Goal 2**: to improve the competitiveness of the participating partners and increase the opportunities for economic development.
- **Goal 3**: to verify the Naturalistic FOT methodology at an intermediate scale, at a larger scale than TSS FOT, and a smaller scale than EuroFOT.
- **Goal 4**: to achieve a close cooperation between Swedish and Michigan partners
- **Goal 5**: to bring SAFER, together with UMTRI, into a leading role in Naturalistic Driving and Field Operational Testing methodology.
- **Goal 6**: to determine how to measure the impact of intelligent vehicle systems with a larger study.

2.2 Constraints

In order to understand the choices made in this workpackage, it is important to understand both the constraints which the analysis had to work within. A brief outline of the main challenges is given here:

- **Very limited resources** – A budget of only about 19 person months was spread out over 16 tasks and 13 partners. From the start it was clear that the resources alloted to analyses in WP5 were small in relation to other FOTs that have been performed. Also, the very fact that this was the first time a project of this nature was done in Sweden meant that there was a shortage in experience and competence for analysis.
- **Many hypotheses prioritized** – In WP2.2 "Definition of objectives, hypotheses, and performance indicators for each function" over 32 research questions and over 63 hypotheses were generated and thereafter prioritized using two iterations of a 100 point-per-partner voting system. Subsequent choices in the task reports were guided by accumulated points. Although only a few hypotheses were actually investigated in the end, resources were spread thin across many topics in a "smörgåsbord" approach – having a "taste" of many analysis topics. The following main conclusions were the outcome of the prioritization process:
    1. Safety and Attention analyses should be prioritized
    2. LDW, Accidentology, ACC, and FCW should be the prioritized applications of the analysis
    3. Within Safety analyses, prioritize analysis of crash-relevant events (i.e. kinematic- and system triggers)
    4. Within Attention analyses, prioritize analysis of eyetracker data in selected situations
    5. Within Usage analyses, prioritize analysis of usage for the LDW, ACC, and ESC functions
    6. Within Acceptance analyses, prioritize analysis of acceptance questionnaires.
- **Data access constraints**: The dataset was well-protected according to the data sharing regulations set out in WP1. A byproduct of these protections is added difficulty accessing the data. For example, OEM constraints (e.g. closed data) set limits for what could be studied, access to data could only be made through well-protected computers at SAFER, etc.
- **Data processing difficulties**: Analyzing a very large dataset brings with it many very practical difficulties. The ease of implementation of a particular analysis, for example limited by technology or limited by degree of tedious manual labour, influences the ability to carry out certain calculations and analyses. Special tools, such as small details in analysis software have a large impact on the amount of time an analysis will take.

2.3 General Analysis Approach

The fundamental question is what is the *reference* (ground-truth) by which we evaluate "the performance of the naturalistic FOT methodology"? How do we know if it is a powerful tool for accident research, evaluation of systems, and for countermeasure development? This is certainly a multifaceted and difficult question and there was some discussion regarding the best approach to take given the constraints outlined above.

Basically, there are three different approaches (1) the broad, *smörgåsbord* approach involving a “taste” of many types of analyses, (2) the narrow, *specialization* approach focusing resources to the depth of few analyses and (3) the *metrology* approach involving quantification of tracebility, accuracy, precision, systematic bias, and measurement uncertainty. For example, is it better to treat a large number of hypotheses on a shallow level, to dig deep for only a few hypotheses, or to design analyses for metrology? The advantage of the smörgåsbord approach would be to gain experience in analysing very different kinds of data for different types of hypotheses, and, in the best of all worlds, to obtain at least tentative answers to a larger number of questions. The advantage of the specialization approach would be that it would enable the analysis team to build up competence in dealing with more time consuming difficult aspects of key research questions. A metrology approach was desirable in some instances, for example in comparing various measurement tradeoffs or data collection tools (e.g. eyetrackers or data aquisition electronics).

In the end, a somewhat specialized smörgåsbord approach was chosen to give a taste or practice with analyses. As we shall see, this choice is witnessed by the diversity of analysis approaches in the 16 WP5 tasks, and by the focus on only a handful of the 63 WP2 hypotheses.

It is proposed that the *evaluation reference* for the “performance of the naturalistic FOT methodology” is the degree to which the six main project goals are achieved by way of a joint assessment from the 13 project partners upon finalization of the WP5 tasks.

3 Data Preparation

3.1 Objectives for data preparation

Before analysis can be carried out, the data needs to be prepared and specific metrics calculated. The purpose of the data preparation work is to achieve data with accurate quality, provide guidelines and templates for the calculation of new measures and to complement the recorded data with new measures. Specific objectives are:

- Procedures for calculation to allow reuse of scripts, easy interpretation and reliable results.
- Quality assessment of the recorded measures in a structured way.
- Add driver ID to the database.
- Complement the data with road data attributes.

3.2 Analysis Calculation Guidelines

The objective of the analysis calculation group is to provide guidelines, advice and in some cases scripts for the calculations that are required for evaluation of SeMiFOT hypotheses. The calculation group keeps track on data quality and can help other analysis groups in the selection of measures e.g. by providing information of which vehicle speed signal that is the best suited for calculating vehicle acceleration. The calculation group will also analyse the performance of the calculations and evaluate the different calculation methods available (SQL, Matlab, etc). Specifically, the analysis calculation group will outline advantages and disadvantages with various calculation and storage methods for future FOTs.

3.2.1 Calculation Procedure

The intention with the procedure is to provide guidelines and specify the task that needs to be performed to arrive at statistically sound conclusions for each hypothesis. An overview of the data processing chain is provided in Figure 1. Definition of measures and performance indicators can be found e.g. in the FESTA glossary.

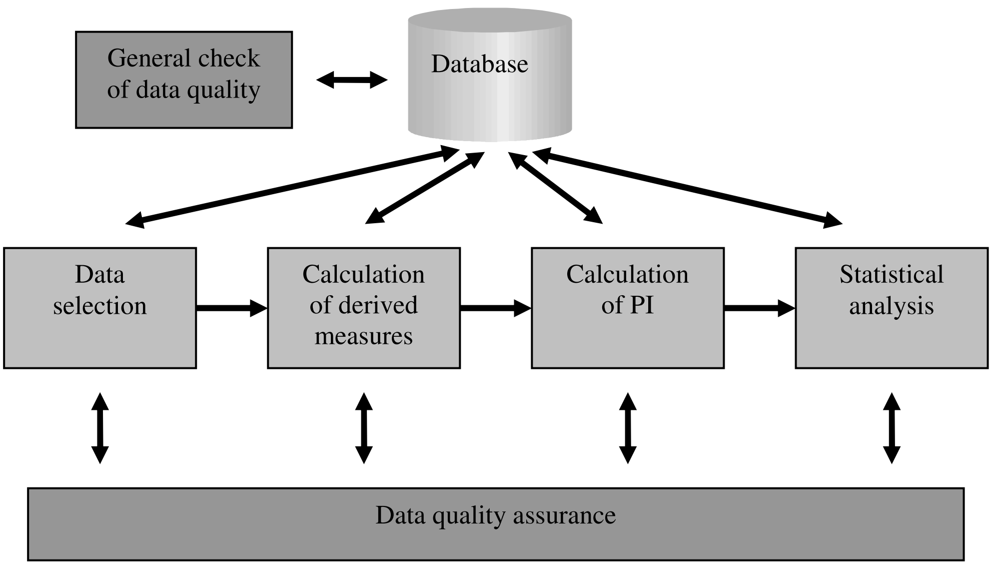


Figure 1 Overview of the data analysis

To carry out the analysis for a specific hypothesis the following calculation process will be followed (the description highlight problems one might run into):

- Data selection
- Calculation of derived measures
- Calculation of performance indicators
- Calculation of events:
- Quality check
- Export data/Statistical analysis

A more detailed schema to follow was developed and used for the vast majority of analysis where hypothesis were being tested.

3.3 Data Quality

This section concerns quality assurance of data in the database. Data quality is a very complex issue to check for. Problems may arise at every step of the data chain ranging from flaws in the study design via data acquisition to the final statistical tests. Data quality is therefore considered as a continuous and iterative process, where each stage has its unique problems. When acquiring a new data set, there are a number of phases that the data flows through. Note that quality assurance have to be maintained after each step in the data flow; after deriving a new measure, after calculating a new performance indicator, after running statistical tests and so on. Concerning the data in the database, a FESTA based quality assurance approach was applied to the database (Figure 2).

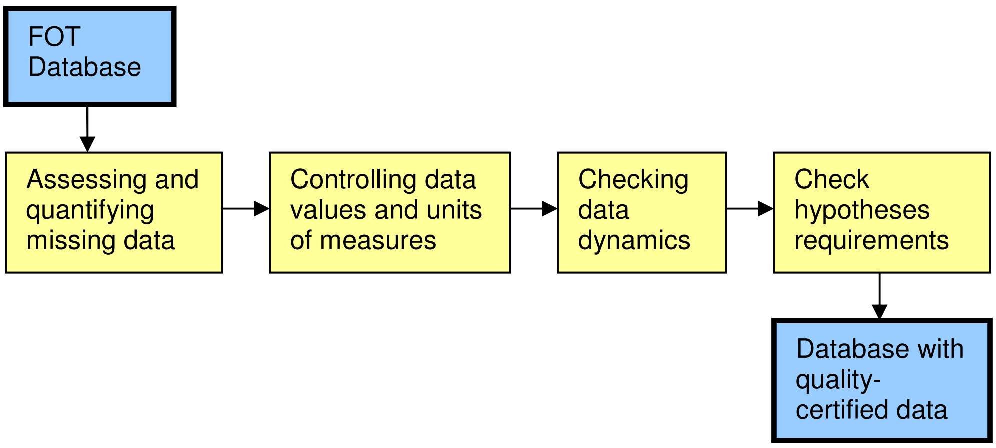


Figure 2: Data quality analysis (FESTA based).

Due to the huge amount of objective data in FOT studies, the data quality assurance process can not rely on visual inspection of raw data. Instead, data visualization must be performed on heavily aggregated data, and preferably, all quality tests should be rule based so that the massive data sets can be assessed automatically. That being said, it is very important to inspect raw data manually from a random sample of the database.

3.3.1 Data quality assessment procedure

To get an overview of how much usable data there is in the SeMiFOT database, a simple test was performed to assess the amounts of trips that were either empty, that only contained a constant value or that only contained zeros (note that such a test does not determine if a sensor is delivering correct values). This estimation of data quality showed that 89% of the collected data was usable (i.e. elligible) for analysis.

It was not possible to assess the quality of all signals in the database because of the large amount of data and the limited time available. Instead, a number of interesting signals were selected for assessment. Three kinds of signals were selected:

- Signals that are frequently used in different kinds of analyses e.g. speed.
- Signals where quality problems are likely to be present e.g. eye tracking data.
- Signals where the quality is difficult to evaluate e.g. signals indicating warnings that are infrequently triggered.

The selected set of measures was: velocity, GPS position, eye tracking, vehicle ahead, direction indicator, ESC, lateral position and steering wheel angle. A quality assessment plan was created for each selected signal. The plan included a description on how to aggregate and visualize data and also a suggestion of tests that could be used in order to find corrupt data. If possible, a quality function was then implemented, based on the findings in the first step.

The quality of the video films was checked manually in parallel with determining driver identity.

3.3.2 Detected quality issues

Many of the quality problems could be corrected in retrospect, but a number of issues inevitably led to unusable data. A selection of the most important quality issues that were found are presented in this section.

There were a number of missing or incorrect mappings in the software that uploaded data from the data acquisition system to the database. This is a severe error since problems in the uploader have a very large impact on the entire database. Fortunately, these problems can be corrected as long as the raw data for the acquisition system is saved. Approximately 150 cases of missing mappings were found. For example, this entailed that all Mobileye data from one OEM was missing and that the distance to the vehicle ahead was not stored in the database. It was also found that some Mobileye data was uploaded to the database without scale coefficients, thus leading to incorrect data in the database. Since mistakes at this stage are so expensive later on, it is essential that this software is checked and double checked by different persons.

Another important quality issue is that documentation about the content in the database is missing or very hard to find. Such metadata is crucial for all analyses. For example, it is necessary to know which numbers represent left and right for the direction indicators and whether positive/negative numbers is left/right for the steering wheel angle. Missing or corrupt metadata slows down the data analysis work substantially and should be avoided. It would also be very convenient if a common data format was used to represent data with regard to units, enumerations, number of signals, signal name in the database etc. The data analysis work becomes unnecessarily complicated and the risk of errors increases when different OEMs represent similar data in completely different ways. An example is the steering wheel angle which have different range for different OEMs and which is sometimes stored in radians and sometimes in degrees.

Proper anti-alias filtering before the sampling of data is an important step in the data acquisition chain, since aliasing effects may be impossible to correct later. In SeMiFOT, acceleration data from the CAN bus was useless due to severe aliasing errors. Fortunately, an external acceleration sensor was used as well and this data was stored with a higher sample rate.

Some trips contain data which have a constant value (often but not always zero) where the range of the constant segment could be parts of a trip, the whole trip or all trips. These issues can have several reasons, but most likely it is a logger problem. Examples include the number of available satellites which is zero for all trips and velocity which sometimes contain non-zero constant segments. A related quality issue is that zero is used as the default value, i.e. the value that is registered when no real data is available. This is very unfortunate, since it will be difficult or impossible to differentiate between missing data and data that actually is zero.

Poor sensor performance that results in noisy and unreliable signals is fortunately quite rare. One sensor that occasionally gives low quality data is the eye tracker. For example, when the driver is looking straight ahead, the eye tracking data sometimes suggest that the head or gaze direction is pointing elsewhere. When the quality indicator, provided by the eye tracker manufacturer, indicates low quality, the raw direction data is quite often ok. Also, at several occasions, head tracking is unavailable even though gaze tracking is ok. This is strange since head tracking should be more robust.

All the video recordings were checked manually. During these checks, it was found that 137 of the video files could not be opened and that 722 face-videos were missing (the total number of trips in the database is approximately 12500). It was also detected that some of the trips were linked to the same video file and subsequent analyses showed that in some cases these trips in the database were identical.

3.4 Driver ID

3.4.1 Introduction

Driver Identification is important for many types of analyses and statistical techniques which build on the possibility to compare individuals. Unfortunately, there was no straightforward sensor or method for assigning a driver an identification. Therefore, a task report “CXXX_DriverID_assigning_v5.doc” was written to describe how the DriverID should be assigned in SeMiFOT. According to the task report the DriverID could be assigned in different manners. This summary describes how the work was actually done and specifies some recommendations for future projects.

3.4.2 Work procedure

According to the task report “CXXX_DriverID_assigning_v5.doc” the DriverID could be defined in three different manners:

1) The OEM supplies all trips (collected drives) sorted by vehicle and driver.
2) The DriverID is defined by using an “operational list” supplied by the OEM.
3) The DriverID is defined by using the analysis tool for video review of each individual trip.

For data collected in Saab vehicles, method 1) was used. Matlab was used to extract minimum 5 pictures per trip for review and assign an ID to the driver. Each individual trip was sorted in a folder structure containing VehicleID and DriverID. All data was then supplied to SAFER on a hard disk drive. At SAFER the data was uploaded to the database with the correct DriverID and VehicleID.

For data collected in VCC, AB Volvo and Scania vehicles, method 3) was used. All trips were uploaded to the database at SAFER with DriverID defined to “0”. This indicates for each individual trip that DriverID has not yet been defined. Each OEM then supplied a “Driver Identification Document” to SAFER which included a picture of each driver connected to project SeMiFOT. All drivers in the “Driver Identification Documents” were assigned a unique DriverID which corresponded to the ID in the consent forms.

Using the video replay analysis tool, called PlayerForm, the video data for each trip in the database was reviewed and compared with the “Driver Identification Documents” in order to determine the driver and assign the correct DriverID. The video data was reviewed at 5 to 10 different time steps, depending on the trip length, to assure the same driver during the whole trip. The DriverID was then assigned by choosing the ID from a drop-down list in the PlayerForm. Issues such as driver change, unknown drivers (not part of the project), missing video views and program crashes were registered in an Excel document, “SeMiFOT – Anteckningar.xls”.

For more information about assigning DriverIDs, please refer to the task report "CXXX_DriverID_assigning_v5.doc".

3.4.3 Recommendation for future projects

If the manual method 3) is used it is recommended to review the data at many different time steps to assure same driver during the complete trip.

The PlayerForm was instable during the DriverID work, i.e. many software crashes. By secure a stable software much time (and frustration) could be saved.

Currently the video views in the trips are stacked on top of each other when retrieved to the analysis tool from the database. Further, the video views are positioned at different places on the screen when new trips are loaded into the PlayerForm. By assuring the video view are placed side by side and at the same position on the screen every time, the DriverID work could potentially be reduced by approximately 1/3.

An alternative to video review would be to automatically generate 10 pictures, at different time steps, from the trip when the data is retrieved to the analysis program. The pictures could be used to determine the driver and the need to review the complete video would be eliminated.

Improvements to the PlayerForm could be to incorporate a stationary list, e.g. check boxes, with the DriverIDs to eliminate the time for scrolling in the drop-down list. An even bigger improvement would be to include the driver pictures from the "Driver Identification Document" into the PlayerForm. When loading a trip the available drivers for that particular vehicle would be shown as pictures with appurtenant DriverIDs. By reviewing the video (or the automatically generated pictures), the DriverID could be defined by choosing the correct picture of the available drivers.

3.5 Map Data

The possibility to use map data for analysis purposes is very attractive as it has the potential to simplify and improve analyses significantly. Some unique possibilities existed to match GPS coordinates with Swedish map databases to extract map data attributes. Therefore a task to add map attributes to the database was undertaken.

3.5.1 Adding map data to the SeMiFOT database

There are three steps in the inclusion of map data based on the SeMiFOT data:

1. Extraction and formatting of GPS coordinates from recorded data.
2. Map data extraction using tool from Triona (a Swedish company).
3. Synchronization and upload in SeMiFOT database.

The following sections describe the steps in more detail. At the end a few bullet points are provided on how to improve similar work in future projects. The reflections can be generalized for any process were part of data is extracted from the database to be completed with more data before uploaded again next to the existing data.

Details of the extracted map data attributes are provided in Appendix "SeMiFOT Map Data Information".

3.5.2 Step 1. Extraction of GPS coordinates from recorded data

The GPS data in the database has been up-sampled from the original 1 Hz to 10 Hz frequency to be able to compare and search with other data (synchronously). Since there is no new information added during the up-sampling, the original recorded coordinates were used for the map data extraction. Using the raw data implies less data to be transferred to Triona's map matching tool and that the map matching process is improved significantly.

All the extracted files were compressed (to about 1/5 of the original size) and encrypted before transfer to Triona over their internal FTP to keep the information secure. In total about 4000 hours of data were kept in a compressed file of about 150 MB.

3.5.3 Step 2. Map data extraction

To be able to analyze the collected data in relation to road characteristics, it is essential to find a method that gives access to the road conditions at every single road segment at the exact time when the vehicle drove there.

The best possible road data in Sweden comes from the National Road Database, NVDB, provided by the Swedish Road Authority, SRA. The company Triona has very extensive knowledge about NVDB, therefore Triona got the task to find out a suitable method, develop necessary software and perform the actual map matching.

NVDB is a model describing road geometry and topology. GPS-points identifying routes for vehicles driving along the roads will never fit exact to the road segments in NVDB. Therefore the most challenging work was to find a method for matching the GPS-points to existing road segments in NVDB.

Further details of how the map data attributes were extracted can be found in an appendix, "SeMiFOT Map Data Matching (Triona)".

3.5.4 Step 3. Synchronization and upload in SeMiFOT database

The map road data attributes extracted by Triona's tool was received in text files containing a number of different parameters as described in Triona's documentation. Among other signals the information includes road number/name, speed limit, road segment, road classification and road width. To synchronize with the 10 Hz signal data available in the SeMiFOT database the map data was up-sampled from 1 Hz to 10 Hz. received from the vehicles with a trip id and a time index.

3.6 Discussion – Data Preparation

The data preparation is an important part in the FOT chain and must be carried out before any analysis to ensure reliable results. During the SeMiFOT project it was found that the data processing and measure calculation is where the majority of analysis time is spent. Thus, anything that can facilitate the work in terms of procedures, templates, structured way of working and similar must be considered. One specific topic that came up during the lessons learned discussion was that the tasks of quality assessment and pre-processing (e.g. calculation of new measures) should be carried out before upload to database if possible. In most cases the required derived measures can be estimated in beforehand and quality can be checked once the data is recorded. By preparing the data with driver ID, map data, quality measures and derived measures as far as possible prior to upload the database will reach a higher level of reliability and usability which naturally will facilitate the analysis.

3.7 Conclusions – Data Preparation

The work carried out in the data preparation phase laid the ground for the actual analysis in the other task reports. In a way the preparation work was indispensable for the analysis by providing driver ID, road attributes, guidelines for operating efficiently on the data and estimations of quality for various measures. However, the foremost conclusion during the preparation work was that an FOT would benefit from preparing the data as soon as it has been collected, i.e. before uploading in the database. Further lessons learned from the data preparation work include the need of better implementation/quality in data logger and the need to further develop the analysis tool used.

4 Analysis of the Naturalistic FOT method as a Tool for Accident Research – Naturalistic-Driving-Study-Style Analyses

4.1 Introduction

Naturalistic driving studies (NDS) represent something as rare as a *new paradigm* in accident- and road safety research. This new paradigm is in the midst of taking form, for example in the SHRP2 study collecting continuous data from thousands of vehicles over a three year period. The naturalistic method "refers to a method of observation that captures driver behavior in a way that does not interfere with the various influences that govern those behaviors." (Boyle et al., 2009). "Naturalistic driving studies are defined as those undertaken using unobtrusive observation or with observation taking place in a natural setting" (Dingus et al., 2006). Typically, drivers are given no special instructions, no experimenter is present, and data collection instrumentation is unobtrusive. They are just let go to lead their lives as normal.

There is now a strong and increasing interest in NDS due to recent, significant advances made in *knowledge*, *methodology* and *technology.* The most intriguing characteristic of NDS is clearly the possibility to record crashes, near-crashes and incidents in great detail, and thereby be able to *objectively understand the true mechanisms of accident causation*. In general, NDS are complementary to existing methodology such as on-site crash investigations, in-depth interviews, and questionnaires.

The central focuses for knowledge advances by NDS are the explanatory factors associated with crashes and the possibility to predict involvement in crashes. The exiting technological advances have to do with the development of data collection instrumentation and data analysis technologies (especially feasability and cost), and are described in the WP3 and WP4 final reports. With regard to methodology, some key advances need to be understood to be able to understand the role that the SeMiFOT work plays in methodology development. These key methodological advances are described below.

One key factor in conducting NDS is *need to collect continuous data over enough kilometres travelled* for there to be enough crashes, near-crashes, and events to be able to draw statistically sound conclusions. The data collection methodology has been heavily influenced by technological advances in data acquisition systems. As technology advances, more data over more km travelled can be collected, resulting in more crashes collected, and this is where the field is developing.

The next key methodological insight is that *it is crucial to have the capability to identify and extract crashes, near-crashes and incidents efficiently*. The landmark 100-car study (Dingus, et al. 2006) was able to examine the detailed causal and contributing factors behind a seemingly unusually large amount of crashes because of the extraction procedure. Everything over and above the 15 police-reported crashes (which is what is to be expected from crash rates per Million Vehicle Miles Travelled) – the 67 non-police-reported crashes, 761 near-crashes, and 8295 incidents – were found because human analysts manually reviewed each video associated with a kinematic event-trigger (e.g. 0.8 g lateral acceleration). Event triggers are thus used to identify potentially interesting data, the video is thereafter reviewed to see if it really was an incident, sort of like instant replay in sporting events.

A third key insight is related to the main result of both the 100-car study and a recent NDS on distaction in commercial vehicle operations (Olson et al, 2009). The main result is the finding that driver inattention is the most prevalent contributory factor in crash causation. Again, it is important to understand that the *method or "sensor" for eye movements in these studies were human analysts data-reducing face videos*. Clearly, this rather time-consuming method would benefit from more automatic methodology.

The final key point is that *Naturalistic Driving Studies and Field Operational Tests are merging methodologically*. To recap, Naturalistic Driving Studies tend to focus on crash-explanatory factors, and Field Operational Tests generally focus on evaluation of systems or functions. They are preferably seen as different methods because (a) the study design is different (participant selection, experimental conditions, vehicle sample etc), and (b) the research questions and hypotheses are different. However, this view is changing due to recent methodological developments, for example:

- Firstly, the recent NDS on distaction in commercial vehicle operations (Olson et al, 2009) was based on a combination of two datasets – the Drowsy Driver Warning System Field Operational Test (Hickman et al., 2006), and the Naturalistic Truck Driving Study (Blanco et al., 2008). That is, a FOT dataset was reused to study crash-causation factors.
- Secondly, the Strategic Highway Research Program (SHRP2), which is commissioning a very large naturalistic driving study (2500 cars over 3 years), is designing the study (vehicle selection) and including research questions for evaluation of the effectiveness of safety countermeasures such as in-vehicle advanced driving assistance systems (ADAS), and road treatments (Boyle et al., 2009). Further, see the selected SHRP2 research questions in the SeMiFOT2 project description below.

In sum, a better, more generic methodological approach is thus to understand that naturalistic data collection can be used to *assess the relationship of static and dynamic Driver-Vehicle-Environment (DVE) factors with crash risk, driving behavior, and countermeasure effectiveness*. In this framework, a study can assess the relationship of an in-vehicle system (dynamic vehicle factor) or age (static driver factor) or distraction (dynamic driver factor) or speed cameras (static environment factor) with crash-risk, driving behavior and/or countermeasure effectiveness.
.

4.2 Crash-Relevant Events Analysis of Accident Causation

A traffic accident is a rare occasion and the likelihood of capturing a crash in a naturalistic driving study of SeMiFOTs proportion is small. However, one valuable way to study the circumstances prior to a "crash" using naturalistic data is to consider crash relevant events (CRE). Crash-relevant-events (CREs) are time segments which contain a particular occurrence that predicts or identifies a crash. For example, CREs can be fatal crashes, non-police-reported crashes, tire strikes, near-crashes, incidents, conflicts, microsleeps, periods of distraction, hard braking, etc. In some cases CREs *identify* crashes (e.g. non-police reported property damage crashes). In other cases CREs *predict* crash outcomes, for example a microsleep of 4 seconds has a certain probability of crash involvement.

One debated issue is though how to define the crash relevant events and how to relate them to real crashes. Hydén (1987) illustrated one view of the relationship between traffic conflicts and accidents in a well-known safety pyramid. Another similar representation is Heinrich's triangle, which for example was applied in the 100-car naturalistic driving study (Dingus et al., 2006). Heinrich's triangle was originally designed for industrial application and the principle behind is that the frequency of occurrence of unsafe acts is related to the occurrence of major accidents. In SeMiFOT, a set of categories and definitions were developed. Figure 3 illustrates a development of the safety pyramid and Heirich's triangle concepts. Here, the crash-relevant event categories are ordered according to *severity categories*, also the point is made that CREs can be related to crash or incident *event types* according to some classification scheme (e.g. the coding scheme developed by Hantula (1994)).

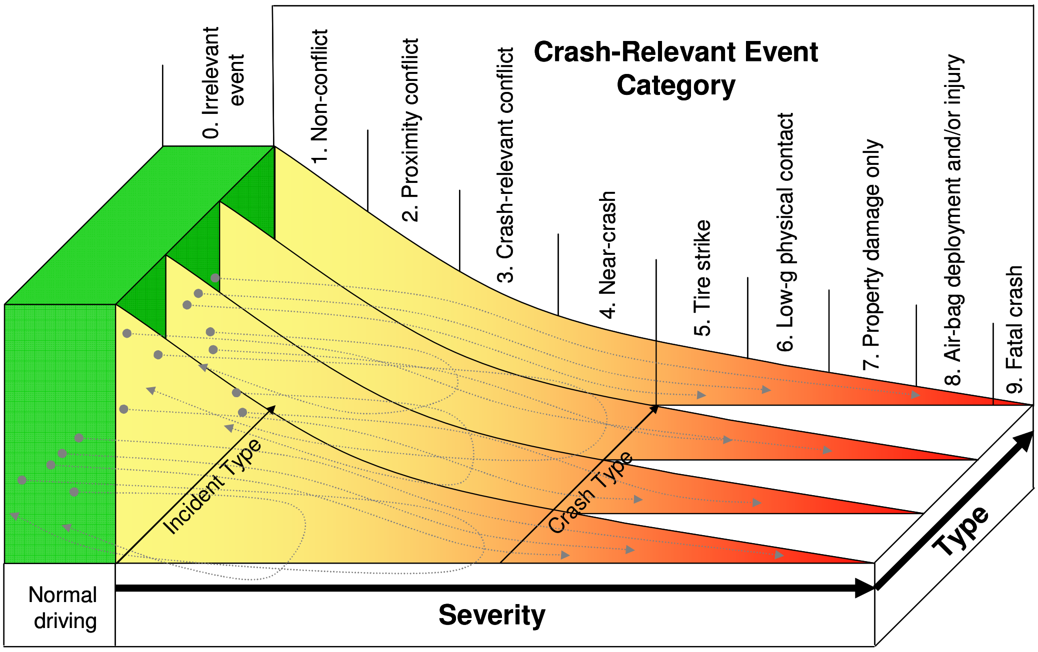


Figure 3. The SeMiFOT crash-relevant event severity categories in relation to event types and normal driving.

The following is the final coding scheme for crash-relevant event severity categories that was used:.

0. **Irrelevant event**: No safety-relevant circumstance is present.
1. **Non-conflict**: Any circumstance that increases the level of risk associated with driving, but does not result in any of the events defined below. Examples include: driver control error without proximal hazards being present; driver judgment error such as unsafe tailgating or excessive speed; or cases in which drivers are distracted to an unsafe level.
2. **Proximity conflict**: Any circumstance resulting in an *extraordinarily close proximity* of the subject vehicle to any other vehicle, pedestrian, cyclist, animal or object where, due to apparent unawareness on the part of the driver, there is no avoidance maneuver or response.
3. **Crash-relevant conflict**: Any circumstance where the subject vehicle performs an *evasive maneuver* to avoid a crash with another vehicle, pedestrian, cyclist, animal or object.
4. **Near-crash**: Any circumstance that requires a *rapid evasive maneuver* by the subject vehicle to avoid a crash with another vehicle, pedestrian, cyclist, animal or object. A rapid evasive manoeuvre is defined as steering, braking, accelerating, or any combination of control inputs that approaches the limits of the vehicle's capabilities.
5. **Tire strike**: A tire strike is when contact occurs with the vehicle's tire only and no damage occurs (e.g. the vehicle is making a right turn at an intersection and runs over the sidewalk/kerb with a tire, or hitting a bump/pothole in the roadway).
6. **Low-g physical contact**: Any low-g contact between the subject vehicle and an object, either moving or fixed. Includes other vehicles, pedestrians, cyclists, animals, roadside barriers or objects on or off the roadway.
7. **Property damage only**: Any contact between the subject vehicle and an object, either moving or fixed, at any speed. Includes other vehicles, pedestrians, cyclists, animals, roadside barriers or objects on or off the roadway. Includes crashes that causes property damage.
8. **Air-bag deployment and/or injury**: Harsh contact between the subject vehicle and an object, either moving or fixed, at any speed. Includes other vehicles, pedestrians, cyclists, animals, roadside barriers or objects on or off the roadway. Deploys airbag and/or injuries someone inside or outside the subject vehicle.
9. **Fatal crash**: Harsh contact between the subject vehicle and an object, either moving or fixed, at any speed. Includes other vehicles, pedestrians, cyclists, animals, roadside barriers or objects on or off the roadway. Fatality inside or outside the subject vehicle.

4.2.1 Triggers providing candidate events

In FOTs event triggers are used to identify potentially interesting data, e.g. crash relevant events. Depending on the data acquisition system used in a study the trigger criteria can play a part in the data collection itself, so that only crash relevant data is saved. When continuous data is collected, as in SeMiFOT, the event triggers can be elaborated and refined after data collection. The triggers used in SeMiFOT were based on the commonly used event triggers: longitudinal acceleration, lateral acceleration and forward time to collision (TTC). At the same time new trigger types were proposed. Specifically it was questioned whether the same trigger threshold should be applied to every driver of if rather "rare events" should be selected as candidate triggers. The events found by an automatic trigger are referred to as *candidate* events.

4.2.2 Verification and classification of events using video

When a threshold for a trigger is overdrawn, a corresponding video clip from a time before to a time after the trigger needs to be manually reviewed to estimate, first and foremost, the severity of the event. To be able to manually code a large number of candidate events in a structured repetitive way there has to be a coding scheme. During this project coding schemes were developed for classifying events as true or false, estimating the severity and categorize the event. The developed methods were based on combined knowledge from schemes in previous studies. During the SeMiFOT project only a limited time could be spent on candidate event classification.

A pilot manual coding was carried out to understand how long time it takes to code an event, provide feedback on the developed coding scheme and to estimate the reliability. Then, two students were given the task to code 935 events representing the worst events for a range of triggers, both driver dependent and not. It turned out that 32 crash-relevant events were found among the selected 935 candidate events (3.4 %). However, when using Kappa statistics the reliability did not even reach a moderate level. Two experienced analysts went through the events again and reduced the number of true events to 19. From the event categorization it could be found that the candidate triggers generated three types of crash-relevant events:

- Conflict with animal or pedestrian on the road/roadside (7)
- Rear-end conflict with vehicle in front (6)
- Vehicle merging from the right (5)

4.2.3 Oops reactions and negative versus positive CREs

The analysts noted that there appears to be a common reaction when something unexpected happens: the driver raises the head a little bit and opens up the eyes wider. Therefore they defined an "oops reaction" as stiffen up, widen eyes, take full control and prepare for action (or brake at once). The analysts also noted that there are positive and negative crash-relevant events. For instance, it is positive when the driver carries out a preventive action because of dangerous circumstances due to other traffic actors. There is a risk of a crash but the preventive action of the driver removed the risk entirely. If no preventive action would have been taken the event would most likely not lead to any crash. However the preventive action is safer than not carrying out a preventive action. Thus a system that generates more such reactions is good. This is somewhat in contraction to expecting systems to generate as few crash-relevant events as possible.

4.2.4 Conclusion of the CRE analysis

The study of critical events is important and can provide understanding to why the drivers' attention may be going away, why everything developed as it did or what situations that are perceived as dangerous. It can then be further investigated how some technology relates to the found behaviour and dynamics. Thus there is a large value in finding and studying crash-relevant events for improving traffic safety.

To ascertain that events are found and classified in a proper way, strict and well thought through definitions are required. The definitions must contain a levelling of proximity to crash and potential severity in the outcome of a crash. To classify well according to the definitions, tutoring of the manual coders are required. The inter-rater reliability need to be studied to attain a sufficient level of training. The results in this study indicate that previous knowledge in driving and accident analysis will improve the understanding of complex situations. Driver interviewing is one further method to better understand the course of events to improve classification.

Automatic triggers have proven to be one efficient way of generating candidate events and they reduce the amount of data that need manual classification. However, there will be events missed and the complexity of a traffic situation may bias the type of events that can be automatically triggered. Therefore it is important to consider random sampling of data to generate further candidate events. By classifying and analysing randomly sampled events any bias in automatic trigger detection can be assessed. During this project the main study did not contain any randomly sampled data and thus no information on potential bias can be provided.

Detection of more events and further analysis of the events found are required to better understand the possibilities with the method of crash-relevant events.

4.2.5 Evaluation of CRE triggers using accident statistics

A separate study was carried out at Volvo Technology on the limited data set available halfway through the project. This study focused on objective evaluation of crash-relevant triggers using accident statistics geographically. A method for evaluating different triggers and their thresholds was proposed and evaluated using more than 11407 reported accidents in Sweden. The method related geographical occurrences of event triggers with the locations of accidents and trends between different trigger conditions and accident frequency was found. Further analysis will be needed, with data including map data, to verify the findings. In any case the method is interesting since it operates on a direct relationship between candidate events and accidents. Thus it may be able to complement the manual coding.

4.3 Visual Behavior Analysis

4.3.1 Introduction

Driver inattention is frequently cited as the most prevalent contributory factor in crash causation. A recent naturalistic truck study with over 3 million miles and over 200 drivers, showed distraction alone was a contributing factor in 71.4% of all crashes, 60% of all safety critical events, and 78% of unintentional lane deviations (Olson et al, 2009). In naturalistic studies, *long off-road glances* greater than two seconds were shown to be especially related to crashes (Olson et al, 2009; Klauer et al, 2006). In particular, *inattention*, or having your eyes-off-road when something unexpected happens seems to be the difference between a crash or incident developing. Secondly, naturalistic studies show that critical events are associated with high eyes-off-road times during the six second period preceding an event onset. That is, these studies show that *accumulated* eyes-off-road time – not just a poor timing of a glance away – is associated with higher crash probability.

Eye movement metrics are considered to be the most sensitive metrics for measuring distraction and workload. The most direct identifier of visual distraction is the visual time-sharing revealed in eye movement metrics. Visual time sharing refers to the stereotypical pattern of glancing back and forth between the road and an object (e.g. a cellphone) during a visual task such as dialing. Visual time sharing induces abrupt steering wheel corrections, large and frequent lane deviations, and slow reaction times to lead vehicle braking. Correlations between eye movement and lane-keeping measures in experiments are typically in the r=.60 to .80 range (e.g. Green, 1999). Similarly, Zhang and Smith (2004b) show correlations between measures of off road glances and brake reaction time of r=.65 in driving simulator experiments. Abrupt steering wheel corrections (2-6 degree corrections) generally occur immediately after the driver looks back to the road (Markkula and Engström, 2006). Here, the term "*oops reaction*" will be used to refer to these abrupt steering wheel corrections, and any other abrupt reaction such as heavy braking, following a glance back at the road. It is possible that a larger degree of steering correction is associated with longer periods of eyes-off-road as can be seen in the later oops reactions (see Markkula and Engström, 2006). These oops reactions could be evidence or predictors of similar loss of control situations involved in crashes and near crashes.

The original analysis plan was to perform correlations and/or multiple regression between attention performance indicators (PIs) on the one hand and lateral- and longitudinal control PIs in selected situations (e.g. visual time sharing, long single glances, high Perclos values, Microsleeps). Some simple analyses were to be performed, wherein the use of selected PIs would depend on how difficult it was to calculate PIs. Due to various difficulties associated with calculation, such as data quality issues and issues with identification of curvature, the following hypotheses were settled upon in the end:

- *H1: Single Long Glances of Head rotation (SLGH)are strongly correlated with measures of oops reactions (MaxReversal, ReversalRate, AbsMaxLatAccel, StdLatAccel, MinLongAccel, and MinTTC)*
- *H2: Measures of visual time sharing (VTSduration and VTSintensity) are strongly correlated with measures of oops reactions (MaxReversal, ReversalRate, AbsMaxLatAccel, StdLatAccel, MinLongAccel, and MinTTC)*

In addition, to perform the analyis of the involvement of inattention in Crash-Relevant Events (CREs), a case study approach was selected, due to resource restrictions.

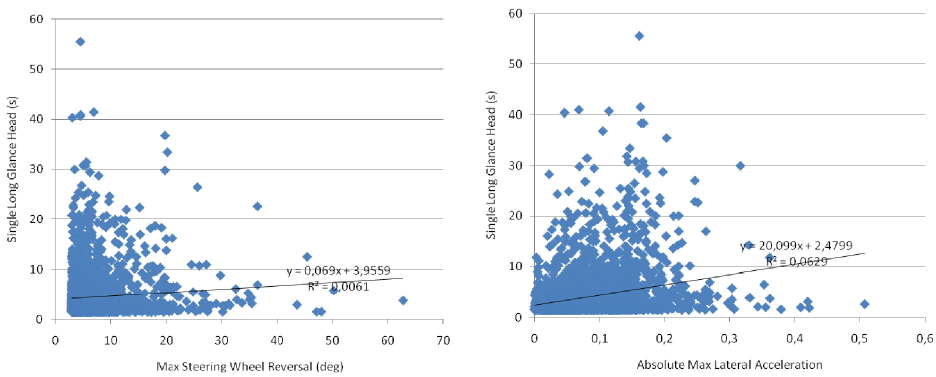
4.3.2 Results and Discussion

In total 3069 Single Long Glance Head events, 9875 Visual Time Sharing events, and 9 Crash-Relevant Events were selected from 47 unique drivers, 13 vehicles, 6308 hrs of data, and 7933 trips from 3 different types of eyetrackers. The analyses included large efforts with regards to data quality and classification algorithm development.

Very low correlations between single long glance head (SLGH) and driving performance PIs were found (Table 1), ranging from r=0.06 to r=0.25 at the highest. Correlations for only the highest driving performance values were also calculated, but these did not inprove the strength of correlation. Extremely low correlations between single Visual Time Sharing (duration and intensity) and driving performance PIs were found (Table 1), ranging from r=0.004 to r=0.12 at the highest. Further attempts to investigate correlations for only the highest driving performance values, for each eyetracker type separately, or for cars and trucks separately did not improve these correlations noticably. Representative plots of the relationships can be viewed in Figure 4.

Table 1. Correlation between visual behavior and driving performance PIs (r).

r =	MaxReversal	ReversalRate	AbsMaxLatAccel	SDLatAccel	MinLongAccel	MinTTC
SLGH	0.08	0.18	0.25	0.18	0.25	0.06
VTS Duration	0.09			0.07	0.12	
VTS Intensity	0.12			0.004	0.05	



a) b)

Figure 4 a-b. The relationship between SLGH and a) maximum steering wheel reversals, and b) absolute max lateral acceleration.

As it was surprising to find such low correlations between SLGH and driving performance indicators, a closer manual validation of both the classification performance of the SLGH algorithm and of the driving performance PI algorithms was undertaken. In sum, 25 out of the 72 validated SLGH events were correctly classified (35%), with DSSR producing 9/24 correct events (38%), AntiSleep producing 2/24 correct events (8%), and SmartEye Pro producing 14/24 correct events (58%). The sample of 72 events out of 3069 SLGH events represents only 2% of the data. Thus, the outcome of the verification cannot be treated as statistically sound or be taken as an indication of which eyetracker performed best because the false classifications were a combination of both eyetracker quality problems and classification algorithm problems.

Overall, extreme SLGH duration is caused primarily by corrupted data, but importantly, also by seating posture changes, and curve negotiation, which can cause extreme duration even though the quality of the head tracking is correct. Inspection of high values of steering wheel reversals seem to indicate that drivers look to the side to negotiate a curve and thus the SLGH segments are wrongly classified as SLGH. Thus, the false classification of SLGH segments in SLGH Duration and driving performance PIs is caused by a two main factors; 1) poor eyetracking quality, and 2) poor classification algorithms for SLGH and Driving Performance. The eyetracking quality filtering is a problem because the false SLGH events were caused by eyetracking data that was classified as high enough quality for analyses. The classification algorithms should be developed to be able to to identify and exclude seating posture changes, head rotations “into” the curves in curve negotiation, etc.

Case studies of nine of the Crash Relevant Events (CREs) from the CRE dataset in WP5.1.1 were also performed. The purposes of these case studies are (a) to investigate the quality of eye tracking recordings during critical events and (b) to see if visual behaviour plays a major role in the series of actions leading up to the event. It can be concluded (from the manual video inspection) that the drivers were seldom looking away from the relevant traffic scene before the actual crash relevant event took place, at least not for a longer period of time. Eyes-off-road seems to be a significant contributing factor in only one or two of the nine crash-relevant events in the case studies. However, more work would be needed to manually code the gaze behavior (on/off road) in a comparable way with previous naturalistic driving studies (Dingus et al, 2006; Olson et al, 2009). As these crash-relevant events are not crashes, but rather incidents, a lower level of involvement from inattention can be expected.

Although large efforts were put into the processing of the eyetracking data by experienced researchers, the eyetracking data included in analyses is in general quite poor. In sum, the data quality and classification issues are so serious that the outcomes of the analyses are very questionable. Therefore, the correlation analyses presented here are likely to underrepresent the true relationship between visual behavior and driving performance or oops reaction metrics. Recall that the systems used in SeMiFOT are very much more close to being mass producable products, that is, the previous eyetrackers that have been used have been more “lab” systems producing higher quality signals. Previous algorithms have been primarily been developed using better signals.

Could it be true that there is a very low correlation between visual behavior and driving performance? The most likely conclusion is that the reliability of the present results is questionable. Next steps should include more verification of the relationship between visual time sharing and single long glances and driving performance deterioration through video analysis, as well as judgement of whether the classification algorithms work properly. This should give a better indication of the frequency of oops reactions, that is, how often a consequence of distraction occurs.

The main conclusion was that the data produced no (or extremely low) correlations between measures of visual behavior and measures of oops reactions. A further conclusion was that the data quality and classification algorithm problems are so serious that the outcomes of the analyses are very questionable. The correlation analyses presented here are likely to underrepresent the true relationship between visual behavior and driving performance or oops reaction metrics. As the eyetracking data is likely the largest dataset in the world. As such, it represents data that could provide true groundbreaking results in suggested further analyses.

4.4 Automatic Speed Camera Analysis

The kangaroo driving analysis presented in this task report aims to investigate whether drivers brake when they catch sight of speed cameras and accelerate after having passed the camera. It further aims to illustrate how naturalistic FOT-data can be used together with data on infrastructure and the problems that may arise.

4.4.1 Limitations of the study

The speed camera data wasn’t available until very late in the project and thus, there was no time for a comprehensive analysis. Unfortunately, a large amount of the speed camera data was found to be of either low or unknown quality and hence, a lot of data was excluded. Yet more data was excluded for other reasons, e.g. if the posted speed limit was not constant in the analysed road segment. The remaining data was not controlled for driver or road and may thus be biased. As a consequence, no statistical analysis was done. The results presented below should merely be seen as examples of kangaroo driving behaviour.

4.4.2 Method and data quality

The data needed for the kangaroo driving analysis is:

- Vehicle speed (CAN)
- GPS position
- Speed camera distance
- Posted speed limit

The quality of vehicle speed and GPS position has been investigated previously (see the task report on Automatic Speed Camera analysis). The quality of these signals is in general good. There are quality scripts for identification of low quality data available and if these quality scripts are used and low quality data is excluded, no considerable quality problems are expected for these signals. With regard to the quality of the map data there were some issues both regard to errors that occurred during the extraction and upload phase but also due to the quality of the data available in the map databases. For more info on this see the task report.

Speed cameras are present in about 9% of all trips (speed camera trips = 1144, total number of trips = 12598)

4.4.2.1 Analysis procedure

Passenger cars and trucks were analysed separately. Two speed profiles were derived for each vehicle type based on posted speed limit: 70 km/h and 90 km/h.

The speed profiles were calculated on road sections of 2 km, starting 1 km before the camera and ending 1 km after the camera. The speed signals were resampled to a resolution of 10 m and all speed profiles were then averaged and visualized in a diagram. In order to be included in the analysis, the road section had to fulfil the following criteria:

- The speed camera distance signal must fulfil the criteria stated in the data quality section above.
- The camera road section must not overlap with another camera road section.
- The posted speed limit must be the same during the whole camera road section.
- The posted speed limit must be either 70 or 90 km/h.
- The speed data must be of good quality (whole quality vector = 1).
- The GPS data must be of good quality (whole quality vector = 1).

Driver identity and road identity was not considered in the analysis, i.e. the result may be biased by an overrepresentation of a certain driver or road.

No statistical analysis was done, because of the possible bias effects and because of the uncertainties in the speed camera data.

4.4.3 Results

Table shows the number of camera road sections included in the analysis, i.e. the number of road sections that fulfilled the criteria given above.

Table 2: Number of camera road sections included in the speed profiles

	Cars	Trucks
Camera road sections included, 90 km/h	83	302
Camera road sections included, 70 km/h	27	11

The mean speed profile for cars on speed camera sections with a posted speed limit of 90 km/h is shown in Figure . There is a clear decrease in speed 50-100 m before the camera. It should be noted that the GPS data is delayed compared to CAN speed with approximately 0-5 s (depending on speed and other factors), i.e. in reality, the speed is probably the lowest very close to the camera. The mean speed profile for 70 km/h roads looks different,. Unfortunately, a lot of this data seems to come from the same road segment and it seems that the speed at the start of the measured section is affected by a side road which clearly affects the averaged speed profile.

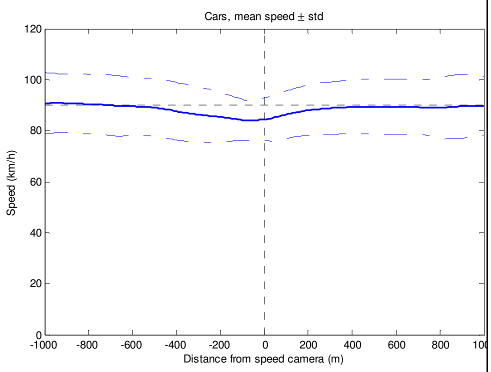
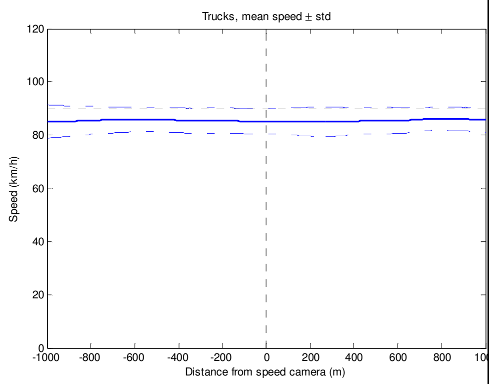
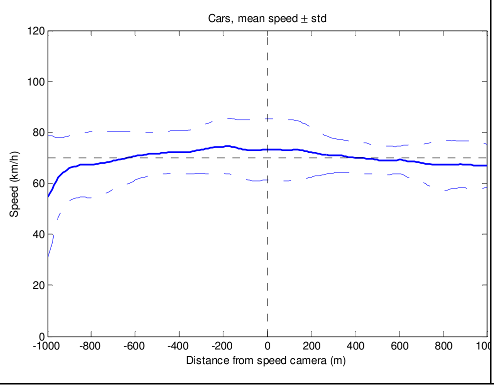
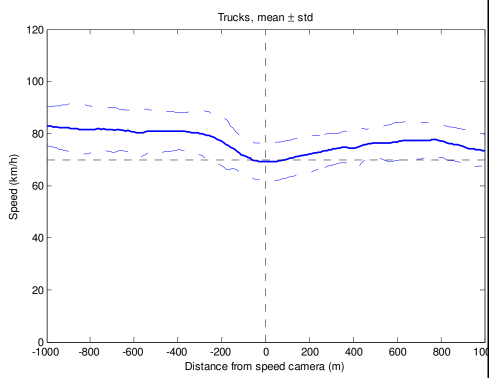
	Cars	Trucks
90 km/h		
70 km/h		

Figure 5 Speed profiles for cars and trucks.

The speed profile for trucks on roads with a posted speed limit of 90 km/h is given in Figure . Since trucks cannot drive faster than 90 km/h, no kangaroo driving can be expected in this case. On roads with a posted speed limit of 70 km/h, there is a clear decrease in speed close to the camera. It should, however, be noted that this profile is based on only 11 camera road sections.

4.4.4 Discussion – Automatic Speed Camera Analysis

This study has shown that it is possible to connect vehicle data with map data containing amongst others information on infrastructure. In this study driver behaviour at automatic speed cameras was shown as a case study but it would have been possible to study behaviour at intersections, roundabouts, specific road types etc as well. For contents of the map data see the automatic speed camera task report. However, the study has also shown that a careful quality check is needed to make sure that data is correct. It has also revealed some problems that have arisen when matching the GPS-coordinates with the vehicle data, some may be solved by a different approach when repeated in another study while some are more difficult to handle.Despite the questions regarding quality of data the results show that there is a kangaroo effect but a more detailed analysis is needed to establish the size of the effect.

4.5 Conclusions – Naturalistic-Driving-Style Analyses

The SHRP2 dataset will dramatically change the way accident research will be conducted in the future. Within a few years, naturalistic data from thousands of vehicles collected over a three year period will be *publically* available for a new generation of researchers to conduct analyses. Today, the time-series data from all the crashes and near crashes from the 100-car study is available at http://www.access.vtti.vt.edu/. The data from the Distraction in CVO study (Olson et al., 2009) is also scheduled for public release in the spring of 2010. Shortly, some of the world's largest databases will be those with naturalistic driving data. Clearly, new methods have to be developed to deal with this new paradigm of research.

Due to immense amount of data and the lack of (experimental) control, the challenge of identifying underlying driving safety mechanisms is a *methodological challenge*. Experimental and epidemiological methods need to be developed or morphed into naturalistic analytical methods. The three tasks reported here all represent development of what we believe are key methods for analyzing naturalistic data.

The work on crash-relevant events (CREs) analysis represents our first foray into developing methods for extraction of CREs. The prerequisite for extaction – a coding scheme for CRE event severity and event typology, including definitions – was developed. The experience with manual review of CREs gave feedback to the development of the coding scheme. Although more data (km travelled) is necessary for obtaining enough events for proper analyses, the work on developing extraction methods a central competence in analyses and there is interesting in its own right. Most known triggers were implemented and some new trigger types were implemented. In addition to these more know-how related results, other interesting data-related results include:

- the positive results using an approach applying trigger thresholds individually to each driver
- the evaluation of CRE triggers using accident statistics geographically
- the observation about postive and negative CREs – the point being that some CREs showing preventive actions are encouragable.

More work is needed on extracting CREs from the dataset and is scheduled to be continued in WP1 of the SeMiFOT2 continuation project. Efficient CRE extraction-, and analysis competence is at the core of future analyses, next steps include the optimization of the CRE hit-rate of triggers.

As noted, driver inattention is one of the most important explanatory factors in accident causation but the current state-of-the art is to manually review and reduce eye movements from video. Therefore, there is a strong interest in whether it is possible to automate the analysis by using eyetrackers. The visual behavior analysis workpackage showed that this is no easy task because of difficulties with the data quality of current eyetrackers and the subsequent problems in making classification algorithms work well despite fluctuating data quality. Although it would have been better if it was an easy task, the analysis methodology developed here arguably represent the state-of-the-art. Eyetracking data of similar quality will be available in the EuroFOT and in SHRP2 datasets so a workable solution for analysis must be developed. The SeMiFOT2 project is well-suited to develop such a well-functioning solution involving improved signal-processing and classification algorithms. Even if this solution involves partial manual review it would be a large step forward from the frame-by-frame video reduction that is the alternative, and would be very useful for SHRP2 and EuroFOT analyses.

The automatic speed camera analysis workpackage represented an important example of how naturalistic data can be used together with map data to assess infrastructure-related research questions. Once map data was matched to the GPS data, the work was able to show that there is a kangaroo effect where drivers slow down before and accelerate after speed cameras, however more analysis is needed to establish the effect in statistical analyses. Again, data quality issues were important to grapple. Now that the road and infrastructure attributes are available matched to the driving behavior, it is possible to formulate continuation studies on the relation between the relation between environment factors with safety. It is particularly interesting to consider that the SeMiFOT dataset can be merged with other datasets, as in Olson et al. (2009), for example comparisons of Swedish and US infrastructure treatments.

In sum, the naturalistic analysis approaches presented here have shown that it is no easy task to analyze naturalistic data, thereby agreeing with Boyle et al (2009). More work is needed to be able to efficiently analyze future datasets such as the SHRP2 and EuroFOT datasets. As intended, the analysis experiences in SeMiFOT represent a valuable step in the development of the competence and methodology needed.

5 Evaluation of selected functions regarding safety, usage, and subjective data – FOT-Style Analyses

5.1 Introduction

This section deals with the analysis of the systems and functions included in SeMiFOT regarding safety, and usage & subjective data. It aims to verify the Naturalistic FOT *methodology* at an *intermediate* scale by performing an evaluation of selected functions.The methods range from more traditional questionnaires to novel simulation techniques. Five tasks are included:

- Events-Prevented Analysis
- Crash-Relevant Events Analysis
- Visual Behaviour Analysis
- Usage Analysis
- Subjective Data Analysis

These tasks differ quite a lot between them both in terms of methodologies used but also in terms of how much focus has been on developing and testing methodologies (as in Visual Behavior Analysis and Usage Analysis), using and adapting existing methodologies to FOT's (as in Subjective Data Analysis) or, as in the case of Events-Prevented Analysis and Crash-Relevant Events Analysis where we have started from page one with developing new theoretical frameworks suitable for analysis of FOT data.

5.2 Events-Prevented Analysis of Lane Departure Warning

5.3 Background

The events-prevented methodology is a new technique which involves the use of multiple computer simulations derived from a single incident in the field operational test data. The simulations are used to generate candidate vehicle trajectories that are used in turn to estimate the benefits of an active safety system in reducing crash risk in the incident analyzed. This capability is important because, while manual review of an incident may provide the reviewer an approximate sense of whether the system performed a useful function – e.g. when a lane departure warning systems alerts the driver to a drift out of lane and the driver responds to correct the excursion – it is not normally possible to objectively quantify the likely safety benefit provided in any particular case. In the events-prevented approach, multiple simulations from the actual event allow metrics relevant to crash-risk to be computed from two classes of events – those with the safety system (lane departure warning) enabled and those without. The aim is to compare outcomes between the two classes and derive a measure of risk reduction.

The simulation approach does require the use of a validated computational model, and in complex crash-relevant cases involving multiple vehicles this may be beyond current capabilities. However, for simple crash-relevant incidents, for example that involving drift out of lane and subsequent recovery, suitable models have already been developed (Gordon et al 2009b). Simulations allow the analysis to explore generalizations of the particular event, and time-step from an initial state (common to all cases) and subject to stochastic disturbances (e.g. visual attention switching) to provide event outcomes in the form of ensembles of simulated vehicle trajectories.

Overall, the goal of the events prevented analysis is to (a) demonstrate the potential utility of lane departure warning (LDW) using real-world events from the FOT carried out in Sweden, (b) estimate the relative risk reduction for the LDW for a specific pre-crash situation, as a first step toward a more comprehensive safety assessment, and (c) gain insight into how LDW may interact with primary crash mechanisms to influence safety benefits.

The method has been developed and applied to a small sample of data from the SeMiFOT data gathered on Volvo Car Corporation (VCC) vehicles equipped with an LDW system. While the scope of this task activity is not mature enough to provide as an objective evaluation of this particular LDW system, and results are largely illustrative of the method, the authors believe that the novelty and predictive power of the technique make the development of considerable value.

5.4 Methods

The approach starts with a manual review of LDW events, intended to screen out irrelevant cases such as deliberate curve cutting or un-signalled lane-changes, and find cases where, in an informal sense, the LDW system appears to have alerted the driver to an unintended drift out of lane, after which the driver responds appropriately. For any such extracted event the simulation model is calibrated based upon a baseline driving episode (typically with the same driver during an earlier part of the same trip) where no lane excursions occurred. From this baseline, key parameters in the driver model are estimated; these include visual preview time, steering control gain and visual attention switching frequency. The driver model is coupled to a model of the road – defining especially the lane-width and horizontal curvature, a model of the LDW system and also a vehicle model. The drift event obtained from FOT data provides a number of initial conditions for the vehicle and driver, especially the visual attention state (typically eyes off the road), the vehicle speed and position in lane, and also the "drift speed", i.e. the component of vehicle velocity perpendicular to the lane centre.

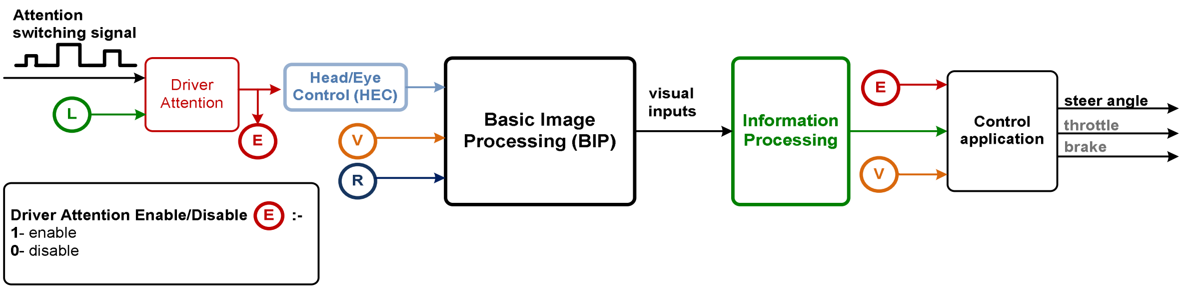


Figure 6: Outline of driver model

The driver model, see Figure 6, is described in more detail in Gordon et al (2009b). Event simulations are started at a point shortly before the instant when an alert driver would have been expected to have applied a steering correction to avoid the drift from lane. This is based on the comparison between vehicle yaw rate and the critical values derived from vehicle motion relative to lane boundaries (Gordon et al, 2009a). The model has a number of stochastic inputs, notably the attention switching and a random yaw disturbance associated with road roughness and aerodynamics. Because of these stochastic inputs, simulated by pseudo-random numbers (PRN) and randomized PRN initial seeds, successive simulations give differing vehicle trajectories.

According to whether the LDW system is enabled, two classes of trajectory are derived; outcomes are evaluated via a simple crash-proportional metric and comparisons are made between the two classes. We note that the functional form of the crash metric can be very simple since the mapping between initial conflict (or pre-conflict) condition is mapped to outcome via the simulation; a "risk-predictive" conflict metric such as time to lane crossing is superfluous to the analysis. In previous work (Gordon et al 2009b) the longitudinal distance travelled with wheels intruded at least 30cm into the adjacent lane was used as the candidate metric assumed to be proportional to crash risk. We do not formally implement this metric here as there remains at present too much uncertainty in two important areas (see below).

5.5 Results

Starting from initial conditions derived from FOT data, the follow sets of trajectories were derived. In Figure 7 the left plot shows an ensemble of 25 simulations with LDW enabled, while the right plot shows the results with the same initial conditions and LDW disabled. The plots show the trajectories of front axle centre, and a lane excursion occurs when this point comes within 0.9 m of the lane boundary, indicating that one of the wheels has left the lane. The increased dispersion (and hence frequency of lane excursions) is expected to provide a direct measure of increased crash risk, and a number of metrics may be used to quantify this. In this initial example there is actually very little intrusion seen in the LDW-enabled case, and by most metrics that equate crash risk to a substantial intrusion into the adjacent lane (e.g. that used by Gordon et al 2009b) the risk reduction associated with the LDW become very high (around 90% reduction using the parameters of the reference). However this figure is very sensitive to two effects that have not been calibrated in this early work: (a) the degree of dispersion expected in the initial conditions, and (b) the stochastic timing of the driver's visual attention switching.

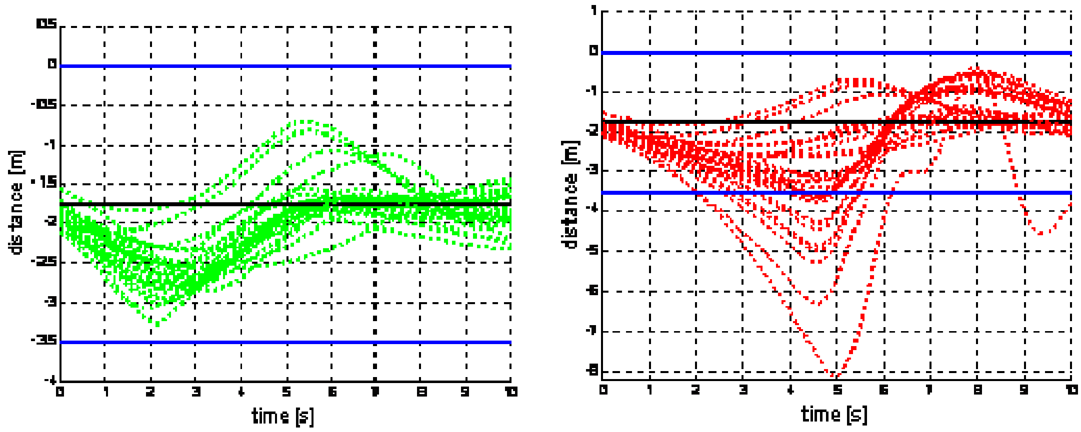


Figure 7: Simulated event trajectories with LDW enabled (left plot) and with LDW disabled (right)

5.6 Conclusions

The "events prevented" analysis was found to be broadly feasible, though labour-intensive, especially in the areas of event review and extraction, as well as parameter tuning using baseline lane-keeping. Once the event had been selected and the model calibrated, the actual event simulation was relatively straightforward to carry out, as was the crash metric estimation. The analysis presented shows significant benefit of the LDW system in terms of the dispersion of simulated trajectories. Current uncertainty over two key processes does however limit the objective validity of the simulations performed and presented above. This motivates conducting further analysis of FOT data to establish a parametric model for visual attention switching from baseline driving data (i.e. outside of crash-relevant events), and also a more formalized set of equally probable initial conditions.

The *events prevented* analysis is important because it is ultimately capable of doing what no human reviewer can do – take an event of interest and answer the question: "what was the system safety benefit in the event we just saw?". In fact the method can be applied equally well in reverse – taking an event with the LDW system inactive and answer the question "what would have been the safety benefit of an LDW system in that case?". The key point is that melding modelling and simulation to actual crash-relevant events provides a systematic and objective way of evaluation that can be fed back into making system design improvements, defining critical driver errors that can be highlighted in education programs etc. etc. And clearly the general approach is not restricted to crash-relevant events associated with lane departures; in the future, similar analyses can be applied to forward crashes, and potentially even more complex scenarios such as lane changes and merging crashes.

5.7 Crash-Relevant Event-based Safety Impact Assessment (CRESIA) Method

5.7.1 Introduction

Naturalistic driving data offers the new and unique possibility to capture crash-relevant-events (CREs) for analysis purposes. However, the methodology for using CREs for evaluations of safety systems must be developed to accommodate this new possibility. Here, it is argued that the evaluation of the safety impact of advanced driving assistace systems (ADAS) and other safety systems is best done by use of crash-relevant events. A novel CRE-based safety impact assessment method (CRESIA) is presented.

5.7.1.1 Objectives

In WP2 prioritzations of hypotheses, which were a summary of the prioritizations of all SeMiFOT partners, a specific research focus within Safety analyses was for analyses of hypotheses related to crash-relevant events. For each system, one typically encounters research questions and hypotheses like: *"R1_LDW: Will incident and crash risk be reduced with LDW?"* , and *"H1_LDW: With the system there will be fewer lane departure incidents, near-crashes, and crashes."*. These types of research questions and hypotheses are arguably the most important ones in safety impact assessment because they capture the main intended effects of safety systems – to reduce the number of unintended incidents and accidents.

5.7.2 An Overview of the Method

This section aims to outline a crash-relevant-event-centered analysis method for estimating the safety impact of a system or function. A number of guiding principles for the analysis approach are presented below. It is expected that further work is needed to define and modify this approach based on detailed analysis of SeMiFOT and other FOT data. This work is expected to occur in the SeMiFOT2 project.

5.7.2.1 Possible safety impacts

An active safety system is likely to effect the driving situation on both macro and micro levels. Macro effects include factors such as individual interest in driving, time of driving, selection of road, speeding tendency and overall sense of safety and the effects thereof. Micro effects are closely related to the operational tasks in driving and include lane keeping behaviour, following distance and similar operations. Active safety systems under evaluation in an FOT are mainly aiming at safety effects on the micro level, typically with warnings for inattentive behaviour and control systems for reducing crash risk. Therefore it is on the micro level where safety impacts of the system are generally to be expected. As an illustration it can be envisioned that a system which reduces the accident risk significantly at dangerous scenarios may at the same time slightly increase the speed in common scenarios, perhaps due to an increased feeling of safety. From a macro perspective, e.g. by looking at the average speed with and without the system, this could be interpreted as if the safety decreases with the system since increased average speed is connected with increased fatality risk. Thus, to assess the safety impact of a system a condition is that the micro level dynamics and events are analysed.

The ideal way of studying safety impact of a system on the micro level is to carry out in-depth analysis of accidents to see how the occurrence of severities and injuries vary with and without a system. Fortunately there are normally not enough accidents occurring during an FOT for statistical significant results. However, conflicts and near-crashes are frequently recorded and can be used as surrogate measures of accidents if they are accurately interpreted. These events are included in the Crash-Relevant Event (CRE) categories. As discussed in the task report on Crash Relevant Events, there is a complex nature in the study of CREs since it must comprise both frequency and severity at the same time. That is, there is one likelihood that an event leads to an accident, and another likelihood that an accident will lead to injury or fatality once a crash occurs. It is important to note that a traffic solution that intends and succeeds in reducing crashes and fatalities may at the same time increase the CREs at another level. For example rebuilding a 4-way intersection into an average sized or large roundabout increases the number of accidents by up to 60%, but reduces the expected total injury outcome by an average of 70% (Englund, 1998). Analyses of safety impact must be able to take into deal with such redistributions in injury risk in addition to accident risk.

5.7.2.2 Overview of the steps in the proposed CRESIA method

The method presented here can be summarized in the following six steps:

5.7.2.3 Step 1. Detection and classification of CREs

As mentioned in the task report Crash-Relevant Events Analysis of Accident Causation (see Section 4.2), candidate CREs can be provided by several methods and manual inspection/classification is required to determine if an event is a true CRE or not. This process of manual video review of CREs can be likened to on-site crash investigation procedures. Typically, a coding scheme is used (see Section 4.2) to determine attributes of the CREs which are particularly useful in downstream safety impact analyses. For example, accident types can be coded according to a chosen typology (e.g. Hantula (1994)). This manual video review can also be used to code other attributes such as the behaviours leading up to a crash occurrence, types of distracting activities, etc. More or less elaborate coding schemes exist (see Section 4.2), however coding schemes are currently immature and can benefit from further developments more targeted to safety impact analyses.

It is particularly important to code and account for the following aspects of CREs:

- **Accident typology coding of possible crash outcomes**. The coding of the likely or possible events that are associated with the CRE can for instance be done by using the INTACT typology (Hantula (1994)) or similar established methods.
- **Crash risk coding: Likelihood estimation of possible crash outcome.** The likelihood estimation of a CRE actually leading to a crash is by far the most difficult part in this methodology. It is very important that the estimation is carried out in robust way with high reliability (independent of a safety system's activation state). For instance, a constant slight overestimation of frequency would likely lead to overestimation of accident risk for both baseline driving and treatment with system on. However, if the reliability is low the variations

in estimation may influence the generated safety impact outcome. A naive estimation of crash risk of a CRE would be to use trained specialists, and thereafter measure the inter-rater reliability. However, the validity of the estimations may be questioned simply because of the use of subjective ratings. To approach an objective method, accident statistics could be considered to provide a background on how many of the accidents with the current typology occur per km (or similar). Further research is suggested to establish a robust method of estimating likelihood of crashes from CREs. A suggestion to how such research can be focused is provided in Section 4.2.5 where the relation between real accidents (in a national crash database) and CREs is compared geographically in Sweden.

- **Injury risk coding: Risk of injury in the event of a crash.** The injury resulting from a crash can be measured according to established scales such as the Abbreviated Injury Scale (AIS, see for instance http://www.trauma.org/archive/scores/ais.html). In order to generate only one probability of injury (rather than one per injury level) it is suggested that a threshold is selected. For instance AIS 6 can be considered to only estimate fatality risk if the AIS scale is used. The injury risk will be a function of several event attributes such as accident type, opponent object, relative impact velocity, seat belt usage, occupant physical conditions and age etc. From accident statistics and previous work within the field of passive safety there are relations available to be able to determine a function for injury risk estimation.
- **Description of circumstances and control factors.** To facilitate work on identifying comparable situations it is suggested that the circumstances of each CRE is described in a uniform way. The attributes described should not be related to the system under evaluation. Items to consider are for instance weather, traffic density, road and environmental categories, vehicle type and load carried.

5.7.2.4 Step 2. Identification of comparable (baseline) scenarios

Due to the naturalistic data collection in an FOT the exposure of various scenarios and conditions may vary between the drivers and over time. To estimate the risk of injury based on a CRE it is therefore important to be able to identify a comparable baseline, i.e. how often scenarios similar to the CRE occur. By using the descriptions and classifications attained during the coding, the CREs can be clustered in according to scenarios. For example, one scenario covering several events may be quantified by two congested lanes merging on a highway during night time in clear weather. Depending on the attributes and factors considered, the selection can be described at an arbitrarily detailed level. A method for comparing and identifying scenarios is outline in a task report appendix.

The baseline can be expressed either in kilometres or hours driven. There may be large variation in the denominator if the detail of the attributes describing the scenarios is changed. Thus, it is important to decide on a sufficient level of detail, and use the same scenario definitions for all treatment conditions (system on, off, disabled etc). It is recommended that all scenarios are included even if the system evaluated is not actively operating in such scenarios. This is the only way to ensure that there are no unexpected or unintended positive or negative effects of a system.

Due to time and resource limitations it may be required to discard several of the scenarios identified. It is suggested that scenarios which include few and less severe CREs are primarily discarded. One further option would be to focus only on scenarios where the system under evaluation is likely to operate. It must however be kept in mind that limiting the scenarios will reduce the conclusions possible of the impact analysis; it will only be possible to make statements for the scenarios selected. Therefore it is recommended that enough scenarios are included to cover the majority of CREs. Ideally the division is carried out using available measures of control factors. Then the division of comparable scenarios can be carried out automatically and no limit on numbers is required.

5.7.2.5 Step 3. Estimation of combined CRE severity

The combined severity of a CRE can be given by the product of crash risk (the likelihood of a CRE leading to a crash) and injury risk (the likelihood of injury outcome in the event of a crash). Depending on the threshold used for severity selection during the CRE classification, the likelihood of injury may range from likelihood of fatality down to likelihood of impact with no injury. Assume that a limit was set using the AIS scale, on the threshold AIS*X*. Then the likelihood of the CRE leading to an AIS*X* injury can be expressed as P(Accident) * P(AIS*X*). By weighting the frequency with the severity in this way, the complexity of CREs can be handled, such as the roundabout example (Englund, 1998).

5.7.2.6 Step 4. Division of data and estimation of treatment categories

Depending on the study design and system configurability, the recorded data can be divided in two or more categories. For instance data from an evaluated LDW system may be able to categorize in the categories:

- System not installed (or disconnected)
- System disabled due to too low speed or similar
- System available but turned off by driver
- System available and operating

It may be of interest to further divide the categories. As an example it could be estimated for the period with system not installed if the driver would have preferred to use the system or not, based on how the system was used when it was available. Such estimations would allow direct comparison between drivers exposed with system turned on and drivers unexposed where they most likely would have used the system. Information from usage analysis can be valuable input in this work.

For each treatment category a matrix can be generated with one column for each scenario selected. An example of such a matrix is shown in table 3 below.

Table 3: Example of evaluation matrix for one treatment category, e.g. system not installed.

	Scenario A	Scenario B	Scenario C
Number of events	54	13	27
Amount of comparable scenarios (km or h)	650000 km 10000 h	92000 km 1500 h	450000 km 8000 h

5.7.2.7 Step 5. Compensation for macro effects on comparable (baseline) scenarios

As previously mentioned, a system may have impact on macro level traffic variables. It is likely that driving interest, selection of route and similar factors will in turn affect the amount of driving in the considered scenarios. Thus, before calculating the risk measures for the various treatment categories and scenarios it may be wise to estimate the macro impacts of the system and use the result to compensate the baseline (amount of comparable scenarios in table 3 above).

5.7.2.8 Step 6. Relative risk and population-attributable risk

The relative risk can be calculated directly from the matrices for the various treatment categories. To understand the population-attributable risk (PAR), see Sahai and Khurshid (1996) for the relevant conditions required for general statistics of road user behaviour. Specifically, estimation of frequency is required for the comparable scenarios selected.

5.7.3 Discussion

The presented method assumes that severe crashes involving injuries and fatalities in general are represented by a large number of CREs that do not lead to severe crashes. This assumption is connected to the shape of Heinrich's triangle or the Safety Pyramid (Hydén, 1987). The interpretation is that slight conflicts are in general more frequent than severe. However, this can not always be certain for various CRE types. Further, it is also assumed that the effects on crash-relevant events from long-term usage – increased experience, and changed micro dynamic behaviour due to penetration rate increase – can be neglected. In any case there is currently no clear theory of how the changes of e.g. penetration rate would affect the micro-level events that are studied. Rather, that is an interesting field of study for upcoming future FOTs.

5.8 Visual Behavior Analysis of Adaptive Cruise Control

5.8.1 Introduction

In marketing material from Honda (presented to the main author), a claim is made that the use of Adaptive Cruise Control (ACC) increases visual scanning and thereby increases safety and comfort. The idea is that the driver doesn't have to fixate the lead vehicle as much because the headway keeping control task is automated. Eyetracking data is shown in the marketing material in the form of density (scatter) plots showing that the distribution of gaze is more spread out when using an ACC in comparison to when the ACC is disengaged.

In the marketing material it was unclear whether the visual scanning patterns to the interior of the vehicle were also increased. That is, it is possible that the drivers were not only looking around more to the exterior, but were also performing more distracting tasks in the interior of the vehicle. Thus, in response to ACC being active, drivers may be more active in their gaze in a positive way – scanning the exterior and traffic environment more – or may be more active in a negative way – performing more distracting tasks. The driver may look around in the traffic environment more or less, and/or the driver may perform more in-vehicle secondary tasks.

5.8.2 Objectives

The aim of the current analysis is to study the change in visual behavior as a response to Adaptive Cruise Control usage, comparing visual behavior with and without ACC active. Two hypotheses were selected for analyses:

- *H4_ACC: Situation awareness (gaze patterns) will be improved when using ACC*
- *H5_ACC: When using ACC, the driver will be more distracted because of more time spent on secondary tasks*

5.8.3 Method

The analysis plan was to provide eye/head-movement "density" plots with and without ACC on, summed across all vehicles and drivers, and to calculate a performance indicator called Standard Deviation of Radial Gaze/head movement (SDRG). SDRG is the standard deviation of the Euclidean distance from the mode (peak) of the cluster of glances towards the forward roadway. However, unfortunately the SDRG measurements proved to be sensitive not only to the spread of the gaze or head movements, but also to the quality of the data. More work is needed to be certain that differences are not being picked up because of variations in data quality (noise) rather than increases in scanning (signal).

In the ACC comparison procedure, density plots of Head and Gaze Angles were created for the five ACC states that can be seen at the right-most section of Figure.8. In the figure it can also be seen that the study is only carried out for VCC cars with an eye tracker system installed.

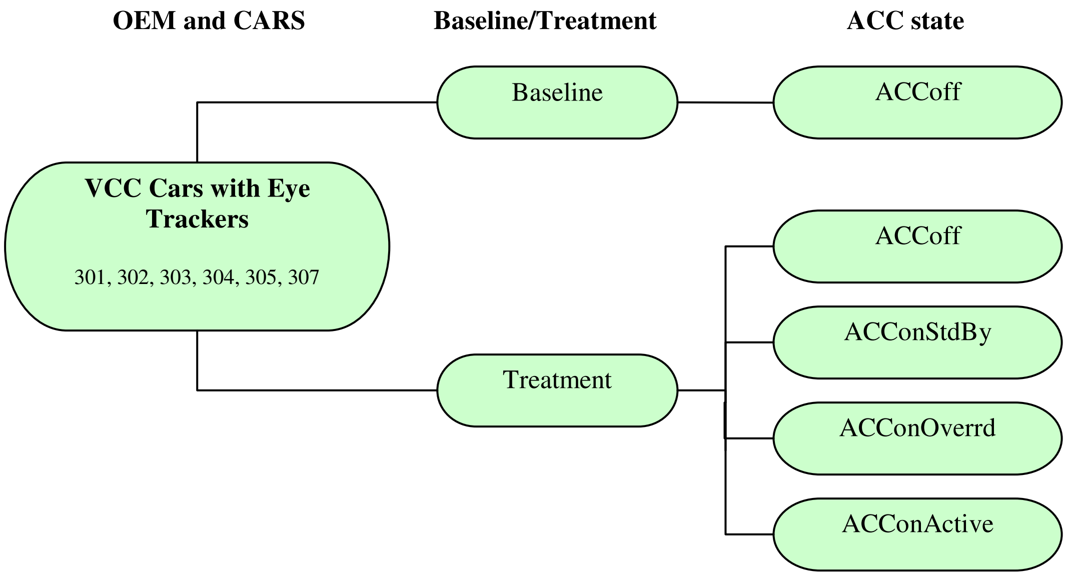


Figure 8: Selection of data for the ACC comparison study

During the baseline period drivers were not able to turn ACC on and thus the *off* state was permanent. During the treatment period when ACC is *on and active* and the driver presses the *accelerator pedal*, the state becomes ACC *on and overridden*.

All the trips from the vehicles listed in Figure 8 are considered for the computation of the densities of head and gaze angles. If a trip has a TRIP_ID in both the main signal-data table (TRIP_XC70_MAIN_10HZ) and in some of the eye tracker tables (TRIP_SMARTEYE_XHZ / TRIP_DSSR_XHZ), the trip is analysed to find segments with different ACC states.

The density distribution for the gaze and head angles are estimated for each trip (in a -40 to 40 degrees 2-D histogram, divided into 128 bins along each axis) and are then aggregated (with data from other trips) into five different matrices, one for each ACC state—that is, 5 matrices for gaze angles and 5 matrices for head angles. Only the eye tracker data samples that pass the quality check below are included in the aggregated results.

The estimated time with eye tracker data from trips that fulfil the selection criteria and that have a trip id in both the main table and the eye tracker table was 182hrs and 27min for ACCoffBaseline, 789hrs and 46min for ACCoffTreatment, 41hrs and 7min for ACConStdby, 52min for ACConOverrd, and 33hrs and 2 min for ACConActive.

5.8.4 Results

Figure 9 shows plots of head rotations in ACC data in inactive states – ACCoffBaseline, ACCoffTreatment, ACConStdBy – in comparison with the active state – ACConActive. As can be observed, there seems to be a visible difference in the concentration of the distribution of the central peak between active states (8a-c) and the inactive state (8d). There is also a visible difference in the lowest blue portions where states with less data show more data bins without gaze data. This is likely simply an effect of less eyetracking data collected.

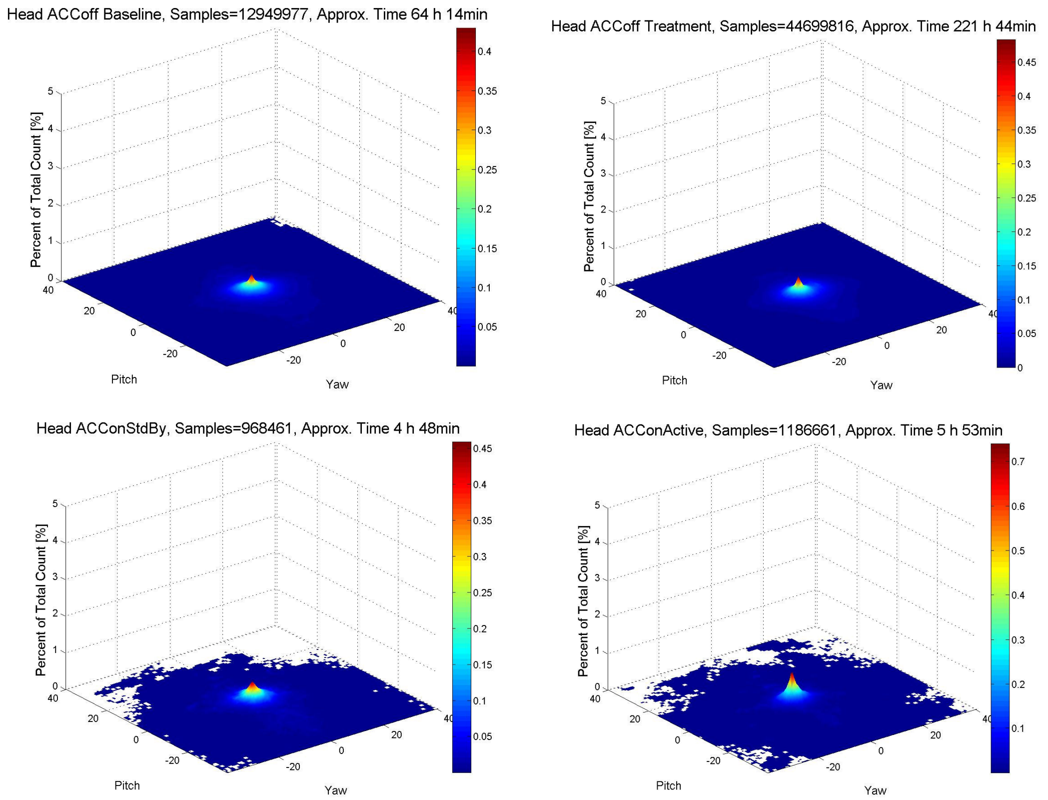


Figure 9. Density plots of ACC data in inactive states (ACCoffBaseline, ACCoffTreatment, ACConStdBy) in comparison with d) the active state (ACConActive).

Plots were generated of gaze ACC data in inactive states – ACCoffBaseline, ACCoffTreatment, ACConStdBy – in comparison with the active state – ACConActive. Visual inpection of these plots showed that there is no visible difference in the distribution of the central peak between active and inactive states. Similar to the head data, the only visible difference is in the lowest blue portions where states with less data show more data bins without gaze data. This is simply an effect of the amount of eyetracking data collected.

5.8.5 Discussion

In summary, the *head rotation* density plots show that there seems to be a *concentration* of the distribution of the central peak when ACC is active (ACConActive), in comparison to the inactive states (ACCoffBaseline, ACCoffTreatment, ACConStdBy). However no statistical tests were performed to show if this effect was significant. The *gaze* plots show no directly visible difference in the distribution of the central peak between active and inactive states.

If the head rotations were shown to be significantly concentrated, then this result would consitute a rejection of the first hypothesis *H4_ACC: Situation awareness (gaze patterns) will be improved when using ACC.* The results here seem to indicate that gaze is less active and more concentrated to the road center when using ACC, but these findings should be statistically tested. Preferably, the intended performance indicator Standard Deviation of Radial Gaze/head movement (SDRG) should be calculated.

As regards the second hypothesis *H5_ACC: When using ACC, the driver will be more distracted because of more time spent on secondary tasks* there seems not to be indications of a markedly increased proportion of gaze of head rotatations towards the interior of the vehicle. More work would have to be done to investigate this in proper statistical tests and other performance indicators such as measures of visual time sharing.

A very small amount of analysis time was alloted to the analysis presented in this report. The analyses presented therefore suffer from the lack of time needed to provide publishable results. Clearly more work is required. Considering the data quality concerns uncovered regarding eyetracker quality in Section 4.3, it is likely that the gaze data is very noisy and may not be able to pick up a difference in the gaze data. Both the head movment and the gaze data should be subjected to validation before performing statistical tests in future analyses. All in all, even though we were not able to reach the analysis stage where statistical tests were able to be performed, it was a very useful exercise to gain experience from various issues associated with extracting the data from the database. Also more work and consideration of issues surrounding selection of comparative datasets for proper statistical testing is needed.

5.9 Usage Analysis of Lane Departure Warning

The aim of the usage analysis was to deliver answers to the hypotheses defined in WP2. In that work package a number of possible hypotheses regarding Usage were proposed. The number of hypothesis assessed during analysis was restricted due to the limited number of hours and the complexity of analysing each hypothesis. The delay of data availability and quality issues resulted in limited actual testing of hypothesis while several calculation procedures were developed.

The analysis calculation procedure was carried out all the way for the hypothesis H7_LDW, “the rate/frequency of system-initiated warning of LDW will vary with different situations”. Specifically it was analysed how the system-initiated LDW warnings were related to the road type. The detailed background and motivation is given in WP2. A comment that well illustrates the hypothesis was given by one of the truck drivers during the final workshop: “as soon as I enter Norway the winding roads lead to so many warnings that I need to turn it off”.

After carrying out the selection of data with relevant control factors, the statistical problem turned out to be complicated to solve due to the absence of warnings on several segments of data. In order to work out a solution the division of Mathematical Statistics at the Department of Mathematical Sciences of Chalmers University of Technology and The University of Gothenburg were consulted. When modelling the relation between the number of warnings and the road characteristics it was found that there is statistical significant evidence that system-initiated LDW warnings are more frequent on winding roads.

The outcome may at a first glance not seem too surprising. In any case, the methods developed and applied during the usage analysis work provide a scientific way of working for hypothesis testing with FOT data. Thus the methods described in this document will provide valuable input to similar hypothesis testing.

5.9.1 Selected null hypothesis

From the outcome of WP2, six specific usage hypotheses were identified. Each of the hypotheses needed to be broken down in specific null hypothesis before they could be statistically evaluated with the SeMiFOT data. After review in the usage analysis group, seven null hypotheses were selected as primary hypothesis to test.

The selection was presented and discussed during one of the overall SeMiFOT WP5 meeting to ensure that they were accepted by all partners. Depending on available resources and data additional null hypothesis were to be considered. As it turned out the delay of map data and limited quality of available signals reduced the number of null hypothesis to one.

In the selection of suggested null hypothesis the background material in the WP2 task reports and the available data in SeMiFOT were considered. A general comment on the analysis is to use GPS speed in the selection since the odometer offset may vary.

5.9.2 Calculation Process for hypothesis

The detailed analysis plan is a break down according to the template from the SeMiFOT analysis calculation group (see Section 3.2.1 on Calculation Procedure). The SeMiFOT hypothesis H7_LDW was selected as example to illustrate how the process will be carried out for usage hypothesis. Definition of measures and performance indicators can be found in the FESTA glossary. A measure can be interpreted as a signal over time.

5.9.3 Curvy or straight road

In order to evaluate if a part of the road is curvy or straight a Matlab script was used. The script uses vehicle speed, time and yaw-rate as input to estimate average curvature of a longer segment of data. The curviness in terms of true or false for each time stamp was then calculated by another Matlab script.
The assumption was that the road is winding unless:
1. The curviness is above a threshold value
2. The curviness has been above the threshold for a certain time.
If the vehicle speed is below normal road speed the value is not calculated as it was decided to difficult to assess these kinds of movements. The different threshold values used was based on trials and observations of the recorded data. The threshold values for curviness varied between different vehicles used in the tests.
If the curviness value is above the threshold value the road immediately becomes curvy. This is different from the transition between curvy to straight roads. The reason for this is to avoid small straight parts of a mainly curvy road to counts as straight.

5.9.4 Data selection

The analysis of the hypothesis was based on the data collected from VCC cars. The data collected from the vehicles and the subjective data from the questionnaires is stored in the SeMiFOT database (WP3). The data is selected only when LDW is active and the speed is above 65km/h. The data is selected by using SQL query and then is imported in Matlab with Matlab script. The whole analysis is then done in Matlab. The trips that fulfil this condition are selected with for each vehicle and driver.
The other information extracted from the database is driver age, driver experience, gender and average number of kilometres that the driver drives during one year as well as the metrics defined during the calculation procedure. The output matrix is then exported to excel document used to do the statistical evaluation of the hypothesis, described in the section *Statistical analyses*.

5.9.5 Statistical analyses

A problem with the data from a statistical analyses point of view was that the actual number of system warnings was quite low. 85% of the road segments did not have any warning at all. (The average road segment length was 6.2km).

5.9.5.1 Model

We have modelled the relation between the number of warnings and the road characteristics with a negative binomial count regression model with logarithmic link function (NB2). Negative binomial regression can be viewed as a generalization of the more basic Poisson count regression that allows for over dispersion. By adding an extra scale parameter the restrictive assumption that the variance equals the mean inherent in the Poisson model can be avoided (Cameron and Trivedi, 1998).
To account for the different interval lengths we used the logarithm of the lengths as an offset. This is the same as including the logarithm of the distance as a covariate but avoiding estimating its coefficient

which is instead fixed at value one. In effect we are actually modelling the rate of warnings per distance travelled.
To make up for the possible dependencies between observations for the same driver we added one Gaussian random effect per driver. That is, we assume that there is an unobserved mean zero normal random variable added to the linear part of the model shared by all observations from the same driver.

5.9.6 Results

Our analysis shows that there is statistically significant evidence of a relation between the type of road (winding/straight) and the number of warnings (p- value 0.0007). In addition to the road type we tried adding weather and light conditions to the model. None of these showed any statistically significant impact on the rate of warnings (when already accounting for the road type). Perhaps is it the road width that is strongly connected to average road curvature and accounts for the increase frequency of warnings?

5.10 Subjective Data Analysis

The subjective data play an important role in a test such as SeMiFOT. Not only is the subjective data necessary to explain the outcome of the objective data collected, but also for testing some hypotheses, for instance regarding an individual’s acceptance of a function.

In SeMiFOT, data on the participating drivers and their acceptance of the functions evaluated were collected by the means of different questionnaires and by means of personal interviews (with a sub-set of the participants.

The questionnaires included:
1. Background Questionnaire, car and truck version;
2. Driver Behaviour Questionnaire (DBQ), car and truck version;
3. Decision Making Questionnaire (DMQ);
4. Traffic Locus of Control Questionnaire (T-LoC);
5. User Uptake/Acceptance Questionnaire, one per function tested, repeated at three occasions;
6. Evaluation Questionnaire, one per function tested.

The decision was taken not to develop *new* questionnaires, specific for the SeMiFOT project unless considered necessary but rather to use questionnaires and/or instruments commonly used in studies of drivers and driving. The questionnaires chosen were consequently a combination of custom-made questionnaires (number 1; part of numbers 5 and 6) and established questionnaires and instruments (numbers 2; 3; 4, and part of 5 and 6) often used in field tests of, for instance, safety systems. Unless originally designed or already available in Swedish, the questionnaires were translated from English into Swedish.

The questionnaires were mainly distributed as web-questionnaires using a commercial questionnaire service - Loopon.

5.10.1 Results

5.10.1.1 Response rates

The response rate differed between the different OEMs. While two OEMs had more or less 100% response rate, the result for the other two OEMs was lower.
Two reasons for the lower response rate can be distinguished. First, the time for the trials were shortened to such an extent that there was no time for two intermediate questionnaires, i.e. the 3rd Acceptance Questionnaire had to be excluded and therefore a major part of those participants who did answer the questionnaires only completed the first, the second and the last one, i.e. what was intended to be the fourth. Another reason could be that in the case of the truck companies, several drivers shared

the trucks equipped with the functions to be tested. Hence not all drivers were exposed to the functions to the same degree and, possibly, some drivers completing the first questionnaire did not drive the equipped trucks for more than a limited part of the test period why their motivation may have been low. A third explanation could be the amount of time spent on monitoring and reminding the participants to fill in the questionnaires. Both car manufacturers had assigned the responsibility to one individual who spent substantial time reminding the participants to complete the questionnaires as requested.

5.10.1.2 Background data

In the field test participated a total of 22 car drivers (12 men and 10 women) and 17 truck drivers (16 men and 1 woman). The participants were recruited among the test drivers of the respective OEMs.
The mean age of the car drivers was 45 years (sd +/-10.4) and for the truck drivers 37 years (sd +/-8.2). Approximately 2/3 wore glasses or used contact lenses, while none used any kind of hearing aid.
The participants were all experienced drivers: the car drivers drove on average 26.500 km/year, the truck drivers obviously a lot more, on average 126.000 km/year with trucks plus 19.000 km/year with their private car. On average the truck drivers had totally driven more than 5 million kilometres.
The car drivers were in general “positive” towards driving. They rated their driving style as “somewhat offensive” rather than “defensive”. The truck drivers were “very positive” towards driving cars and especially positive towards driving trucks. They rated their truck driving style as “slightly defensive” while the considered their car driving style as “slightly offensive”.
Questions were asked regarding the participants’ previous knowledge and experience of different systems. The car drivers had, in general, knowledge of the different functions (with some exceptions), but their use experience of the functions was less. Also, if they had experience of use, this was mainly associated with their company car, not their private car. Also the truck drivers had, in general, knowledge of different functions (though again with some exceptions), but their experience of the functions was less. Their use experience originated mainly from functions installed in their trucks, rather than systems installed in their private car. Of the functions tested in SeMiFOT, the truck drivers appeared in general more familiar with LDW and ISA than did the car drivers, whereas the car drivers were more familiar with reverse warning and BLIS (blind spot detection).
The results from the Driver Behaviour Questionnaire compares well with other, comparable, studies of driver behaviour in that the participants in SeMiFOT rated themselves as good drivers. For instance, as a group they “never” or “very seldom” missed a traffic sign, activated the wrong function in the vehicle when driving, or forgot what gear was in place. They admitted, however, to speeding.
The Traffic Locus of Control Questionnaire (T-LoC) was used instead of the general Locus of Control Questionnaire (LoC). T-LoC was, however, only filled in by seventeen participants. Overall the participants who answered the questionnaire appeared to attribute accidents equally to themselves and other drivers but with a tendency to attribute accidents to other drivers. Overall “drivers” rather than “vehicles” or “environmental circumstances” were considered to be the cause of traffic accidents. The participants’ responses to two items can be noted: the negative impact of alcohol and the important of keeping a distance to the car in front. Thus it appears that different information campaigns targeting these factors and their negative influence on traffic safety have had an effect on the driver’s awareness (if not necessarily their behaviour).

5.10.2 User acceptance

User acceptance was measured four times during the test by means of the van der Laan Acceptance Scale; before the test, after a short trial period, after a longer period of time, and after the test. The van der Laan Acceptance Scale consists of nine semantic differentials that are used to calculate two values, one indicating how usable the respondent finds the product/function and one indicating how satisfactory it is perceived to be. The scale ranges from -2 to +2 with -2 being the “best score”.
Most functions were considered moderately usable and satisfactory, with FCW, ACC, and ESC receiving the highest score and BLIS and DSS-R the lowest. (The low score of DSS-R is to be expected since the system is not designed to be of use for the driver, but is rather a tool to study the drivers’ behaviour.)
The results also show that FCW and ACC received higher ratings the longer time the participants used the system. On the other hand, the participants’ satisfaction with the LDW function reduced over time

whereas perceived usefulness remained at the same level. This could be interpreted as though the participants though that LDW is a “good function” as such, but that the performance of the specific systems installed was less satisfactory.

5.10.3 Design of questionnaires and instruments

A number of the questionnaires used were established, validated questionnaires or instruments. However, even though many of them have been – and are still – used for collecting data on drivers in order to investigate differences between drivers with different personality, driving styles, etc., it became apparent that some of the instruments have aged regarding, for instance, the formulation of the items. This was one of the reasons why the Sensation Seeking Scale (SSS) was withdrawn for instance. Overall, it appears as though the instruments need updating and modernisation for the 20th century.

It also appears as though the standardised questionnaires have typically been developed for addressing drivers of private car, not professional drivers in their capacity of, for instance, truck or bus drivers. It could be claimed that the questionnaire could reflect on the driver in his/her capacity of car driver and thus the individuals’ driving in general. However, it can be argued that if the safety systems to be tested are to be introduced into the professional driver’s truck or bus, it is more appropriate to consistently focus on the driver and his/her driving style in this vehicle. Based on the drivers’ self-assessment of their driving styles, the driving styles differ between, for instance, the truck and the private car. Given this observation, the questionnaires could evidently have been modified and new items formulated in order to better fit the professional drivers.

Most of the questionnaires/instruments used were originally designed in a language other than Swedish, most often in English. There are some instruments translations available in Swedish, a few of which are validated. One problem in translating from one language to another (and perhaps back again) is evidently to find the exact correct wording for each question/item in order not to change their intended meaning. To distribute questionnaires in English to Swedish test participants was not, and is not, regarded as a solution. It cannot be assumed that the participants’ knowledge in English is good enough not to result in fundamental and serious errors in the interpretation. Thus, given that questionnaires/instruments to be used in a field operational test need to be translated, substantial work has to be allocated the translation as well as further verification of the translations, before the actual data collection can begin. Based on the experiences from SeMiFOT, the phase in the project could have been allocated more time and resources. Regarding the van der Laan Acceptance Scale, a Swedish version exists but already the English version is problematic in that it has been translated from Dutch to English and, further, that the adjectives used for the construction of the semantic scale can be debated as to their being the best antonyms.

Related to the issues of updating and translating the questionnaires/instruments is the issue of culture and differences between cultures. Particularly in project involving studies in several countries, issues related to possible cultural differences should be considered and pilot tests of for instance questionnaires should be carried out so that these cultural aspects may be identified.

The questionnaires used included some open-ended questions. As expected, few participants added additional comments. In most cases these were short, ranging from a single word to a single sentence but hardly more. Thus the web-based solution does not appear to motivate the participants to elaborate their answers to any higher degree than do the traditional paper and pen solutions.

Open-ended questions require time in order to, and may also be difficult to, encode and analyse. Already the responses from the 39 drivers participating in SeMiFOT took a substantial amount of time to compile and analyse and in some cases it would have been quite feasible to use a closed answer alternative instead of the open alternative chosen. However, in order to provide relevant alternatives, substantial piloting is required before the final questionnaire design is decided upon. Nevertheless, while most often motivated, in particular when reasons behind a rating or ranking are sought, open-ended question may still be feasible to use in a study with few participants. The use of open-ended questions in FOTs with several hundred participants must, on the other hand, be discouraged unless considerable effort and resources are put on pre-processing, translating and/or encoding the verbatims according to a predefined code key. An alternative approach to elicit the necessary more in-depth information could be conducting personal or group interviews with a limited number of sampled test participants.

5.11 Discussion – FOT-Style Analyses

Field trials like this are not anything new and in Sweden we have a strong tradition in this kind of work through the ISA field trials that has been carried out. In the ISA trials however the task was much simpler since there was only one system tested. There were different HMI's tested but the system was the same, i.e. a system intended to make drivers keep to the speed limit in urban areas. To make it even easier each study was also limited geographically to one city, and not even big cities. The results from those studies where primarily mean speed and standard deviation of speed without ISA compared to with ISA. The results where presented for all drivers but divided on different road types. The results were very much on a macro level and even though some analysis was carried out for different groups of drivers it is still on a macro, or semi-macro level. The connection between the performance indicators and safety was based on knowledge on speed changes on macro level using for instance the power model.

Compared to the ISA studies SeMiFOT has much more data available and data for a larger geographical area. This of course gives that the possibilities for analysis are much greater, but also that the complexity is much greater. There is here a need to analyse on microscopic level and the safety indicators we have available today (mostly based on macroscopic data such as mean speed and accidents) are not good enough. It can for instance be hypothesized that a FCW or an ESP leads to higher driving speed since the driver feel safer and based on speed - accident relations this is bad for safety. At the same time there are studies showing that ESP is good for safety (without indicating any change in mean speed so let's assume it is unchanged). So how should the presence of a system such as ESP or FCW be evaluated in terms of safety. Another effect that can be hypothesized is that the presence of ACC and LDW makes drivers more prone to driving on motorways which may lead to an increase in safety. How should this be measured and categorized and is it even possible to compare safety before / after when secondary effects like this may appear? The tasks presented above regarding FOT-Style analyses have in many ways illustrated the complexity of analysing FOT data.

The aim of the FOT-Style Analyses (section 5) was to analyze the *methodology* at an *intermediate* scale by performing an evaluation of selected functions regarding safety, usage and subjective data. It is important to recognize that it was therefore *not the intention to perform a full evaluation of the safety benefits* of the safety functions in SeMiFOT. The methodology development was at the center of attention. To be able to perform a true evaluation, a full study such as EuroFOT would have to be designed. SeMiFOT was envisioned as a sort of pilot and methods development precursor to full-evaluation projects.

A novel method which was developed within SeMiFOT, the events-prevented analysis, was found to be feasible and was able to show a significant benefit of the lane departure warning system in terms of the dispersion of simulated trajectories. This method is important because it is ultimately capable of answering both whether a system acting in a event had an effect in that *particular* event, and of answering if the system would have had an effect in an event where no system action was enabled. Extensions of this "what if a system could have acted?" method will be developed in the SeMiFOT2 continuation project.

The task Crash-Relevant Event-based Safety Impact Assessment (CRESIA) Method presents a theoretical discussion and suggestion for an implementable method on one of the most difficult and underdeveloped, yet most sought after, areas within the FOT methodology, i.e. the detection and use of crash relevant events and their relation to actual risk. The work describes an analytical approach which uses a comparison of crash-relevant events with and without a safety system active, and takes into account some of the problems mentioned above. Like the events prevented analysis, it also represents a novel method and can be seen as a first step to evaluate safety benefits of a system by use of CREs. A few gaps remain which need to be filled, where links based on objective data are missing one has to rely on expert ratings. The method however is a good first approach and the classification of events is vital to be able to create an objective link between events and their relevance for safety.

Most of the tasks regarding FOT-Style Analyses in Section 5 have encountered the problem of determining the baseline or what to use as a ground truth. It is easy to say “we measure everything, therefore we can analyse everything” but the analysis work here has shown the need both of a careful study design to make sure that there is the possibility to find baseline data but also careful consideration of inclusion and exclusion criteria. Figure 8 regarding Visual behaviour analysis of ACC illustrates some of the considerations that need to be taken when analysing ACC in terms of system status. In addition to this the usage analysis shows some of the considerations that need to be taken in terms of the environment. Still there are more things that need to be considered to be able to do a correct comparison of data at this detailed level.

The tasks above have also shown the importance of careful quality checks of data. In for instance Visual behaviour the sometimes poor quality of data has made it very difficult to draw any conclusions and to say whether there is an effect or just a random effect due to the noise in the data.

A consequence of both the findings above is that there is often a need to check the data manually and even screen data manually as is the case in Events prevented. This will lead to possible bias not to mention that it is time consuming. For a study of this size it may be feasible to do some manual work but when it comes to scaling up to large scale trials other means of screening and control need to be developed.

The time spent for analysis in the above tasks has in many cases by far exceeded the expectations and the manual work is one part of the explanation but no the only one. The other parts of data control which covers quality control but also the next step which is the investigative work in terms of finding out why there are missing signals, why there are spikes in the data, why all cars for one manufacturer indicate that they are turning left when they are braking etc. has also taken a lot of time.

The “lessons learned” above are of course not all, there are many more mentioned in the individual reports that has been of great value in terms of better understanding the problems and potential of the FOT methodology. The work in this section concerning the evaluation of systems was strongly influenced by the focus on methods development, in contrast to simply executing an evaluation of systems based on established knowledge. With very limited resources, attention was paid to developing competence and methodology. The outcome of this approach was successful as can be seen in

- the development of innovative methods (events-prevented, and CRE of Safety Impact),
- the first application of automatic visual behavior analysis in an FOT,
- a successful usage analysis which included the development of a statistical model, and
- an evaluation of the acceptance of all systems, and of subjective methods (questionnaires and instruments)

In closing, please note that the mistake should not be made to interpret the outcome of the SeMiFOT project in terms of whether it evaluated the safety impact of the systems or not. This was not the intention and this task is left for future full-FOT studies. The question still remains however, whether we have enough knowledge today to do that efficiently and properly. There is a need to do more work in line with what was carried out when the conflict technique was developed and validated, i.e. researchers working together with their respective data sets to take the method further.

6 The Opportunities for FOT/Naturalistic Data Contribution of Safety Knowledge in the Product and Infrastructure Development Process

6.1 Objectives

The objective with this task (Section 6) is to analyze how FOT-data can contribute to the implementation of safety knowledge in the product and infrastructure development process. The results are based on discussions and written contributions involving different SeMiFOT-partners. This summary will give a brief overview while the full task report provides more detail on the subject. The first part describes the safety working processes for the automotive industry and for the Swedish road authority separately since the development process of vehicles and infrastructure is quite different. The second part points out the potential of driving data for further analysis and contribution in the safety development of vehicles and infrastructure respectively by providing a few specific examples. Finally a description of how the results from project are actually implemented up to date in the safety work by the different partners is presented.

6.2 The safety working processes - Automotive industry

Safety development is described as a circular process starting with real world data which can be either be accident investigations, driving data, incidents or even more general as driver habits, attitudes etc. A combination of in-depth and statistical data together with human data (biomechanical, behavioural) provides information for priorities as well as more in-depth understanding about the mechanism behind injuries, accidents or unsafe driving. The knowledge gained forms the basis for internal safety requirements which are included at the very beginning of product development projects. Physical or virtual prototypes and test methods can be used during the product development. Field effect estimates can be done based on analysis of technical performance and available field data. The final product is then verified with test methods and criteria derived from the safety requirements, before going into production. Once new products are on the road, new field data can be collected and analyzed, and the process continues.

6.3 Road authority safety working process

There are three general areas in which FOT-data have their main benefits for the road infrastructure authorities – Swedish Road Administration (SRA) and municipal road owners. They are 1) planning of new infrastructure, 2) improvements and maintenance of road infrastructure and 3) winter road maintenance.

The infrastructure development cycle is a relatively slow process at the SRA. The first step is the acquisition of real world data and potential safety concerns may arise. Hereon after the data/information is used in a decision making process in the infrastructure planning procedure. Many different aspects are weighed into the planning and construction of new roads and design specifications of the road infrastructure are also regulated by the SRA’s construction manuals. These manuals may need amending or updating which is a separate process.

6.4 Implementation of FOT-data in the vehicle product development process.

FOT can provide insights into different driver safety related needs (or problems) in accidents-incidents-driving. Examples could for instance be high demand driving situations, distraction due to secondary task engagement or reduced ability due to drowsiness. Driver information needs while driving or operational performance in critical situations can also be studied. Driver needs may be statistically quantified partly based on driving data. The FOT-data can also provide more detailed understanding and scenario description (e.g. traffic situation, velocity, distance, visibility, driver and road conditions) of each problem area which is a key factor in the creation of *product requirements*. Driving data can also be used for estimating the safety potential of different conceptual solutions or to make effect analysis on more mature technical solutions.

6.5 Implementation of FOT-data in infrastructure development.

There are a number of examples where FOT-data can be very beneficial to the road authorities e.g. black spot identification; road user behaviour and road traffic safety countermeasures. This could be of a general or a specific nature e.g. drowsiness in general; speeding in certain specific areas; distraction from road-side eye-catchers; systematic errors at specific intersections etc.; traffic flow data and driver behaviour in relation to traffic information and congestion in urban areas.

6.6 Actual implementations made at the end of the SeMiFOT project.

The implementations made can be divided by methodology and analysis:

Methodology in this case refers to instrumentation of vehicles, processing of data to different measures. The principle steps of formulating hypothesis, defining performance indicators and estimating impact on safety, environment etc.

- The experience and methods established are implemented and further improved in ongoing larger scale FOT-project (e.g. EuroFOT, all Swedish partners).
- The same methodology is in near future being partly used for vehicle verification methods, such as experiments on-road or on a closed test track. (Vehicle manufacturer).

Analysis in this case is using the SeMiFOT data in order to explore hypothesis that are outside the scope of the project.

- There are several ongoing studies aiming at understanding driver behaviour in relation to e.g. warnings, choice of speed etc. The studies vary from research to development of next generation safety systems. (Vehicle manufacturers).

7 Analysis of consumer- and research-driven data collection opportunities for full-scale FOTs

On-board monitoring systems are used to gather information about driver behaviour and/or vehicle performance. A wide range of systems are today commercially available. This task report includes a state-of-the-art review of available on-board monitoring systems that enables data collection for FOTs. The report starts with a description of some EU systems, followed by US systems, Japanese systems and other. The report illustrates the wide range of on-board monitoring systems available on the market. The output given from some of the reviewed systems is used to give feedback to the driver or a fleet manager in real-time or at some time after the drive, other systems are primarily used for research and insurance purposes. Some of the reviewed systems include continuous data recording, whereas others record data only when an acceleration threshold is exceeded. Some systems record vehicle signals as well as signals from external sensors and video and are complex to install, whereas others are hands-on aftermarket products easily installed by the driver. These aftermarket systems principally include accelerometers and cameras.

Subsequent to the review of systems the insurance company Volvia Försäkringar discuss if and how the different kinds of systems are usable for insurance purposes. Today's arguments for insurance premiums have the benefit of being easy to handle, and a new customer can easily take out insurance without answering a lot of personal questions. This however gives a rather blunt classification of the customer, and the possibility of more detailed information about the customers driving, such as actual driven kilometers, could lead to improved customer segmentation. If an on-board monitoring system should be used for insurance purposes, it's crucial that the system cannot be manipulated by the customer. Neither should he/she have the possibility to decide when to share the actual data, for example, not only share the data to prove his/her innocence, which seems to be the case with the Japanese end user (consumer) systems. With a system that registers unsafe driver behavior, it could be possible to also have more detailed arguments for insurance premiums based on risky driver behavior.

At last, the convenience of the different kinds of systems for a large-scale FOT is discussed. What systems are suitable for large-scale FOTs? It depends on the purpose. Is data from serious accidents essential, or is continuous higher-fidelity information needed? How can event data recorders be distributed and for what fundamental purpose (other than research)? In Japan, video-based event data recorders are installed by fleet owners in cooperation with universities and whereas the universities use the data for research the fleet owners can use the information if vehicles are involved in crashes. The same strategy could be employed in Sweden where taxi fleets, car pools, truck fleets and insurance companies can be involved.

8 Benefits of Origin and Destination Information in Vehicle-Infrastructure Integration (VII) Data Set

Trip origin and destination (O-D) data plays a crucial role in various transportation activities. This information not only includes the starting and end points of a trip, but also information that can be obtained through the ability to track vehicles across a network. Collected information may include the location and speed of a vehicle every second, a record of the links entered during a journey, the time a vehicle has entered each link, the time taken to travel a link, etc. Despite some recognized benefits, there are still concerns, primarily related to privacy, about whether or not probe vehicles should be allowed to collect O-D trip information. Within this context, this report examines various issues associated with the potential collection of O-D data by IntelliDrive probe vehicles.

Research findings presented in the report include:

- An overview of current uses of O-D data in activities related to transportation planning and system operations.
- A description of the methods currently used to collect or generate O-D trip data.
- A description of issues associated with current O-D data collection and generation methods, as well as how O-D trip data are applied in transportation applications.
- A discussion on how O-D information collected from IntelliDrive probe vehicles can enhance existing applications and enable the development of new applications.

- A discussion on the potential benefit-cost ratio offered by the inclusion of O-D trip information in IntelliDrive data sets.
- A discussion of privacy issues regarding O-D data collection. This includes a review of current concerns, emerging policies regarding driver privacy in IntelliDrive systems, methods currently promoted to ensure driver privacy, impacts of promoted privacy methods on IntelliDrive applications, and methods available to mitigate privacy concerns.
- A simulation study with the Paramics microscopic traffic simulation model demonstrating the benefits to individual vehicles and network operations of using O-D data reported by IntelliDrive vehicles to provide dynamic route guidance around an incident.

9 Analysis of Requirements and Interest for a Next Phase – SeMiFOT2

9.1 Introduction

Throughout the project lifetime, project partner interest in continuation projects were brought to the attention of the SeMiFOT Steering Group. Some of these interests in continuation projects of similar nature to SeMiFOT were channelled into separate projects and some of the interests were channeled into a continuation project called SeMiFOT (presented below and in SeMiFOT2 Task Report). When the project proposal for SeMiFOT was written, it was envisioned that SeMiFOT Phase 1 would be a sort of a pilot project for a larger Field Operational Test in which the safety impact of selected functions would be tested. However, as is often the case, partner interests and funding opportunities have to be successfully matched in order to win projects.

Various opportunities presented themselves for (a) new SeMiFOT-independent projects and (b) a continuation project called SeMiFOT2. As a matter of interest for the reader, new projects that were somehow inspired, build upon, or connected with the work done in SeMiFOT include EuroFOT (www.eurofot-ip.eu), FOT-NET (www.fot-net.eu), BasFOT (SAFER project), TeleFOT (www.telefot.eu), and DREAMi (SAFER-Japan project). The continuation project called SeMiFOT2 is described in summary below and in detail in a separate Task Report.

SeMiFOT2 will provide additional analysis and methods development, over and above the original SeMiFOT Phase 1 project. The additional analysis goals are as described below. Most notably, the SeMiFOT2 project will considerably increase the competence and strength of a SAFER consortium to apply for identified US analysis projects, and potential future EU projects. The SeMiFOT2 project is an extension to the existing SeMiFOT Phase 1 project. As such, SeMiFOT2 uses the established SeMiFOT Phase 1 project agreements and management frameworks.

9.2 Goals and Relevance of the Project

At a general level, three main Swedish needs are adressed by this project:

1. **Knowledge about what causes accidents and what can be done about them** – There is a strong need for an objective understanding of pre-crash data, particularly related to driver behavior (see e.g. Boyle et al, 2009).
2. **Competitiveness of Swedish vehicle industry in crash countermeasure development.** – The capability of the Swedish vehicle industry to take action based on knowledge of accident causation is strongly supported by this project, thereby stimulating growth of safety technology and services. Countermeasure development includes identification and prioritization of countermeasures based on pre-crash data, evaluation of new technology/solutions, and innovation and development of new technology based on project results.

3. **Cooperation with stong international actors** – Through cooperation with leading international actors, Swedish partners become stronger. In particular, two ongoing international collaborations are targeted:
    a. **The SHRP2-SAFER Memorandum of Understanding (MoU).** The current project constitues a realization of the next steps for joint research activities that were identified by SHRP2 and SAFER. These next steps were identified within the context of the MoU signed by VINNOVA / the Swedish Road Administration and the Transportation Research Board of the National Academy of Sciences to facilitate joint research activities between SAFER and SHRP2. Recall SHRP2 is a very large Naturalistic Driving Study (about 2500 cars collecting data for 2 years).
    b. **Sweden-Michigan Partnership agreement.** The current project is a further realization of work within the Sweden-Michigan partnership agreement as signed by VINNOVA, Swedish Road Administration, Michigan Department of Transportation, and Michigan Economic Development Corporation. Specifically, the University of Michigan Transportation Research Institute (UMTRI) will conduct further work within the project.

9.3 Plans

The SeMiFOT2 project builds on ongoing activities within the SeMiFOT Phase 1 project. It will provide additional analysis and methods development, over and above the original SeMiFOT Phase 1 project. For a relatively small amount of time and resources spent on analysis at the end of the project, a very significant added value can be added to resources already spent. Strong results from the SeMiFOT projects will be used to further advance Sweden's position in the traffic safety field by establishing the value of analyses based on the Naturalistic FOT method. Traffic safety research in Sweden has previously been lacking a component to understand objectively pre-crash driver behaviors and how active safety systems influence the pre-crash and crash outcomes. If strong results are shown, the SeMiFOT projects can be very effective in raising the profile of Sweden (SAFERs partners) by facilitating the establishment of other projects (EU, USA and Japan).

The current SeMiFOT2 project will make SAFER, in collaboration with UMTRI, better suited to apply for and perform Strategic Highway Research Program 2 (SHRP2) analysis project(s). SHRP2 is a very large Naturalistic Driving Study (about 2500 cars collecting data for 2 years). The SeMiFOT2 project would considerably increase the competence and strength of a SAFER consortium which would apply for a SHRP2 analysis project. SHRP2 has subsequently confirmed their support in emails and telephone conversations for the current project proposal as a realization of the next steps in the MoU that both SAFER and SHRP2 have agreed upon.

To be aligned with the goals of SHRP2, it is impotant that the SeMiFOT2 analysis has "emphasis on questions that require data about drivers, those that have the potential to support safety interventions, and those that address large-scale morbidity and mortality consequences." (Boyle et al 2009, p12). Boyle et al (2009) further state that research questions should relate to crash risk and driver behavior, should analyze data beyond what is currently available, should be best suited for naturalistic data, should be associated with a "straighforward intervention", i.e. an intervention for which we know (more or less) what to do (e.g. infrastructure improvements, in-vehicle system enhancements, educational, policy), and should provide some fundamental understanding of the basic mechanisms of motor vehicle crashes and driving behavior that can be generalized to other situations.

Specifically, the SeMiFOT2 project will develop the competence and expertise that is needed for answering the following SHRP2 global research questions (see Boyle et al 2009) in future analysis projects. Notably, the first question can be seen as a prerequisite for understanding the "influence on crash likelihood" in the latter 3 questions.

1. What explanatory factors are associated with crashes or crash surrogates and what analytical models can be developed to predict crash or crash surrogates? (SHRP2-GRQ9b)

2. How does driver distraction influence crash likelihood? (SHRP2-GRQ3)
3. How do advanced driver support systems influence crash likelihood? (SHRP2-GRQ6)
4. How do dynamic driver characteristics [inattention, fatigue, workload, etc], as observed through driver performance measures, influence crash likelihood? (SHRP2-GRQ1)

In line with these research goals, the following analysis workpackages have been selected (see SeMiFOT2 task report for details). The workpackages are selected because of their potential to develop competence and expertise within analysis areas which are particularly relevant for both the SHRP2 project goals and for the three main Swedish needs outlined above – Knowledge, Competitiveness, and Cooperation.

- WP0 Management
- WP1 Development of quantative definitions of crash-relevant-event severity
- WP2. Exploration of new statistical and analytical approaches for the analysis of Crash-Relevant Events
- WP3 Extraction of Crash-Relevant Events
- WP4 Analysis of the impact of visual behavior on driving performance and crash-relevant events
- WP5 Expansion of SeMiFOT simulation techniques
- WP6 Project reporting and paper writing

10 Discussion and Conclusions – WP5 Evaluation of Methodology

The overall aim of this workpackage was to evaluate the Naturalistic FOT *methodology*, rather than an evaluation of the safety systems that were included in the FOT. A smörgåsbord approach was chosen as the general analysis approach in order to give a "taste" and experience with implmenting many different types of analyses. The size of smörgåsbord was however reduced by excersizing restraint on the number of hypotheses investigated. A diversity of analysis methods were applied, ranging from the fully innovative to the established. This was achieved against the backdrop of some serious constraints, such as limited resources (19 person months for 16 tasks and 13 partners), many safety systems, many hypotheses prioritized, restricted data access procedures, and one of the largest datasets collected in Swedish safety research.

Nine tasks were directed squarely at fullfillment of the first overall goal for the project:

> ***Goal 1****: to further develop the Naturalistic FOT method into a powerful tool for a) accident research, and b) evaluation of safety, efficiency, and usage & acceptance, and c) countermeasure innovation and development.*

The naturalistic FOT method as a tool for accident research was analyzed by three diverse tasks. The work on *crash-relevant events (CREs) analysis* was successful in developing and implementing a method for extraction, coding, and validation of CREs. Further, it showed positive results by applying driver-individual trigger thresholds, and by combining CREs and accident statistics geographically. The *visual behavior analysis* encountered serious difficulties with the data quality of current eyetrackers and the subsequent problems in making classification algorithms work well despite fluctuating data quality. Despite this, development of automatic inattention detection will continue to be highly prioritized because of the central role inattention plays in crash causation. The *automatic speed camera analysis* showed how naturalistic data can be used with map data to assess infrastructure-related research questions and was able to show that there is a kangaroo effect whereby drivers slow down before and accelerate after speed cameras. These analysis experiences represent a valuable step in the development of the competence and methodology needed to analyze the explanatory factors in accident causation.

Five tasks were directed at goal 1b) regarding evaluation of safety, usage, and subjective data. The aim of these tasks was to analyze the methodology and *not* to perform a full evaluation of safety functions. Here, a novel simulation method – the events-prevented analysis – was found to be sucessful in being able to determine "what-if" a system didn't act, or conversely if a system had acted. A novel method for CRE-based safety impact assessment was presented but not tested. The visual behavior analysis of adaptive cruise control also suffered from the aforementioned data quality issues, but illuminated the problem of determining the baseline. The usage analysis was able to successfully answer a hypothesis on the influence of road curvature, and the subjective data analysis was able to evaluate acceptance for all the SeMiFOT safety systems.

Before analysis could be carried out in these nine tasks above, the data was prepared in order to achieve data with accurate quality, provide guidelines and templates for the calculation of new measures, and to complement the recorded data with new measures. This preparation work was indispensable for the analysis by providing driver ID, road attributes, guidelines for operating efficiently on the data, and estimations of quality for various measures. Perhaps not surprisingly, the time spent for analysis exceeded initial expectations as the majority of analysis time was spent on data processing and measure calculation. Therefore, the strongest conclusion was that analysis efficiency would improve greatly from preparing the data before uploading it to the database.

Goal 1c) on countermeasure innovation and development was targeted directly by the analysis task in Section 6 on the contribution of safety knowledge in the product-, and infrastructure development process. In direct support of Goal 1c) and Goal 2 (to improve the competitiveness of the participating partners and increase the opportunities for economic development), the naturalistic FOT methodology is currently being used in the development processes in various ways from research to development of next generation safety systems.

The cooperation with UMTRI in WP5 was substantiated in a, for the project, very beneficial general advisory role in all manners of analysis-related issues. UMTRI responded with great success to the idea of developing an events-prevented analysis methodology. In another analysis task, the *Benefits of Origin and Destination Information in Vehicle-Infrastructure Integration (VII) Data Set* various issues associated with the potential collection of O-D data by IntelliDrive probe vehicles were examined. These tasks supported Goal 4 to achieve a close cooperation between Swedish and Michigan partners.

With regards to Goal 6 to determine how to measure the impact of intelligent vehicle systems with a larger study. SeMiFOT was originally intended to be a pilot project for a larger Field Operational Test. EuroFOT and TeleFOT became such full FOT projects for six SeMiFOT partners quite early on in the project. In determining the next step after SeMiFOT, a continuation project (SeMiFOT2) was created to target natrualistic driving analysis opportunities (SHRP2 and EU).

Further support in future projects is the analysis task of a large number of commercially available on-board monitoring systems from EU, USA, and Japan were reviewed with regard to various aspects of suitability for future studies. An insurance company – Volvia Försäkringar – analyzed the usability of these systems for insurance purposes, perhaps enabling a business model involving cost sharing between drivers, insurance, and research funding.

In sum, the naturalistic FOT method continues to hold great promise as a method providing new data and insights regarding safety issues which cannot be reached with other methods. However, the analysis methodology is still currently in an immature phase requiring futher development.

11 References

Boyle, L. N., Lee, J. D., Heyens, K. M., McGehee, D. V., Hallmark, S., Ward., N. J. (2009). *S02 Integration of Analysis Methods and Development of Analysis Plan: DRAFT Phase I Report.* SHRP2 SO2 Phase I Report, University of Iowa.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J. et al. (2006). *The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment* (Rep. No. DOT HS 810 593). Blacksburg: Virginia Tech Transportation Institute.

Draskoczy, M. ; Carsten, O. M. J. ; Kulmala, R. (1998). *Road Safety Guidelines*, Deliverable B5.2 of the CODE Project, Telematics Application Programme.

Englund, Anders, ”Trafiksäkerhet: en kunskapsöversikt”, *Kommunikationsforskningsberedningen*, KFB and Studentlitteratur. Sweden, 1998.

Engström, J and Ljung Aust, M. in press. Adaptive behaviour in the simulator: Implications for active safety system evaluation. To appear in D. Fisher, M. Rizzo, J. Lee and J. Caird (Eds.) Handbook of Driving Simulation for Engineering, Medicine and Psychology. CRC Press.

Gordon et al 2009a: Yaw Rate Error – A Dynamic Measure Of Lane Keeping Control Performance for fhe Retrospective Analysis of Naturalistic Driving Data, ESV Conference Proceedings, Stuttgart, www-nrd.nhtsa.dot.gov/pdf/esv/esv21/09-0326.pdf

Gordon et al 2009b: Safely Impact Methodology, Final Report of the Advanced Crash Avoidance Technologies Program (ACAT), for the National Highway Traffic Safety Administration by Ford, Volvo and UMTRI. (to appear, NHTSA, 2009)

Hantula, L. (1994). *Accident Data Collected by Road Accident Investigation Teams in Finland. The OECD Workshop on Accident Data System.* Jurmala, Latvia.

Hickman, J.S., Knipling, R.R., Olson, R.L., Fumero, M.C., Blanco, M., and Hanowski, R.J. (2005). Phase I: Preliminary analysis of data collected in the drowsy driver warning system field operational test – Task 5: Preliminary analysis of drowsy driver warning system field operational test data. Contract No. DTNH22-00-C-07007.

Hydén, C. (1987). The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique. *Bulletin 70*, Lund University, Lund, Sweden.

I. Isaksson-Hellman, H. Norin. HOW THIRTY YEARS OF FOCUSED SAFETY DEVELOPMENT HAS INFLUENCED INJURY OUTCOME IN VOLVO CARS. 49TH Annual Proceedings AAAM, 2005.

Ljung Aust, M. and Engström, J. in preparation. A conceptual framework for requirement specification and evaluation of active safety systems.

McGehee, D. V., Raby, M., Carney, C., Lee, J. D., & Reyes, M. L. (2007). Extending parental monitoring using an event-triggered video intervention in rural teen drivers. *Journal of Safety Research, 38,* 215-227.

Najm, W., Stearns, M., Wowarth, H., Koopmann, J., & Hitz, J. (2006). Evaluation of an Automotive Rear-End Collision Avoidance System. *National Highway Traffic Safety Administration*. DOT-VNTSC-NHTSA-06-01, DOT HS 810 569

Ross, S. (1993) *Introduction to probability models.* Academic Press.

Sahai and Khurshid (1996), *Statistics in Epidemiology*. CRC Press. ISBN-10: 0849394449

Sayer, J., Devonshire, J. M., Flannagan C. A. (2005). *The effects of secondary tasks on naturalistic driving performance* (Rep. No. UMTRI-2005-29). Ann Arbor: The University of Michigan Transportation Research Institute.

Shankar, V., Jovanis, P. P., Aguero-Valverde, J., and Gross, F. (2008). Analysis of Naturalistic Driving Data. Prospective View on Methodological Paradigms. Transportation Research Record: Journal of the Transportation Research Board No 2061. Transportation Research Board of the National Academies, Washington, D.C. DOI: 10.3141/2061-01

Therneau, T. M. & Grambsch, P. M. (2000). *Modeling survival data, extending the Cox model:* Springer

Vadeby, A., Forsman, A, Kecklund, G., Åkerstedt, T., Sandberg D., Anund, A. (2009, in press) Sleepiness and prediction of driver impairment in simulator studies using a Cox proportional hazard approach. *Accident Analyses and Prevention*

Yee, S., Nguyen, L., Green, P., Oberholtzer, J., & Miller, B. (2007). *Visual, auditory, cognitive and psychomotor demands of real in-vehicle tasks* (Rep. No. UMTRI-2006-20). Ann Arbor: The University of Michigan Transportation Research Institute.

SAFER

VEHICLE AND TRAFFIC SAFETY CENTRE AT CHALMERS

VOLVO

SCANIA

SAAB

Autoliv

vti

FINDING A BETTER WAY

CHALMERS

SP

M

UMTRI

LINDHOLMEN SCIENCE PARK

Länsförsäkringar

VOLVIA

Gör en bra bil bättre.

Vägverket

VINNOVA

MDOT

Michigan Department of Transportation