SEVENTH FRAMEWORK PROGRAMME
THEME 7 TRANSPORT

Grant Agreement Number:	233597
Acronym:	PROLOGUE
Full Title:	**PRO**moting real **L**ife **O**bservations for **G**aining **U**nderstanding of road user behaviour in **E**urope
Funding Scheme:	Collaborative project small scale
Project Co-ordinator	SWOV Institute for Road Safety Research
Duration:	01/08/2009 to 31/07/2011
Project Website:	www.prologue-eu.eu

Summary and integration of a series of Naturalistic Driving field trials

Deliverable D3.7

Authors: Agathe Backer-Grøndahl (TØI), Tsippy Lotan (Or Yarok) and Ingrid van Schagen (SWOV)

Status: Final draft submitted to EC

Date: 19 July 2011

Dissemination Level: Public

Project Consortium

1		SWOV Institute for Road Safety Research (project co-ordinator)	NL
2		CERTH/HIT Hellenic Institute of Transport	GR
3		KfV Kuratorium für Verkehrssicherheit	A
4		Loughborough University	UK
5		Or Yarok	ISR
6		Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek - TNO	NL
7		TØI Institute of Transport Economics	NO
8		Test and Training International Planning and Service GmbH	A
9		Universitat de València	ES

Acknowledgements

Description

We would like to thank all field trial leaders and co-workers for their valuable insights and suggestions for points of discussion. We also like to thank the members of PROLOGUE's Advisory Board for their useful suggestions for improving the report.

Document Reference

Backer-Grøndahl, A., Lotan, T., Van Schagen, I. (2011). *Summary and integration of a series of Naturalistic Driving field trials.* PROLOGUE Deliverable D3.7. TØI Institute of Transport Economics, Oslo.

Abstract

The main objective of the EU project PROLOGUE (PROmoting real Life Observation for Gaining Understanding of road user behaviour in Europe) is to assess the usefulness and feasibility of a large-scale Naturalistic Driving study in Europe. In order to address this objective, various means have been used, including several literature reviews, questionnaire studies, discussions with potential stakeholders, and, last but not least, five small-scale Naturalistic Driving field trials. The aim of these trials was twofold:

- To explore the potential and get a better feeling of the technological aspects of Naturalistic Driving research, and the strengths and weaknesses of various approaches.
- To illustrate the potential usefulness of naturalistic observations for various aspects of road safety.

The current report presents a summary of these five field trials and integrates and discusses the results and the lessons learned. The five field trials differed substantially with regard to technological aspects such as type of equipment, composition of sample and sampling strategy, methodology, analysis methods, et cetera. The reason for this diversity was that it resulted in a wide range of experiences when performing Naturalistic Driving studies in practice; experiences associated with, for example, data acquisition and storage, study design, participants, data reduction, and privacy and ethical issues. The five trials also widely varied in the reseach areas and research questions they addressed. The reason here was that the trials gave some impression of the wide range of application areas of the Naturalistic Driving approach.

The five trials contributed to the identification of the essential elements of the naturalistic driving approach and, consequently, of the important aspects of a possible future large-scale Naturalistic Driving study.

Table of Contents

Executive Summary

Introduction

The main objective of the EU project PROLOGUE (PROmoting real Life Observation for Gaining Understanding of road user behaviour in Europe) is to assess the usefulness and feasibility of a large-scale Naturalistic Driving study in Europe. In order to address this objective, various means have been used, including several literature reviews, questionnaire studies, discussions with potential stakeholders, and, last but not least, five small-scale Naturalistic Driving field trials. The aim of these trials was two-fold:

- To explore the potential and get a better feeling of the technological aspects of Naturalistic Driving research, and the strengths and weaknesses of various approaches.
- To illustrate the potential usefulness of naturalistic observations for various aspects of road safety.

The current report presents a summary of these five field trials and integrates and discusses the results and the lessons learned.

Overview of the trials

The five field trials were conducted in Israel, Austria, The Netherlands, Spain, and Greece. The trials were set up in such a way that they differed substantially with respect to technological aspects such as type of equipment, composition of sample, type of cars used, research method, analysis methods, et cetera. The reason for this diversity was that the trials were meant to provide a wide range of experiences with conducting Naturalistic Driving (ND) studies in practice; experiences associated with, for example, data acquisition, storage capabilities and data download, study design, participants, data reduction, and privacy and ethical issues. All field trials dealt with the issues of data collection and storage, and data reduction and interpretation. In two of the trials two independent data collection systems were integrated. While the Israeli trial integrated two off-the-shelf in-vehicle systems measuring two types of different behaviour, the Dutch trial integrated data from an in-vehicle system with site-based continuous data monitoring. Three trials explicitly looked at the relationship between ND information about driving behaviour and self-reports, and the same three also applied the ND approach to evaluate a particular intervention. On all other aspects the trials differed.

Furthermore, the trials varied widely in the research areas and research questions they addressed. The reason here was that the trials were also meant to give an impression of the wide range of application areas of the ND approach, and the capabilities of different technologies and methodologies to address various research questions. The field trial in Israel looked at novice drivers; the trial in Austria at driver training, the Dutch trial on the interaction between cars and bicycles, the Spanish trial on the use of an in-car information system, and the Greek on the effects of active safety systems.

It should be noted that most of the trials were not ND studies in the strictest way, i.e. in the sense that they collected data on normal driving behaviour without any intervention. For example, in the Austrian and Israeli trial the naturalistically collected data were also used to provide feedback to the drivers and, subsequently, it was studied whether that type of feedback affected driving behaviour. In the Greek trial, the ND data was used to asses the effects of the implementation of two warning systems in the car by looking at the naturalistic behaviour in an instrumented car before and after the implementation. This latter trial therefore came very close to a Field Operational Test with an ND type of data collection. However, all trials included a naturalistic data collection phase in which no intervention or feedback to drivers was administered.

As the trials were small-scale, and the samples small and often not representative, the trials were not meant to provide sound scientific conclusions about road safety behaviour.

Conclusions and lessons learned

The variations among the field trials resulted in different experiences and contributed to a further understanding of the various aspects to consider when conducting an ND study.

Data collection and data storage

With respect to data collection and storage, the trials showed that is was possible to collect data on a continuous basis and to store it. However, the main challenge here is to decide what type of technology should be used: just G-based vehicle parameters, or also event-triggered video data, or even continuous video data; and if video is included, how many cameras are desirable and at what positions. It was concluded that a basic and relative simple and cheap off-the-shelf G-based data acquisition system (DAS) provides very useful data for many research questions. Attention must be paid to the reliability and validity of the identified safety-related events, i.e. avoiding false alarms and missed events.

Continuous video data is important for accomplishing the full potential of ND studies. This is particularly true for estimating the relative risk (odds ratios) for various risk factors. Continuous video is also needed for identifying events or near-crashes where drivers had not undertaken action while they better had, e.g. because they were sleepy or distracted. Furthermore, continuous video data can be important for validating the G-based events by identifying false and missed events.

However, a very advanced DAS, including continuous video registrations by several cameras requires much storage space, and consequently very frequent downloads. Since downloading of this amount of data cannot be done wirelessly, participants have to go to a download station for manual downloads on a regular basis. This puts a burden on the participants and also threatens the desirable unobtrusive character of the trial since the manual download reminds the participants of the study.

Data reduction and interpretation

For data reduction and interpretation, the type and amount of data that is collected is crucial. In particular, the manual processing and analysing of large amount of video data is very time-consuming. Automatic processing and analyses tools are not yet sufficiently reliable, but developments in this area are promising.

Regarding data interpretation, various types of complementary data are helpful, including complementary data of the driving surroundings, such as characteristics of the infrastructure, traffic flow, and weather, and complementary data of the participants, such as personal characteristics, socio-demographic data, and crash history.

System integration

Though only studied in a very limited way, the trials showed that it is difficult and time-consuming to integrate different DASs. This is true both for integrating systems with the aim to compare data from different (national) studies as well as for integrating systems to get additional, complementary data. Therefore, it would be advisable to use one DAS that collects and inherently integrates all the required data.

Integrating site-based video information and in-vehicle data proved to have added value when studying the interaction between cars and bicycles. One issue here is that the use of fixed cameras restricts the observations to a limited number of locations and, consequently, that a long period of data collection is required to get sufficient 'events of interest'.

ND data and self-reports

Self-reports in ND studies cannot only be used to verify the ND data and reduce data loss, but also to complement the ND data. From methodological point of view, studying the correspondence between ND data and self-reports are important when tackling the issues of verification of self-reported behaviour. Furthermore, self-reports contribute to the understanding of ND observations, as ND data does not contain enough information to understand the motives for underlying specific behaviours. Hence, the combination of ND and self-reports can provide strong evidence to not only *how* road users behave, but also to *why* they behave the way they do.

Evaluation of interventions

Whereas an ND study in the strictest way does not include any interventions and just monitors drivers in their everyday driving, ND data can also be used in a useful way to evaluate interventions or to compare the behaviour of different driver groups or in different situations. The ND data can also be used as source of information as such for the driver to confront them with their own behaviour, aiming to change that behaviour. This is not only applicable for road safety, but could also be used for eco-driving training.

In summary

In summary, it can be concluded that the five small-scale field trials successfully demonstrated the usefulness and feasibility of conducting ND studies for gaining knowledge in various road safety areas. Furthermore, the experiences gained in the trials showed that there are several important and crucial factors that need to be seriously considered before conducting a large-scale ND study and indicated several directions for solutions.

.

1 Introduction

1.1 The PROLOGUE objective

Within the EU project PROLOGUE (PROmoting real Life Observation for Gaining Understanding of road user behaviour in Europe), researchers from several European countries and disciplines have been investigating and exploring the field of Naturalistic Driving (ND) studies. Typically, in an ND study passenger cars, generally subjects' own cars, are equipped with devices that continuously and unobtrusively monitor various aspects of driving behaviour, including information about vehicle movements - e.g. acceleration, deceleration, position on the road, driving speed -, about the driver - e.g. eye, head and hand movements -, and about the direct environment - e.g. traffic densities, time headway, road and weather conditions. This technique makes it possible to observe and analyse the interrelationship between driver, vehicle, road and other traffic in normal situations, in conflict situations and in actual crashes.

The main objective of PROLOGUE is to assess the usefulness and feasibility of an ND study in Europe and its value for increasing our knowledge about, in particular, road safety, but also about other traffic related issues such as eco-driving and traffic management.

1.2 The set-up of PROLOGUE

Several approaches have been used contributing to attaining the project's main objective:

- Review of previous experiences with ND studies and similar techniques (Backer-Grøndahl et al., 2009)
- Review of technical, methodological and ethical considerations (Welsh et al., 2010; Groenewoud et al., 2010)
- Workshops & newsletters to inform potential stakeholders and get their support (Eichhorn & Winkelbauer; 2010; Eichhorn & Van Schagen, 2011; Eichhorn, 2011)
- Theoretical analyses and questionnaires to identify relevant research questions and interests of these potential users (Sagberg & Backer-Grøndahl, 2010; Van Schagen et al., 2010)

In addition to these activities, five small-scale field trials were conducted in five different countries. These field trials are the focus of the present report. The aim of the trials was twofold:

- To explore the potential and get a better feeling of the technological aspects of ND research, and the strengths and weaknesses of various approaches.
- To illustrate the potential usefulness of naturalistic observations for various aspects of road safety.

The trials were set up in such a way that they differed substantially with respect to technological aspects such as type of equipment, composition of sample, type of cars used, research method, analysis methods, et cetera. The reason for this diversity was that the trials were meant to provide a wide range of experiences with conducting ND studies in practice; experiences associated with, for example, data acquisition, storage capabilities and data download, study design, participants, data reduction, and privacy and ethical issues. Furthermore, the trials varied widely in the research areas and research questions they addressed. The reason here was that the trials were also meant to give an impression of the wide range of application areas of the ND approach, and the capabilities of different technologies and methodologies to address various research questions.

The trials and technologies were chosen based on prior experience and proven expertise of the partners. It is important to note that due to time and budget constraints, all trials were small-scale trials in order to prove feasibility of the approach presented and provide input to the discussion on trade-offs to be considered when planning a large-scale study.

Eventually, all project activities, including the theoretical and literature work, the communication with stakeholders, the experiences during the field trials, as well as advanced insights and experiences elsewhere in the world, fed into the recommendations for a large-scale ND study in Europe (Sagberg et al., 2011).

1.3 The present report

The present report summarizes, integrates and discusses the experiences gained by the small-scale trials. First, the report briefly presents each of the individual field trials (Chapter 2) giving not only an indication of the technologies used, but also of the variety of research questions that can be addressed by an ND type of research. Then the report discusses the experiences related to the technologies (Chapter 3) and the experiences with methodological and organizational issues (Chapter 4). Chapter 5, finally, sums up the main conclusions that could be derived from the trials and the lessons learned.

2 Summary of PROLOGUE field trials

2.1 General overview

As described in the previous Section, the five PROLOGUE field trials were set up differently and addressed different research questions with the aim to ensure a wide diversity of experiences. Some elements however were addressed in all or some field trials, allowing for specific comparisons. This is shown in Table 2.1 where the rows correspond with each of the field trials, the country and its main area of interest; and the columns correspond with the main research issues that were addressed.

All field trials dealt with the issues of data collection and storage, and data reduction and interpretation. In two of the trials two independent data collection systems were integrated. While the Israeli trial integrated two off-the-shelf in-vehicle systems measuring two types of different behaviour, the Dutch trial integrated data from an in-vehicle system with site-based continuous data monitoring. Three trials explicitly looked at the relationship between ND information about driving behaviour and self-reports, and the same three also applied the ND approach to evaluate a particular intervention. On all other aspects the trials differed. Some of the trials were set up such that some specific comparisons among trials regarding the applied technologies were possible.

As the trials were small-scale, they were not meant to provide sound scientific conclusions about road safety behaviour. The samples were small and often not representative for drawing significant conclusions. When discussing results in the following paragraphs, it must be kept in mind that they are just meant as an illustration.

Table 2.1 Overview of the main elements of the PROLOGUE field trials

	Data collection and storage	Data reduction & interpretation	System integration	Relation ND data & self reports	Evaluation of interventions	Area of interest
Israel (IL)	x	x	x	x	x	Novice drivers
Austria (AT)	x	x		x	x	Driver training
Netherlands (NL)	x	x	x			Cyclists
Spain (ES)	x	x				In-vehicle information systems
Greece (GR)	x	x		x	x	Active safety technology

It should be noted that most of the trials were not ND studies in the strictest way in the sense that they collected data on normal driving behaviour without any intervention. Rather some trials, as part of the experimental design, included an intervention phase. For example, in the Austrian and Israeli trial the naturalistically collected data were also used to provide feedback to the drivers and, subsequently, it was studied whether that type of feedback affected driving behaviour. In the Greek trial, the ND data was used to asses the effects of an intervention by looking at the naturalistic behaviour before and after this intervention. These trials showed that an ND approach can also be used for evaluation purposes. In particular, when the intervention concerns the implementation of an in-car function or system, the ND study comes very close to the aims of a Field Operational Test (FESTA, 2008) with an ND type of data collection. However, all trials included naturalistic data collection phase in which no intervention or feedback to drivers were administered.

Collecting naturalistic type of data has great potential, as demonstrated in the PROLOGUE project as well as other naturalistic studies conducted world-wide. However, these studies are very expensive compared to traditional data collection methods (mainly – self-reports). Three of the PROLOGUE field trials therefore addressed the complementary of information from self-reports. These self-reports included self-evaluation of exposure, driving behaviour as well as safety related parameters. The trials provide a glimpse into the type and extent of differences between the ND and the self-report methodologies, as well as to the methodology for measuring these differences.

The next paragraphs provide a brief summary of each of the five trials, providing information about background and objectives, technology used, methods, results, and relevant lessons learned as identified by the researchers of that trial. Please note that these summaries do not provide detailed technical descriptions. For this information and all other details we refer to the Deliverable of the individual trials:

- For the Israeli field trial: D3.2 (Lotan et al., 2010)
- For the Austrian field trial: D3.3 (Gatscha et al., 2010)
- For the Dutch field trial: D3.4 (Christoph et al., 2010)
- For the Spanish field trial: D3.5 (Valero-Mora et al., 2010)
- For the Greek field trial: D3.6 (Touliou & Margaritis, 2010)

2.2 The Israeli field trial

The Israeli field trial consisted of four sub-trials which mainly focused on young, novice drivers. They also had in common that they all used an off-the-shelf G-based data acquisition system (DAS), called the GreenRoad (GR) system. This system is relatively cheap and easy to use, both in terms of installation and in data processing and handling. The system does not include video cameras, but it does include GPS data. The various sub-trials were meant to demonstrate the potential usefulness of such a relatively simple DAS for providing ND information and to assess the possibility and value of integrating this system with a vision-based DAS and a Geographic Information System (GIS) database. Most sub-trials collected new data; other sub-trials (partly) used larger data sets of previous studies for analysis and comparisons.

The GR system has been developed by GreenRoad Technologies and collects data of all trips made by the vehicle. Each driver has to identify him or herself with a magnetic key. The system provides four layers of data collection and analysis:

- Measurement: Raw data on speed and position.

- Identification: Manoeuvres, i.e. braking, accelerating, turn handling, lane handling and speeding.
- Analysis: Driving risk indices of individual drivers, based on the identified manoeuvres (called events). Based on these risk indices, drivers can be classified as cautious, moderate, or aggressive.
- Reporting: Based on the information gathered in the previous layers, feedback can be provided to the drivers. This can be done
    - in real-time, through in-car display or warnings on aggressive behaviour or significant deviations from "normal driving patterns",
    - via text messages to the drivers or others (fleet managers, parents etc.),
    - by reports that summarize and compare the driver, vehicle, or fleet information to be viewed as printed reports or on a dedicated website.

The next four paragraphs briefly describe each of the Israeli sub-trials.

2.2.1 Integration of G-based and video-based event-triggered DAS

Objectives:

The main objectives of this sub-trial were:

1. To explore the technical integration of the standard G-based DAS and a video-based event-triggered DAS.
2. To identify the added value of this integration for providing a broader understanding of road user behaviour in risky situations.

Technology:

The described GR system was used in conjunction with a video-based event-triggered DAS, developed by Mobile Eye (ME). The ME system was developed as a vision-based time headway, forward collision, and lane departure warning system. The required data is collected by a single camera. The camera is integrated with a processing unit that is located on the windshield inside the vehicle. The system captures the outside of the vehicle and the interaction with other four-wheeled cars and road markings.

In this trial the two DASs were integrated. Since the ME system was designed as a warning system only, this required the construction of a system that actually recorded, transmitted, and stored the collected data.

Method:

Ten employees of the road safety organisation Or Yarok participated as drivers in the field trials. Therefore, the sample is probably biased with regard to driving behaviour. The participants drove the equipped car for four months.

Results:

Regarding the ability to extract ND information, the results showed that meaningful information can be extracted both on driver level (driving behaviour) and trip level (risky events) using the two integrated DASs.

Regarding the integration of the two systems, the results showed that the systems provided different information about exposure (trip times and durations). One reason was that the two systems used different clocks; another reason was that the two systems used different predefined algorithms to detect trip starts and trip ends.

The two systems addressed different types of 'events'. The GR system looks at vehicle-based events (e.g. braking and accelerating), whereas the ME system looks at events in interaction with other vehicles and the road. One consequence is that both systems resulted in different numbers of 'events'. Another consequence is that both systems provided different information about (proxies of) near crashes. For example, during this trial the GR system classified seven events at the highest severity level; none of these events coincided with an ME event. It can be concluded that when using two systems in combination, the definition of thresholds for triggering events or near crashes has to be jointly calibrated.

Lessons learned:

The DASs used in this trial do not include continuous video-recordings, and consequently they do no provide visual information about the drivers' behaviour nor the surroundings. That is a weakness. On the other hand, the systems are relatively cheap, they are quick and easy to install, and the data provided is relatively easy to handle in terms of transfer, storage, reduction, et cetera.

The trial demonstrated that integrating two systems requires a large amount of effort and would require joint calibration if comparable data is required. It is emphasised that it would be a huge advantage to use an already integrated and analysis-ready system.

2.2.2 Spatial analysis of novice drivers' driving behaviour

Objective:

The main objective of the second sub-trial was to investigate the usefulness of the GPS location measurements and a map-matching tool for getting information about geographical and spatial aspects of driving and the locations of safety-relevant events.

Technology:

Again the GR system was used, including its GPS location measurements. For processing the behaviour data and the corresponding locations a map-matching utility was developed that linked the collected data to the geographic features that are available in a Geographic Information System (GIS) database.

Method:

The analyses were performed on data that was collected in a previous study. This concerned data of almost ninety young drivers and their parents. The analyses included:

- Travel distance and time; total and per road class.
- Comparison of the exposure to different road types between young drivers and their parents.
- Comparison of safety-relevant events of young drivers and their parents per road class.

Results:

This sub-trial showed that it was possible to link the ND data to a GIS database and as such to get information about the spatial distribution of exposure and behaviour, e.g. in relation to different road categories. One example of the results generated in this trial was the finding that, accounting for distance travelled by road type, the parents generated a higher rate of speeding events than their children, the young drivers. However, linking these speed events to the location, it was found that parents speeded more often on non-urban motorways, i.e. roads designed for high speed. The young drivers on the other hand, speeded more often on non-urban arterial road and urban highways, i.e. roads that were designed for lower speeds.

Lessons learned:

This trial showed that it was possible to link driver behaviour and events to specific locations, and to get meaningful results from it, also when comparing different driver groups. One problem with regard to linking the DAS and the GIS database relates to the time resolution of the GPS information. In the current trial the resolution was insufficient to identify travel routes for short trips and trips with missing information. A higher resolution is possible, but would increase computational and communication demands. The challenge is to find the right balance between resolution and system demands.

2.2.3 Experienced young drivers, and ND versus self-reported data

Objective:

The objective of the third Israeli sub-trial was related to driving behaviour of experienced young drivers and the relationship between ND and self-reported data.

Technology:

Driving behaviour and event rates were measured by the GR system and by means of self-reporting.

Method:

The data came from just over thirty participants. The participants were young drivers who were recruited from a sample that had participated in a previous study during their first year of driving. This additionally allowed for comparing their behaviour in the first year and after 2 to 4 years with a driving license. Data was collected for 8 months.

During the study, the participants received web-based questionnaires via email. The questionnaires consisted of two parts. Part 1 asked for information about exposure (e.g. number and length of trips) of the last 48 hours, which was also available from, and hence comparable to the DAS information. The second part of the questionnaire included questions on driving behaviour and safety-related events.

Results:

Substantial differences were found between the DAS data and the self-reported safety-related events. The participating young drivers perceived themselves as safer drivers than they were according to the more objective DAS data. As for exposure, however, the participants were quite accurate in their reports.

Comparing the ND data in the first year of licensing with the ND data two to four years later, the results show that, overall, the rate of safety-related events had not changed and that the driving patterns had remained the same. However, a closer look at the characteristics of the events revealed increased event rates during the weekend.

Lessons learned:

This trial showed that self-reported information about driving behaviour seems to over-estimate the quality of the driving and its safety level. Hence, ND data collection has added value. Though the current trial indicated that information about exposure could be based on self-reports, it must be kept in mind that the participants were only asked about the past two days. It is likely that longer recollection periods would result in less accurate information. Again it was shown that ND provides useful and meaningful information about driving behaviour, also when applying a within-subjects design comparing behaviour of the same persons over time.

2.2.4 Between and within cultural aspects of driving behaviour

Objectives:

The fourth and last sub-trial in Israel looked at yet another application area of ND research, i.e. assessing differences and similarities in driving behaviour of drivers of the same culture (within) and drivers of different cultures (between).

Technology:

The GR system was used.

Method:

For the within cultural aspects, data from 11 young novice drivers, all living in the same Christian Arab local community, was used and compared to data from trips performed by other drivers of the family cars. To get an indication of the actual social and family relationships, all participants were also interviewed (e.g. who is your best friend; with whom do you drive?). For the between cultural differences and similarities the data of the 11 young Arabs were compared to data of 11 young Jews from the second sub-trial (Paragraph 2.2.2).

In addition, the feasibility of using different DASs (the GR system and the pdrive system® (see Section 2.3)) for collecting comparable data in different countries was assessed by conducting two test drives with a professional driver who deliberately drove in an aggressive way - under safe conditions -, with both systems on-board.

Results:

Within culture: Based on stated friendships and shared driving between the participants, a social network was mapped and the driving behaviour of the identified subgroups was compared. No effect was found of this type of social relationships on driver behaviour. However, the driving behaviour of the two outliers of the group, i.e. the "best" and the "worse" driver in terms of safety scores, revealed that the most “popular” person in the group was also the best driver, whereas the least popular person was the worst driver. This finding can contribute to understanding the good driving norms of a group.

Between cultures: Comparisons of the behaviour of the young Arabs and the young Jews indicated that the groups differ with regard to driving exposure and event rates.

The group of Arabs drove more often, drove longer and had more safety-related events than the young Jews. However, looking at the safety-related events, individual differences were larger amongst the Arab youngsters. One explanation for this difference between the two groups is that in the Jewish group the feedback information had also been made available to the parents, whereas this had not been the case in the Arab group since this group refused.

First attempts to integrate and calibrate two different DASs in such a way that they would provide comparable data showed that this is possible, but taking relatively large efforts.

Lessons learned:

This trial showed the potential of ND data when looking at differences and similarities within and between social and/or cultural groups. However, when using different DASs, much time has to be reserved for integration and calibration to make the data comparable.

2.3 Austrian field trial

Field of interest:

The Austrian field trial addressed the driving behaviour of young drivers immediately after they had received their licence. More specifically, it looked at the impact of video-based feedback during the practical driving lessons on the behaviour during the first two months after licensing.

Technical objectives:

The technical objectives of this trial were to assess

1. The usefulness of ND data of basic driving parameters such as acceleration and speed for deriving risk indicators, and, from there, risk scales.
2. The usefulness of ND video data for providing (visual) feedback during driving lessons.
3. The comparability of ND data and self-reported data.

Technology:

The DAS that was used in this trial was the pdrive system®. This in-vehicle data recorder system records various g-based data, including speed, acceleration, deceleration, and steering manoeuvres. The system also records the GPS signal and it has two digital cameras, one directed towards the driver, one to the situation in front of the car. Since continuous video recording would result in too much data for the system in question to handle, the video recordings were event-triggered, i.e. only images of a short period before and after a predefined event, such as sudden harsh braking, were actually registered.

The system is powered from its own internal rechargeable battery or from the vehicles power supply. Data and video are stored on a memory card.

Identification of the driver was done by a video clip of the driver 60 seconds after the trip start.

Method:

In the study 12 novice drivers participated, half of them had had video-based feedback during the practical driving lessons, the other half had not. Thus, this specific field trial included an intervention/experimental condition.

All drivers were 18 years old, and in the first months of solo driving after having obtained their driver licence. The DAS remained installed in the participants' cars for 1 to 3 months. The participants had to complete on-line trip diaries to get self-reported information about trips and events.

Pre-processing routines were developed for getting the raw data from the data acquisition system into an SQL database. Three pre-processing steps where performed manually to achieve this, i.e. checking raw trip data, identification of test subjects and mapping trips on a road map.

Results:

Database queries were applied to extract segments of trips. Various threshold levels of lateral and longitudinal acceleration were used to classify type and severity levels of safety-related events per trip segment. Frequencies of these event severity levels per hour driven were used as risk scales for group comparisons.

The results indicate that in urban areas, the group without video-based feedback during the training had more safety-related events per hour driven than the group with video-based feedback. On motorways there were no indications of differences in this respect, though, when looking at the risk scales, again, the group without feedback scored worse than the group with feedback. The study also showed various differences between male and female drivers.

Lessons learned:

One contribution of the Austrian trial, besides testing and proving the usefulness of the applied DAS, was that it showed the potential of using ND video data for providing feedback to drivers.

Furthermore, the trial showed that the ND data is suitable for identifying risk indicators that could be combined in a risk scale. However, data reliability remained a problem and would need further attention.

The Austrian trial used event-triggered video data that were analyzed manually. This turned out to be very time-consuming.

As for the technology, originally there was a problem with the autostart and autostop functionality of the system, resulting in empty batteries. Reliable power supply for the DAS is important.

Finally, there were some problems with getting the participants to show up for transfer of data, resulting in full data storage cards and loss of data. Data transfer by WiFi could be solution for this. Nevertheless, it shows the importance of gaining and maintaining participants' loyalty and commitment to the study.

2.4 Dutch field trial

Field of interest

The Dutch field trial looked at the interaction between cars and cyclists at an intersection with a separate cycle track with a potential conflict between car drivers turning right and cyclists going straight on.

Technical objectives

The trial aimed at

1. Assessing the added value of combining naturalistic in-vehicle data and site-based information for studying vehicle-bicycle interactions.
2. Comparing and validating speed measurements from the site-based and the in-vehicle study.
3. Comparing and validating manual (image-based) analysis and automatic (video-based) analysis of the site-based data.

In addition, each of the observation methods served as a separate study in itself.

Technology:

In-vehicle observations: The same DAS as in the Austrian field trial was used, i.e. the pdrive system®. The system continuously recorded GPS data, and GPS derived variables like speed and acceleration. The system also registered video data, triggered by either the occurrence of a safety-related event or by GPS, i.e. when entering the intersection with site-based cameras. As in the Austrian trial, the pdrive system had two cameras: one facing the driver and one facing the forward road.

Site-based observations: an urban intersection was equipped with two cameras. The two cameras were positioned with different angles to provide a broad picture of the area of interest.

Method:

The two fixed cameras were placed at a four-armed traffic-light regulated intersection in an urban area with a separate cycle path. Since cars and cyclists get green light at the same time, there are potential conflicts between right turning cars and straight-on going bicycles. The fixed cameras were active for two weeks and were only switched on when some movement was detected. For the in-vehicle part of the trial, eight cars of people regularly commuting through this particular intersection were equipped with the DAS. The DAS remained in the car for eight weeks. The equipped cars were marked with a big white “dot” on the roof to enable matching the site-based and in-vehicle information. Detailed data analysis of the site-based data set focused on a randomly selected one day period from 06:00 am until 08:00 pm.

Results:

The information from the site-based, high-mounted cameras showed that the non-halted encounters (arriving when the light is green) are relatively more critical than the halted encounters (accelerating after the red light turns green). The in-vehicle measurements showed that drivers look more and longer in the direction of a potential cyclist in a halted situation than in a non-halted situation. Moreover, the fixed camera data made it possible to define the exact position of both the car and the bicycle enabling, for example, quantifying the time-to-collision. The in-vehicle data, on the other hand, provided more explanatory data. So, both methods yield complementary information about car-bicycle interactions.

Regarding the validation of information obtained by the two observation methods, the results showed that quite similar speed profiles were obtained by the site-based (manually derived speed profiles) and the in-vehicle GPS-based data. Even though there were some differences, the overall similarity indicates that the two measures are valid.

Furthermore, based on the site-based material the manual processing appeared to be more accurate than the automatic processing, but demanded relatively more in terms of processing effort. The automatic processing on the other hand, was less accurate but much faster.

Lessons learned:

First of all, this trial illustrated the added value of combining in-vehicle and site-based information when studying the interaction between cars and bicycles, since both methods yielded different but complementary behavioural information. Similar added value is also likely to be found when studying the relationship between driver behaviour and road design or when studying various driver aspects related to traffic flow and traffic management.

Automatic analysis of video data is to be preferred from an efficiency point of view. However, the accuracy requires attention. One problem with automatic processing, as identified in this trial, is the difficulty of identifying different road users. A top view camera or more cameras with more viewpoints might solve this problem.

Technically, this trial encountered the same problem with the power supply of the DAS as in the Austrian trial. Furthermore, since data needs to be retrieved from the vehicle on a regular basis, it is important that the process is easy and quick in order not to burden the participants unnecessarily. Finally, there were some problems with the quality of the data due to small defects. This illustrates the need for regular quality checks of the data.

2.5 Spanish field trial

Field of interest

The Spanish field trial focused on the effects of a navigation system on road user behaviour.

Technical objectives:

This trial was conducted with an instrumented car, the ARGOS car. The two main objectives were:

1. Test the suitability of a highly instrumented car for ND research.
2. To compare G-based and video-based identified safety-related events or incidents.

Technology:

The ARGOS car is highly instrumented allowing for recording a wide range of variables. These variables relate to:

- Dynamics of the car (e.g., distance travelled, lateral and frontal position, speed),

- Driver vehicle interaction (e.g., steering wheel rotation, steering wheel rotation speed, brake pressure),
- Comfort systems for the driver (e.g., window control, parking assistant, internal lights),
- Indicators in the car (e.g., indicator of ABS activation, oil temperature, water temperature, level of fuel),
- Environmental conditions (e.g., outside and internal temperature, interior noise),
- Data about the driver (eye-movement measurement),
- Video recordings (7 video cameras covering the inside and outside of the car).

The ARGOS car also has the potential for additional data collection that were not applied in this field trial: GPS, sound and noise registrations, recordings of lateral position, and eye tracking.

Method:

Five drivers participated in the study. They drove the instrumented ARGOS car rather than their own car which actually made this trial not really ND. The participants drove the instrumented car for four consecutive days without an experimenter being present. This made the trial as “natural” as possible. During these four days, the participant was instructed to drive to different destinations, finding a number of specific addresses in various parts of the city of Valencia and its direct surrounding. Drivers were free to plan their route, using either the on-board navigation system or a map. The participants were allowed to take a break whenever they wanted, but due to power limitations in the system’s batteries, they had to return within three hours.

Data was registered continuously during each trip, and was downloaded from the car daily after each journey.

Results:

With regard to the first objective, testing suitability of in instrumented car like the ARGOS car for ND studies, the results were mixed. One obvious problem is the lack of “naturalistic” driving given the highly instrumented car. Obviously, such an instrumented car can record a wide range of interesting driving parameters. However, at the same time, this large amount of recordings set high requirements concerning on-board data storage and download frequency that seem hardly feasible in a real naturalistic study.

An interesting aspect of this trial was the validation of the identification of safety-related events or incidents based on vehicle parameters. It was found that this method resulted in many false alarms: when looking at the video data of these events, often nothing special could be detected. In addition, after incidents were identified using video data, vehicle and driving parameters were investigated for these events.

A total of 16 incidents were recorded during the trial period. Out of these, six were crash relevant conflicts, one unintentional lane deviation and nine illegal manoeuvres. Out of these 16 incidents, 11 happened while the driver did not use the navigation system or any other nomadic device, e.g. a mobile phone. It must be noted, however, that the participants hardly used the navigation system at all while driving. They normally stopped the car to do so or to use the mobile phone. This may be due to the non-naturalistic nature of the car and the trial.

Lessons learned:

The main lesson learned concerning the technical equipment, is that it was not completely reliable, and that quality checks of the data must be performed as often as possible, for example, after half a day. As for the video cameras, the authors suggest a better solution for the position of the various cameras.

Regarding analysis, manual coding of events/incidents by investigating video was necessary. This was time consuming.

Based on the prevalence of incidents in this trial, it was estimated that in order to get 100 incidents, one would need about 250 hours of recorded driving, and between 100 and 120 days of field trials.

The main result is that the identification of incidents cannot only be based on the numerical parameters of the car. This would result in many false alarms while at the same time many true alarms would remain undetected. Even though the data collected in this trial might be the basis for further refinement of the criteria used for identifying incidents, detailed examination of the parameters suggests that the patterns could be very complex and that further research is needed. Besides, interesting incidents are also those where the drivers do not react at all because they are distracted, and consequently there is no evidence from the vehicle parameters of this behaviour.

2.6 Greek field trial

Field of interest

The field of interest of the Greek field trial was the effect of two advanced driver assistance systems: Forward Collision Warning (FCW) and Lane Departure Warning (LDW). Among other things it was studied whether people compensate for the increased safety associated with the warning systems by driving in a more risky manner, and whether the systems affected the prevalence of distraction.

Technical objectives:

Like the Spanish trial, also the Greek trial used an instrumented car in a relatively naturalistic setting. The main objective of this trial was testing the usefulness and feasibility to extract particular meaningful results out of the various vehicle and video data and to define instrumentation priorities for future ND studies.

Technology:

The main vehicle parameters recorded in this trial were lane deviation, indicator activation, steering angle, brake activation, time headway, and acceleration. In addition, cameras were installed in the vehicle allowing for recording the driver, as well as the forward and backward road view. The system also recorded GPS.

Method:

Five drivers participated in the trial, and each of them drove the instrumented car for three consecutive weeks. Each day of this three weeks period, the participants had to drive a pre-defined, familiar route of around 40 minutes, including peri-urban, rural and highway sections. In order to investigate the research questions, a quasi-experiment was conducted with different conditions in each of the three weeks, in a fixed order:

- Week 1: baseline; no systems on, but warnings recorded by CAN-bus
- Week 2: FCW activated
- Week 3: LDW activated

The participants were acquainted with new technologies. In addition to the vehicle data, participants were also asked to complete three types self I reports/questionnaires: the Driver Behaviour Questionnaire (DBQ), a System Usability Scale (SUS), and a daily driving diary.

Data was downloaded on a daily basis, both as a quality assurance precaution and because of hard drive memory restrictions.

Results:

During this trial, drivers were hardly involved in behaviour or actions that were not primarily related to operating the car. When FCW and LDW were activated the number of secondary behaviours was somewhat higher. However, this could be a familiarity effect as the baseline condition was always the first week. It was also found that when the driver looked at the scenery while driving, this coincided with larger speed variation and larger steering wheel movements.

From the self-reported data it could be concluded that the participants in this trial belonged to the category of what the DBQ qualifies as “safe” drivers. The participants assessed the usability of the assistance systems above average. From the trip diaries, it became clear that the participants attributed the self-reported events to other road users rather than to themselves. An example of a self-reported event is close following by another vehicle.

Lessons learned:

The Greek field trial showed that self-reported trip data and questionnaire data are useful additional instruments to reveal biases in the sample. They are also useful to uncover particular biases towards participants' own interpretation of their behaviour and the traffic situation.

Regarding vehicle and video data collection, it was first of all, and again, concluded that manual processing of video data is very time-consuming. In addition, it was concluded that the number of parameters should be decreased to the ones that are found most promising. The subsequent data reduction is best to involve more than just one person, in order to guarantee a sufficient level of reliability.

3 Discussion of the technological experiences

As shown in Chapter 2, the five field trials varied substantially in the technology used. This Chapter integrates and discusses the experiences with data acquisition, data storage and data reduction. The trials also varied on all sorts of methodological and organizational. These experiences are integrated and discussed in Chapter 4.

Both Chapter 3 and Chapter 4 aim to contribute to the overall aim of the PROLOGUE project, i.e. assessing the feasibility and usefulness of conducting a large-scale ND study in Europe. An important focus is the way the various factors contribute to or limit the reliability of the results, as well as both internal and external validity.

As a general note, the discussions first and foremost focus on the experiences – strengths and weaknesses – that need to be taken into account when conducting further ND studies. “Lessons learned” do not necessarily constitute problems experienced during the field trials. Rather, these are issues or factors that have been brought to our attention because of the research questions addressed, the possibilities or limitations of the technologies, samples, study designs, et cetera, and that are important to take into account for a large-scale trial.

3.1 Overview of the applied technologies

One prerequisite for conducting an ND study is to have a technological system that unobtrusively records a minimum of driving-related parameters. The system also has to be able to store and transfer the data to a data reduction tool, so that the collected data become eligible for analyses. Various such technological systems, data acquisition systems (DASs), exist with different possibilities as for what can be measured and recorded. In the reported field trials, the following DASs were used:

- Green Road (GR) (Israel)
- Mobile Eye (ME) (Israel)
- Pdrive system (Austria and the Netherlands)
- Instrumented car (Spain and Greece)

In addition to these vehicle-based DASs, the Dutch trial included a site-based observation with stationary cameras.

One way to categorise the technological systems, is according to their level of complexity. From least to most complex, the categorization starts with DASs that just record the basic driving parameters, followed by DASs that also include event-triggered video data, DASs that continuously record video data, and finally DASs that record specific additional data, such as eye tracking information. This is illustrated in Figure 3.1, specifying the PROLOGUE DASs that fell within those categories.

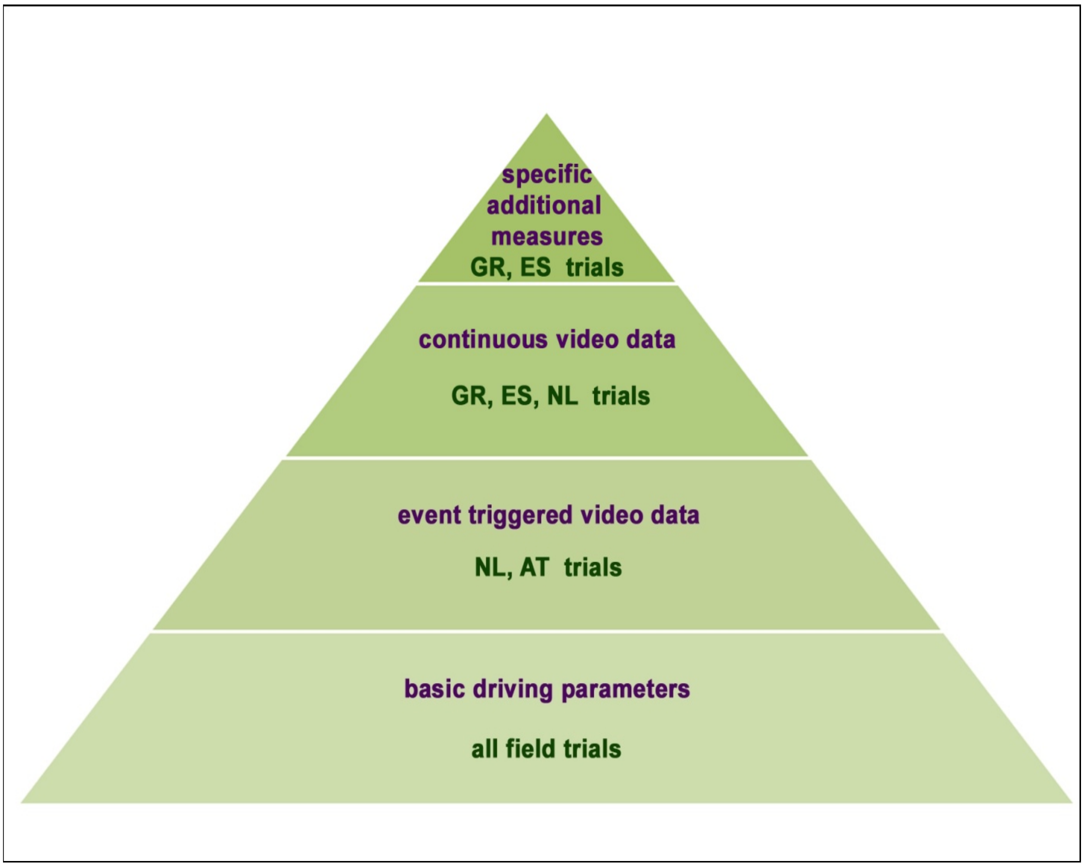

Figure 3.1: Level of complexity of the technologies

3.2 Basic driving parameters

All field trials used a DAS that at least recorded the basic g-based driving parameters such as acceleration, deceleration and speed. Based on this type of data, it is possible to define specific thresholds to identify safety-relevant events. Based on frequency and severity of events, it is possible to, for example, identify drivers with different driving styles. This was done in the Israeli field trials by distinguishing between “cautious”, “moderate”, and “aggressive” drivers. The Austrian trial used frequency and severity of events to determine the risk profile of a driver. Moreover, based on the severity of events, it is also possible to define a near-crash which in turn can be considered as proxy for an actual crash. A first step towards identifying such proxies was made in the Israeli trial. A general consideration here is that there is not yet scientific consensus to what extent the conditions of a near crash are comparable to the conditions of a real crash. Finally, the g-based raw data can be used to study behavioural differences over time or over locations, as was done for example in the Greek trial by looking at the differences in steering wheel position and speed when the driver was either or not involved in a secondary task.

In conclusion, research that provides information about the basic g-based driving parameters allows for various meaningful analyses. One important question that needs to be answered here is which thresholds must be applied to define an event at several levels of severity. In the current field trials, only the Israeli Green Road system had predefined thresholds for events, based on previous experiences. In the other field trials, the researchers had to explore themselves where to set the thresholds. As was demonstrated in the Spanish field trial, identification of events by just g-forces may result in false alarms, i.e. identified events that were actually not an event. Additional inspection of qualitative video data was found to help to avoid on the one hand false alarms and on the other overlooking events that are not triggered by the thresholds. In addition, video data also allows for the identification of events where the driver does not undertake any action where (s)he actually should have done. Self-reported data seem not to be a good validation of the g-based events, since the Greek trial found that drivers tend to only identify events that were, according to these drivers, caused by another road user.

3.3 Video and specific additional data

In addition to the basic g-force data, some of the trials also collected video data, either event-triggered (the Dutch and Austrians trial) or continuous (the Spanish and Greek trial).

Video data is essential if one wants to investigate the behaviour of the driver that cannot be derived from the basic driving parameters. For example, if one wants to investigate secondary behaviours, looking behaviour/glances of the driver etc., video data of the driver is essential. Using naturalistic video data is thus particularly promising with regard to studying driver distraction, inattention, and fatigue. Whereas these have previously been studied by means of controlled experiments or self-reported data with inherent limitations such as low external validity and subjective biases, naturalistic video data opens up the possibility for recording objective and direct measures of, for example, secondary behaviours performed by the driver.

In the Spanish field trial, the continuous video data was of particular relevance in order to identify events, as the g-force triggers turned out to produce many false alarms as well as failing to identify relevant events. The video data was thus inspected to identify safety-relevant events, and the g-force based driving parameters were analysed for these events afterwards. Ideally, it should be the other way around; triggers indicate that there is a safety-relevant event and the video data of the driver and the context can be investigated and analysed for the events in question.

Either way, continuous video data demands a large amount of time and effort when performing manual inspections and analysis. This issue was also discussed during the naturalistic data workshop with the Virginia Tech Transportation Institute (VTTI) in Tel Aviv in June 2010. VTTI gained much experience with this issue during the 100-car ND study in the USA. A unanimous opinion of the PROLOGUE field trial researchers was that coding and analysing large amounts of video data requires at least partly automatic processing. As this is emphasised based on the experiences with the relatively small-scale field trials, the argument for automatic analysis tools in larger ND studies is strong. However, it is important to recognize that automatic analysis may not be as accurate as manual analysis, as was also found in the Dutch field trial. An automatic video analysis tool needs to be validated in order to assure accurate and reliable identification of the measures that one is interested in. Of course, reliability needs to be assured when applying manual analysis as well, and at least two independent observers should go through the material. However, it is worth noting the experiences of other researchers in this regard, and a small discussion on the issue from SHRP2 moderates the perspective:

> "To answer a number of questions, manually reviewing samples of video data will be required regardless of the degree of sensors/automation attained [……]. This is because an understanding of the larger driving context and/or sequence of driving behaviours is extremely valuable. Despite initial reactions to the contrary, our experience has shown that this is neither particularly daunting nor expensive in the overall scope of the project."[1]

[1] SHRP2, S05 Task 1: Development and Prioritization of Research Questions: Preliminary results.

One challenge in the field of driver distraction and inattention has been to estimate actual odds ratios or relative risk for various risk factors, partly because it is difficult to estimate exposure to the risk factors, for instance how much of the driving time is spent while engaged with secondary tasks, such as eating, smoking, and taking care of passengers. Using continuous video data, one can estimate exposure to the risk factor in question by investigating random time slots in the video data (Klauer et al., 2006). Odds ratios can be estimated based on such exposure measures and the presence of the risk factor in question in safety-relevant events. However, continuous video data demands a large storage capacity and/or possibility to transfer data from the DAS on a daily basis. For instance the pdrive system does not currently have the required capacity. As an alternative to continuous video data, it is also possible to use event-triggered video, so that the video only starts recording in case of a safety-related event (triggered by predefined thresholds of particular g-based data) or at specific times or locations (triggered by GPS). Although event-triggered video does not allow for estimation of relative risk, this kind of data is important for understanding what happens in or just before specific situations. An example of a location-triggered video application was the Dutch trial, where the behaviour of drivers at a particular intersection was studied. An example of an event-triggered video application was the Greek trial.

Another potential of naturalistic video data tested in the Austrian field trial, is for monitoring and advising novice drivers in their learner period. Using naturalistic video data as an educational tool involves the driving teacher going through the video data with the driving pupil. Thus, the pupils can see their own driving behaviour in various situations, which in turn is assumed to be of driving educational benefit.

In the Dutch trial video data was used, among other things, to investigate the visual behaviour of drivers in a situation where a cyclist may be present. The video data showed the face of the drivers which allowed for the identification of both direct glances (with head and/or body movement) and indirect glances (only eyes). As discussed by Welsh et al. (2010), this method is relatively cheap and “simple” and has been used to great effect in previous studies – and supported again by the Dutch field trial. A more advanced method is the use of eye tracking systems which, for example, include the measurement of exact fixation points as well as sleepiness metrics (Welsh et al., 2010). Advanced eye-tracking systems designed for in-vehicle use are relatively rare and consequently expensive. Nevertheless, using such equipment in an ND study could be of great importance for development of for instance the knowledge field of driver fatigue.

Finally, requirements regarding the number and exact location of the on-board cameras depend on the research questions that are investigated and the available time and budget for analysis. This became clear from the Spanish trial where five out of seven cameras in the ARGOS car, were operational. In principle, these five cameras must be able to capture the most important information. However, based on experiences with both the cameras capturing the outside of the vehicle and the cameras capturing the face and the arms/hands/lap of the driver, the researchers proposed more cameras and still a better positioning of the cameras.

3.4 Level of complexity of the DAS

Three of the trials – the Israeli, the Austrian and the Dutch – involved private cars equipped with a DAS for a limited period of time. This is in line with the definition of an ND study, i.e. including unobtrusively observing normal drivers in their normal driving context while driving their own vehicles (Backer-Grøndahl, 2009).

Both the Spanish and the Greek field trial made use of a highly instrumented car, and all participants drove the same car. Thus, according to the definition above, these trials cannot be denoted as purely naturalistic, even though in both trials the car was used in a more or less naturalistic way. However, since the instrumentation of the Spanish and the Greek cars was far more complex than the DASs in the other trials, it gave an opportunity to assess the usefulness of this type of high-end DAS.

It is a common notion in social science that there often is a trade-off to make between internal and external validity. It can be argued that the internal validity, in terms of determining causal effects, is higher when more varied data can be collected, and advanced analyses can be performed which is the case for the high-end DAS. This allows for more control of potential confounding variables and in turn for a better basis for inferring causal relationship between variables.

On the other hand, various aspects of using a highly instrumented car limit the external validity, i.e. the generalizability of the results. For instance, the participants are not used to the car and the instrumentation of the car they have to drive, which makes the driving not “naturalistic”. It is possible to counter this effect to some degree by letting the participants drive the car for some time before the study starts, in order to familiarise them with the car and the instrumentation. Behaving in a naturalistic way may further be threatened if the equipment is visible for the drivers. If the equipment is visible for the drivers of other cars, this may in turn affect the way they respond to the instrumented car. In the Spanish field trial, the latter problem was solved by tinting the windows of the backseat of the instrumented car.

Another limitation with using highly instrumented cars relates to the economic costs. Highly instrumented cars are very expensive and that limits the number of cars that can be instrumented. As a consequence participants cannot use their own cars. Moreover, the more extended and advanced the equipment, the more time and resources of skilled personnel is needed for installation. Moreover, when using highly instrumented cars, it is important to consider legal issues and insurance conditions. The cost of the car and insurance issues also restrict where the cars can be stored, and in both the Spanish and the Greek field trial the cars had to be returned to the researchers’ facilities each day. In this respect, it is important to consider the rapid technological development taking place, and that what is considered complex, rare and expensive technological equipment today, may be standard and cheap in a couple of years or even months.

For the time being, however, it must be concluded that instrumented cars may provide very useful extra information as compared to the more simple DASs, but that the fact that participants cannot drive their own cars and have to return the car to a specific place each day, substantially limit the 'naturalistic' aspects of the studies.

On the other end of the scale, quite simple off-the-shelf existing technologies were used in the Israeli trials. These technologies were relatively cheap, easy to install and included a very low level of interference for the driver. Moreover, the technology offered built-in data reduction, and had potential to be used as effective and smart triggers for data reduction of continuous data. Furthermore, data was transmitted in real-time and no manual data download was needed. Thus, data on basic driving parameters was collected, without severe threats to the external validity of the trials, and at a relatively small cost. On the other hand, the technology used in the Israeli field trials does not allow for video recordings which, as described in the previous section, are essential for some research areas. Moreover, when working with number and severity of events as identified by g-based information only, a good validation process preceding the actual data collection is essential.

3.5 Different technologies and integrated systems

If one wants to measure and record various types of behaviour, one option is to invest in one complex and integrated DAS, while there is also the possibility to integrate two or more technologies. For example, in one of the Israeli sub-trials the g-based Green-

Road technology and the video-based MobilEye were integrated, allowing for complementary data. The experiences here showed that integration of different systems is not so easy and may require substantial effort to guarantee consistency of even quite basic but essential information such as time, and start and end of trips.

Rather than integrating two systems with the aim to get *complementary data*, it is also possible to calibrate two different systems with the aim to get *comparable data*, for example in two different countries. The possibilities and difficulties of such an effort was touched upon by a small-scale attempt to make the GreenRoad technology in Israel comparable with the pdrive system® in Austria. The main issue to be solved was to calibrate the two systems in terms of precise manoeuvres and threshold triggers and that appeared to be very difficult and within the current project even impossible.

A third type of integration was performed in the Israeli sub-trial on spatial analysis of young driver's behaviour. Here, mapping behaviour and events to specific locations, more specifically to various types of roads, was performed successfully by linking the DAS and the GIS database. However, there was a problem with the resolution of the GPS location records that made it difficult to identify travel routes for short trips and trips with missing information.

When looking at the integration issue, the overall conclusion is that it is preferable to use one and the same integrated system to collect the required data. Post-hoc integration appears to require much time and effort with uncertain outcomes in terms of reliability and comparability.

3.6 Site-based and in-vehicle observation

Yet another type of integration of systems is the integration of an in-vehicle and site-based study, as realized in the Dutch field trial. Since this is a fairly new approach, it is described in a separate section.

It appeared that the combination of site-based and in-vehicle data provided more information about the situation and the interaction between road users than just one of these sources, for example when studying the interaction between vulnerable road users and cars. In addition, observing behaviour from different perspectives made it possible to validate the data obtained by each of the techniques, e.g. video-based and GPS-based speed data. However, as the Dutch trial showed, various factors need to be considered before conducting such a combined study.

First, the site-based cameras need to be situated at a location that the participants of the in-vehicle study pass regularly. Second, it must be possible to identify the equipped cars on the site-based video, e.g. by marking the vehicles with a white "dot" on the roof. Third, it must be ensured that in-vehicle data collection is operational at the fixed camera sites. For this, in the Dutch trial a location-based event trigger was developed and uploaded to the DAS which ensured that recording of video data started when the driver with the equipped car entered the predefined area from a given direction.

One lesson from the Dutch field trial was that the situations of interest, in this case an interaction between a car and a bicycle, were very rare, and consequently there were very few situations with eligible data for the research questions and analyses that were proposed. Thus, even though one had assured that the participants drove through the section in question on a regular basis, a huge amount of data is needed to get sufficient events from both the site-based cameras and the in-vehicle DAS.

As for the site-based study, analyses were conducted both manually and automatically. Whereas the automatic analysis is greatly advantageous in terms of time, this type of analysis was less accurate than the manual. Obviously, this also applies to the use of automatic analysis of in-vehicle video data as well. Thus, at this time it seems not yet feasible to depend entirely on automatic processing and analysing video data; however, developments in this area are going fast.

3.7 Other technological issues

This Section briefly discusses a few remaining, more specific technological issues that came up during the field trials.

3.7.1 Driver identification

Since a car is often also used by other drivers and not just by the participant of the ND study, it is important to reliably identify the driver. Identification can be done automatically by taking a video shot of the driver immediately after the trip start and manual identification afterwards, as was done in the Dutch and Austrian trial. Some more advanced applications can automatically identify the pictures taken, these, however, were not used in the current trials. In the Israeli trials, the drivers had to identify themselves by a magnetic key at the beginning of each trip, requiring an active behaviour of the participant. The trials showed that automatic identification is preferable. Otherwise, a participant may, consciously or unconsciously, forget to identify himself or someone else may use the identification key resulting in a false identification. The Israeli trial showed that missing or false identifications make it difficult to interpret the results. For instance, in the study on social connections among young drivers, only the young drivers – and not the other drivers in the families – had to identify themselves. A fair amount of the trips were unidentified trips, and there is no way to check whether these unidentified trips were carried out by family members or that the young drivers intentionally failed to identify themselves on specific trips. This could be one explanation of the low event rates in these data.

3.7.2 Power supply

Power supply to the DAS and switching between power modes are also issues that need consideration. Using the pdrive system, both the Austrian and the Dutch in-vehicle field trials experienced power supply problems influencing the overall quality of the data in terms of faulty recordings, and inconveniences to the participants in terms of, for instance, flat vehicle batteries. The problems experienced were mainly caused by inherent properties in the GPS signal which determined the mode of the DAS, and the fact that the DAS was supplied with power directly from the vehicle's starter battery.

3.7.3 Storage capacity and data transfer

A last issue to mention here relates to storage capacity and data transfer. The applied DASs varied in terms of storage capacity and how often data had to be transferred from the DAS to an external hard drive. The storage capacity is related to the complexity of the DAS. In the Israeli trial, data was automatically transmitted on-line and in real time to a central server, and hence no manual data download from the DAS was needed. This, of course, eased the operation and quality checks and provided minimal intrusion to participants. On the other hand, in the Greek and Spanish trials that used highly instrumented cars and registered continuous and varied data, downloading and transfer of data had to be done manually on a daily basis. In the Dutch field trial, data had to be downloaded on a weekly basis, whereas the participants in the Austrian trial could drive for two weeks before data had to be downloaded. For a large-scale study wireless data transfer would be preferable, since this does not burden the participants. For basic parameters state-of-the-art equipment is already able to do so. However, video data, in particular continuous video data, produce very large data files that cannot yet be dealt with through wireless transfer.

Though a larger burden for the participants, a regular download of the data allows for regular quality checks of the data. Quality checks of data are very important in order to ensure the technical system(s) are properly working, and that various triggers are activated at the correct thresholds, times or locations.

4 Discussion of methodological and organizational aspects

Next to the technological aspects the small-scale trials shed some light on various methodological and organizational of ND studies. These aspects are briefly discussed in this Chapter.

4.1 Populations, samples and participants

As the field trials were just feasibility pilot studies with small samples, no conclusions about the issues of sample size and sample representativeness could be drawn.

One experience from the trials that is important to consider when drawing a sample for an ND study, is to assure that the participants in the sample can be expected to drive on a regular basis in the study period. For example, in the one of the Israeli studies, the majority of the participants were in military service when the study took place, and consequently they hardly had an opportunity to use the equipped car at home. Something similar was experienced in the Austrian trial, in which the participants were 18 years old and had to study for the final exam during the study period. Consequently, the participants probably did not drive as often as they normally would. In addition, the study period included the summer holiday and a school trip, both factors likely to influence driving exposure in terms of frequency, duration and geography.

Some of the field trials are directed towards a specific population, e.g., novice drivers in their learner period in the Austrian field trial, and young experienced drivers in the Israeli trials. Both trials showed that ND studies are well suited to acquire objective information about specific groups of drivers and their driving behaviour. Sampling, of course, has to reflect the population of interest.

Another finding was the importance of a participant’s loyalty in order to avoid participant withdrawal. In an ND study in which the participants use their own cars, it is important that the equipment installed in the cars does not involve any inconveniences on the part of the participants, such as the flat batteries in the Dutch and Austrian trial. This is a factor that can contribute to participants' withdrawal. Furthermore, in Austria the participants appeared to be insufficiently involved in the study to take the time to fill in driving diaries and other self-report measures. Solutions to this problem could be to make sure the participants have a feeling of obligation to fulfil their “duties” as participants, and/or to offer some kind of incentive for participating. It is important, however, to assure that the participants still behave as they would under normal circumstances, and that they don’t behave as they expect that the researchers want them to.

4.2 Research questions and study design

In an ND study in the strictest way, participants are just required to drive as they usually do and data is collected and analysed afterwards. The research questions determine for a large part which data are collected and, consequently, what type of DAS is required.

In addition to this general ND approach, it is also possible to apply the ND technology in a more 'experimental' way to address specific research questions, e.g. by comparing different groups of drivers, by comparing the behaviour in different traffic situations or by comparing the behaviour with and without a particular on-board information or warning system. The common denominator is that the participants' behaviour is observed in otherwise naturalistic conditions. Looked at it this way, the difference between a 'real' ND study and a Field Operational Test is indeed very small, with the ND approach as a research tool rather than a research method.

In the PROLOGUE field trials, different study designs were used, showing the potential of using ND approach as an observation tool for addressing a variety of research questions. The Austrian field trial, for instance, used a quasi-experimental design comparing two groups of drivers. Both groups consisted of novice drivers who had received different types of driver training. The ND approach was used to assess the effect on everyday driving behaviour. Most of the Israeli trials also involved a comparison of driving behaviour between drivers. For instance, one trial looked at the difference in driving exposure and driving behaviour between young drivers and their parents. Another trial looked at the behavioural differences of drivers who belong to a different culture. A third trial did not look at different groups of drivers, but looked at the development of the same drivers over time, i.e. comparing their behaviour in the first period of solo driving and after a few years of experience. In all cases, even though some sort of an experimental study design was applied, the observations were done in the driver's normal and daily driving, i.e. naturalistic, context. Furthermore, it is important to include an initial baseline phase in the design in which participants drive completely unobtrusively and "naturalistic".

It must be noted again that the Spanish and Greek trial were far less naturalistic, since participants drove highly instrumented vehicles and got specific instructions concerning the route (Greek trial) or the destinations (Spanish trial). These trials were not included to assess the usefulness of an ND approach, but more to assess the added value of advanced and very elaborate DASs.

For many research questions complementary data to the direct observational data is useful, e.g. data on self-reported behaviour or information about drivers' personality characteristics or other relevant background information. Self-reported data can, for instance, serve as a basis for validation of the naturalistic data. Additional background data can also provide complementary data that cannot be observed by means of naturalistic observation and that can help to interpret differences between groups, including sample biases. This extra information was collected and used successfully in some of the Israeli trials and the Greek trial. The Austrian trial, on the other hand, showed that participants do not automatically provide the requested additional information; a lack of motivation may be a cause here. While additional data may enrich the observational data, it must be noted that the data collection effort may interfere with the naturalistic nature of the data collection by reminding the drivers of the participation in a study. This may be especially the case if they have to complete questionnaires or diaries on a regular basis.

4.3 Privacy, legal and ethical considerations

Privacy, legal, and ethical issues need to be seriously considered and ensured in ND studies.

As for the legal issues, it is for instance important to make sure that all drivers actually have a driver's license (explicitly stated in the Dutch and Austrian informed consent form) and that the vehicle is insured, both when it is one's own car (Dutch field trial) or when it is a company car (Spanish or Greek field trial). Importantly though, legal issues vary from country to country, and the requirements for studies will differ accordingly.

As the aim of ND studies is to observe participants in their everyday driving, this directly affects privacy requirements. Including video and/or audio data makes the privacy considerations even more important. Participants need to be informed about the type of data that will be collected, for how long it will be stored, who will have access to the data etc., and each participant has to sign an informed consent form stating that they are aware of all these issues. The informed consent form should be tailored to the type of study, i.e., the type of data that is collected, the type of vehicles used, etc. in the study in question.

Furthermore, participants should have the right to withdraw at any point of time from the study and have the data erased.

In cases where the technology also monitors other subjects (e.g. passengers) they should be clearly informed.

Databases should be strictly protected.

5 Conclusions

The aim of the present report was to summarise and discuss the five small-scale field trials that have been conducted within the PROLOGUE project in terms of the usefulness and feasibility of a large-scale ND study in Europe. The trials were not meant to provide an exhaustive overview of what is possible in a technical, methodological and organizational way. They were meant to explore the potential and get a better feeling for the strengths and weaknesses of various approaches. The trials were also meant to illustrate the potential usefulness of ND studies for various road safety research areas. Therefore, the field trials were set up in such a way that they varied on the technology used, the study design, and the areas of interest.

Together with the other project activities that included several literature reviews, workshops and questionnaires, the practical experiences in these trials will feed into the final recommendations for a large-scale European study.

The following questions have served as guidelines for focusing the discussion of the field trials:

- To what extent the small-scale trials illustrated the usefulness and value of conducting ND studies?
- To what extent the trials showed that a large-scale study is technologically and organizationally feasible and what lessons were learned?

In summary, it can be concluded that the five small-scale field trials successfully demonstrated the usefulness and feasibility of conducting ND studies for gaining knowledge in various road safety areas. Furthermore, the experiences gained in the trials showed that there are several important and crucial factors that need to be seriously considered before conducting a large-scale ND study.

5.1 Usefulness and value

The five field trials covered various areas of interest and, even though most of the trials were too small to lead to reliable conclusions, they did show the potential value of an ND approach to obtain knowledge about various aspects of road user behaviour. For example, one of the trials found indications that young drivers speed on different, more dangerous roads than their parents. Another trial found indications that the most popular youngster in group was a safer driver than the least popular one. Yet another trial found indications that feedback to learner drivers based on video images of their own driving behaviour resulted in less risky behaviour in urban areas once they were driving independently. When studying the interaction between motorized vehicles and vulnerable road users, it was found that additional fixed cameras at locations where many of these interactions happen, provide useful extra information, for example about the time-to-collision in car-bicycle encounters. The trials also illustrated that the ND approach can be used to study the use of in-car information systems and to study the desirable and undesirable effects of Advanced Driver Assistance Systems. Again, it must be stressed that the trials were meant as pilots and that small samples were used. Therefore, the findings are just indications and illustrations of what type of research is possible and the results can certainly not be generalized.

Another aspect of drivers' behaviour that was touched upon and explored was the ability of ND and complementary data to "tell the story": what is the sequence of events that is correlated with specific behaviours and what can explain it. This was demonstrated, for example, by administering questionnaires relating to personal characteristics and trip purpose, by analysing GPS location of data, and by looking at in-vehicle behaviour in conjunction with site-based data.

From a methodological point of view, the trials as performed in PROLOGUE only slightly resemble the 'true' ND study, where cars are equipped and driven for a long time without any intervention. The overall project duration of two years as well as the relatively unknown features of European equipment made this unfeasible. However, the usefulness of these long ND studies for studying relevant road safety aspect like fatigue and distraction was already proven by the 100-cars study in the USA (Dingus et al., 2006).

5.2 Feasibility and lessons learned

Very broadly speaking, technologically and organizationally the trials showed that it is feasible to carry out an ND study in Europe. To discuss the lessons learned, the overview of the five field trials and their special focus as presented in Chapter 2, after having changed the columns and the rows, seems to be a usable framework. In the next Sections each of the elements at the rows will be briefly discussed.

Table 5.1 Revisiting the main elements of the PROLOGUE field trials

	Israel	Austria	Netherlands	Spain	Greece
Data collection and storage	x	x	x	x	x
Data reduction & interpretation	x	x	x	x	x
System integration	x		x		
Relation ND data & self reports	x	x			x
Evaluation of interventions	x	x			x
Area of interest	Novice drivers	Driver training	Cyclists	In-vehicle information systems	Active safety technology

5.2.1 Data collection and storage

All trials looked at elements of data collection and storage. Whereas, not surprisingly, the trials showed that is was possible to collect data on a continuous basis and to store it, the main challenge here is to decide on what type of technology should be used and, more specifically, what is the most appropriate degree of complexity of the technological equipment? Of course, the research questions and type of data needed for answering these questions have to be guiding here. At a more detailed level the main lessons learned in this area were:

- A basic and relative simple and cheap off-the-shelf G-based DAS provides very useful data for many research questions. Attention must be paid to the reliability and validity of the identified safety-related events, i.e. avoiding false alarms and missed events.

- For linking driver behaviour to road characteristics such as crash history or design elements, it is useful and time saving to be able to automatically link GPS and GIS data, e.g. by means of a map matching utility.
- Continuous video data is important for accomplishing the great potential of ND studies. This is particularly true for estimating the relative risk (odds ratios) for various risk factors. It is also true for identifying events or near crashes where drivers had not undertaken action while they better had. Continuous video data can also be important for validating G-based events, i.e. by identifying false and missed alarms/events.
- A very advanced DAS, including continuous video registrations by several cameras requires much storage space, and consequently very frequent downloads; if downloads cannot be done wirelessly, this puts a burden on the participants and also threatens the desirable unobtrusive character of the trial since the download reminds the participants of the study.
- Data must be stored in strictly protected databases to guarantee compliance with the European and various national privacy regulations.
- Strict quality assurance procedures must be applied through the duration of the data collection phase to make sure that data is correctly collected and data loss can be kept to a minimum.

5.2.2 Data reduction and interpretation

For data reduction and interpretation, the type and amount of data that is collected is crucial. In particular, the manual processing and analysing of large amount of video data is very time-consuming. Automatic processing and analyses tools are not yet sufficiently reliable, but developments in this area are promising.

Issues that relate to the interpretation of collected ND data and need to be considered carefully are

- Reliable, automatic driver identification to avoid missing information and, in case of deliberate lack of identification, biased results and interpretation.
- Collection of data to analyse the processes proceeding events or conflicts, or generation of sequence of events to gain understanding of specific behaviours.
- Collection of external complementary data to provide better understanding of the observed behaviour. This data includes data on the characteristics of the infrastructure, traffic flow, weather, crash history of specific locations, et cetera.
- Collection of complementary data relating to the observed participants. This data includes personal characteristics (e.g. attitude towards risk), socio-demographic data, crash history, et cetera.
- Collection of data relating to the social environment of the participants. This data includes, for example, data on the presence and behaviour of passengers, data on social relationships among drivers, and driving culture and norms.
- Collection of detailed driving diaries for specific time intervals. This data includes driving exposure, trip purpose, evaluation of safety related events, et cetera.

5.2.3 System integration

Though only studied in a very limited way, it was found that it is difficult and time-consuming to integrate different DASs. This is true both for integrating systems with the aim to compare data from different (national) studies as well as for integrating systems to get additional, complementary data. Therefore, it would be advisable to use one DAS that collects and inherently integrates all the required data.

Integrating site-based video information and in-vehicle data proved to have added value when studying the interaction between cars and bicycles. One issue here is that the use of fixed cameras restricts the observations to a limited number of locations and, consequently, that a long period of data collection is required to get sufficient 'events of interest'.

5.2.4 ND data and self-reports

Self-reports in ND studies have proven to be a valuable extra source of information in different ways. First, they can be used to roughly verify the ND data and ensure no data loss, but more importantly they can be used to complement the ND data.

From methodological point of view, studying the differences between ND data and self-reports are of great importance when tackling the issues of verification and update of self-reported behaviour to realistically correspond to actual observed behaviour.

The added contribution of self-reports to the understanding of ND observations is of utmost importance, as ND do not contain enough information to understand the motives underlying specific behaviours. Hence, the combination of ND and self-reports can provide strong evidence to not only *how* road users behave, but also to *why* they behave the way they do.

5.2.5 Evaluation of interventions

Whereas an ND study in the strictest way does not include any interventions and just monitors drivers in their everyday driving, the current trials have shown that ND data can also be used in a useful way to evaluate interventions or to compare the behaviour of different driver groups or in different situations. The ND data can also be used as source of information as such for the driver to confront them with their own behaviour, aiming to change that behaviour. This is not only applicable for road safety, but could also be used for eco-driving training.

6 References

Backer-Grøndahl, A., Phillips, R., Sagberg, F., Touliou, K., Gatscha, M. (2009). N*aturalistic driving observation: Topics and applications of previous and current naturalistic studies.* PROLOGUE Deliverable D1.1. Institute of Transport Economics, Oslo, Norway.

Christoph, M., Nes, N., Pauwelussen, J., Mansvelders, R., van der Horst, A. R. A., Hoedemaker, M. (2010). *Dutch field trial.* PROLOGUE Deliverable D3.4. TNO, Soesterberg, The Netherlands.

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.* Report no: DOT HS 810 593. National Highway Traffic Safety Administration (NHTSA), Washington DC.

Eichhorn, A. (2011). Final PROLOGUE Workshop Minutes. PROLOGUE Deliverable D5.13. Kuratorium für Verkehrssicherheit KfV, Vienna, Austria.

Eichhorn, A., Van Schagen, I. (eds.) (2011) *Minutes of six regional workshops about Naturalistic Driving.* PROLOGUE Deliverable D5.11. Kuratorium für Verkehrssicherheit KfV, Vienna, Austria.

Eichhorn, A., Winkelbauer, M. (2010) *1st PROLOGUE Workshop Minutes*. PROLOGUE Deliverable D5.6. Kuratorium für Verkehrssicherheit KfV, Vienna, Austria.

Gatscha, M., Brandstätter, C. Pripfl, J. (2010). *Video-based feedback for learner and novice drivers: a small-scale ND field trial in Austria.* PROLOGUE Deliverable D3.3. Test & Training Internationa, Teesdorf bei Baden, Austria.

Groenewoud, C., Schoen, E., Malone, K., Jonkers, E., Hoedemaeker, M., Hogema, J. (2010). *Methodological and organizational issues and requirements for ND studies*. PROLOGUE Deliverable D2.2. TNO, Soesterberg, The Netherlands.

Klauer, S. G., Dingus, T. A., Neale, V.L., Sudweeks, J. D., Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study.* Report No: DOT HS 810 594. National Highway Traffic Safety Administration (NHTSA), Washington D.C.

Lotan, T., Grimberg, E.,et al. (2010). *Specification and Planning of Field Trials.* PROLOGUE Deliverable D3.1. Or Yarok, Hod Hasharon, Israel.

Lotan, T., Albert, G., Ben-Bassat, T., Ganor, D., Grimberg, E., Musicant, O., Hakkert, S., Toledo, T. (2010). Israeli field trial. PROLOGUE Deliverable D3.2. Or Yarok, H

Sagberg, F., Backer-Grøndahl, A. (2010). *A catalogue of applications and research topics for future naturalistic driving studies.* PROLOGUE Deliverable D1.3. TØI Institute of Transport Economics, Oslo, Norway.

Touliou, K., Margaritis, D. (2010*). Small scale field trial in Greece: Focus on assessment of FCW and LDW.* PROLOGUE Deliverable D3.6. CERT/HIT, Thessaloniki, Greece.

Valero-Mora, P. M., Tontsch, A., Pareja-Montoro, I., Sánchez-Garcia, M. (2010). *Field trial in Valencia.* PROLOGUE Deliverable D3.5 INTRAS/University of Valencia, Valencia, Spain.

Van Schagen, I., De Craen, S., Van Nes, N., Eenink, R. (2010). *The areas of interest of potential users for naturalistic observations studies.* PROLOGUE Deliverable D1.2. SWOV Institute for Road Safety Research, Leidschendam, The Netherlands.

Welsh, R., Reed, S., Talbot, R., Morris, A. (2010). *Data collection, analysis methods and equipment for naturalistic studies and requirements for the different application areas.* PROLOGUE Deliverable D2.1. Loughborough University, Loughborough, UK.

List of Abbreviations

ADAS:	Advanced Driver Assistance System
DAS:	Data Acquisition System
DBQ:	Driver Behaviour Questionnaire
FCW:	Forward Collision Warning
FOT:	Field Operational Test
GIS:	Geographic Information System
GPS:	Global Positioning System
GR:	GreenRoad technology
LDW:	Lane Departure Warning
ME:	MobileEye technology
ND:	Naturalistic Driving
SQL:	Structured Query Language
SUS:	System Usability Scale