VTI rapport 664A
Published 2009

www.vti.se/publications

Advantages and disadvantages of different methods to evaluate sleepiness warning systems

Anna Anund
Katja Kircher

vti
FINDING A BETTER WAY

Publisher:

vti

SE-581 95 Linköping Sweden

Publication:
VTI rapport 664A

Published:	**Project code:**	**Dnr:**
2009	40649	2006/0165-26

Project:
DROWSI

Author:
Anna Anund and Katja Kircher

Sponsor:
IVSS

Title:
Advantages and disadvantages of different methods to evaluate sleepiness warning systems

Abstract (background, aim, method, result) max 200 words:

This is a methodological paper with the aim to discuss pros and cons related to different tools and environments when evaluating the effect of warnings given to sleepy drivers. There is no simple answer to the question which platform is most suitable. It depends on the research question asked, and it is possible that different aspects of the problem should be approached with different methods. A driving simulator has clear advantages when high control and repeatability are paramount. A simulator can also be used when the driver has to be put into a potentially dangerous scenario. How ecologically valid the results obtained from a simulator in fact are depends very much on the fidelity of the simulator. A test track study is based on real driving and should have a higher degree of ecological validity. On the other hand, the test track most often consists of an unrealistic environment. For assessing the prevalence of drowsy driving in real traffic, and in order to investigate what drivers actually do when they receive a sleepiness warning, it is absolutely necessary to study their natural behaviour when they go about their daily routines. Here field operational tests or naturalistic driving studies are most suitable. A disadvantage is the lack of control.

Keywords:
Evaluation, sleepiness warning system, method, driving simulator, experiment, test track, real road, FOT

ISSN:	**Language:**	**No. of pages:**
0347-6030	English	38

Utgivare:

vti

581 95 Linköping

Publikation:
VTI rapport 664A

Utgivningsår:	Projektnummer:	Dnr:
2009	40649	2006/0165-26

Projektnamn:
DROWSI

Författare:
Anna Anund och Katja Kircher

Uppdragsgivare:
IVSS

Titel:
Fördelar och nackdelar med olika metoder i samband med utvärdering av trötthetsvarningssystem

Referat (bakgrund, syfte, metod, resultat) max 200 ord:

Detta arbete är ett metodarbete med syfte att diskutera för- och nackdelar med olika verktyg för att utvärdera varningssystem för trötta förare. Det finns inget entydigt svar på vilken metod eller plattform som är den mest lämpliga. Det beror på forskningsfrågan man vill besvara.

En körsimulator har fördelen att erbjuda en hög grad av kontroll och möjlighet till upprepning. Ytterligare en fördel är möjligheten att nyttja scenarion som kan vara farliga under verklig körning, till exempel körning i mycket trött tillstånd. Hurvida simulatorn erbjuder en ekologisk validitet är i hög grad beroende av simulatorns kapacitet och rörelsemekanism. En studie på en testbana är en studie där man kör på riktigt och graden av ekologisk validitet är sannolikt högre än i simulatorn. Å andra sidan är en testbana inte en riktig väg. Om man vill utvärdera prevalensen av körning i trött tillstånd i riktig trafik, men även för att utvärdera hur förare reagerar när de får en varning i trött tillstånd, krävs försök i form av naturalistisk körning där föraren agerar som han eller hon normalt gör i vardagen. I detta sammanhang är så kallade "field operational tests" eller naturalistisk körning sannolikt mest lämpliga. En nackdel med dessa är dock den låga graden av kontroll.

Nyckelord:

Utvärdering, varningssystem för sömnighet, trötthetsvarning, metod, körsimulator, verklig trafik, testbana, FOT

ISSN:	Språk:	Antal sidor:
0347-6030	Engelska	38

Foreword

This methodology study has been possible to perform thanks to the national project DROWSI with fundings from IVSS.

I would like to thank Katja Kircher, VTI, the co-author, for valuable discussions and contribution to the report, Jan Andersson, VTI, for reading and giving valuable comments and Gunilla Sjöberg, VTI, for the support with layout.

Linköping December 2009

Anna Anund

VTI rapport 664A

Cover: Hejdlösa bilder and Katja Kircher, VTI

Quality review

The research director Jan Andersson, VTI, reviewed the report. Main author Anna Anund revised it accordingly and Jan Andersson, VTI, examined and approved the report for publication on 30 November 2009.

Kvalitetsgranskning

Forskningschef Jan Andersson, VTI, har granskat rapporten. Huvudförfattare Anna Anund reviderade rapporten enligt önskemål och Jan Andersson, VTI, granskade och godkände rapporten för publicering 2009-11-30.

Table of contents

Advantages and disadvantages of different methods to evaluate sleepiness warning systems

by Anna Anund and Katja Kircher
VTI (Swedish National Road and Transport Research Institute)
SE-581 95 Linköping Sweden

Summary

Sleep related crashes have received increasing attention during the latest decade. During the last years there has been an increased interest in developing driver support systems that identify sleepiness. These systems normally consist of sensors for measuring physiological and behavioural changes, as well as algorithms to quantify such changes and predict risks. Lots of efforts have been addressed to this area. However, less effort has been targeted at the warning strategies, and how to provide the driver with feedback and/or a warning in a way that the sleepy driver considers the received signals and actually does something to resolve the problem. It is important, however, that the effectiveness of the feedback/warning should be considered in relation to user acceptance in order to make a real step forward. Even the most sensitive algorithm or detection system will do no good if the driver does not understand or accept the warning.

It is difficult to evaluate the effect of a given warning and eliminate confounding factors. Different experimental settings like driving simulators, experimental vehicles, but also environments like test tracks or real roads can be used for the evaluation of the effect of warnings addressed to sleepy drivers. For each of them there is a relation between the realism of the situation and the possibility of controlling the test scenario and confounding factors. This is a methodological paper with the aim to discuss pros and cons related to different tools and environments when evaluating the effect of warnings given to sleepy drivers. There is no simple answer to the question which platform is most suitable. It depends on the research question asked, and it is possible that different aspects of the problem should be approached with different methods. A driving simulator has clear advantages when high control and repeatability are paramount. A simulator can also be used when the driver has to be put into a potentially dangerous scenario. How ecologically valid the results obtained from a simulator in fact are depends very much on the fidelity of the simulator. A test track study is based on real driving and should have a higher degree of ecological validity. On the other hand, the test track most often consists of an unrealistic environment. For assessing the prevalence of drowsy driving in real traffic, and in order to investigate what drivers actually do when they receive a sleepiness warning it is absolutely necessary to study their natural behaviour when they go about their daily routines. Here field operational tests or naturalistic driving studies are most suitable. A disadvantage is the lack of control.

Fördelar och nackdelar med olika metoder i samband med utvärdering av trötthetsvarningssystem

av Anna Anund och Katja Kircher
VTI
581 95 Linköping

Sammanfattning

Trötthetsrelaterade olyckor har fått en ökad uppmärksamhet under senare år. En åtgärd är förarstöd som påkallar förarens uppmärksamhet om trötthet har detekterats. Dessa system består vanligtvis av sensorer för att mäta förarens fysiologi, till exempel ögonrörelser eller körbeteenderelaterade förändringar, de har även matematiska modeller för att kvantifiera förändringar och predicera risk. När det gäller vad man ska göra med själva varningen så har mindre arbete gjorts. Hur ska en varningsstrategi se ut för att få en trött förare att stanna? För att få ett effektivt system är det helt avgörande med ett system som har förarens acceptans. Den mest känsliga och perfekta matematiska modellen kommer inte att vara till nytta om inte föraren förstår och accepterar varningen. Det är ytterst svårt att utvärdera om givna varningar är effektiva eller inte och att i det sammanhanget undvika sammanblandning med andra faktorer. Olika plattformar kan användas: simulatorer, experimentella bilar, men även olika typer av miljöer som testbana eller riktig väg. För var och en av dessa möjliga utvärderingsmiljöer och scenarion finns det för- och nackdelar vad avser realism och möjligheten till kontroll av försökspersoner, scenarion och andra faktorer som kan bidra till sammanblandning av effekter.

Detta arbete är ett metodarbete med syfte att diskutera för- och nackdelar med olika verktyg för att utvärdera varningssystem för trötta förare. Det finns inget entydigt svar på vilken metod eller plattform som är den mest lämpliga. Det beror på forskningsfrågan man vill besvara.

En körsimulator har fördelen att erbjuda en hög grad av kontroll och möjlighet till upprepning. Ytterligare en fördel är möjligheten att nyttja scenarion som kan vara farliga under verklig körning, till exempel körning i mycket trött tillstånd. Hurvida simulatorn erbjuder en ekologisk validitet är i hög grad beroende av simulatorns kapa¬citet och rörelsemekanism. En studie på en testbana är en studie där man kör på riktigt och graden av ekologisk validitet är sannolikt högre än i simulatorn. Å andra sidan är en testbana inte en riktig väg. Om man vill utvärdera prevalensen av körning i trött tillstånd i riktig trafik, men även för att utvärdera hur förare reagerar när de får en varning i trött tillstånd, krävs försök i form av naturalistisk körning där föraren agerar som han eller hon normalt gör i vardagen. I detta sammanhang är så kallade "field operational tests" eller naturalistisk körning sannolikt mest lämpliga. En nackdel med dessa är dock den låga graden av kontroll.

1 Introduction

Sleep related crashes have received increasing attention during the latest decade. The National Transportation and Safety Board (US) has pointed out that sleepiness while driving is one of the most important contributing factors for road crashes (NTSB, 1999). Epidemiological studies based on self-reports or in-depth crash investigations show much higher figures compared to official crash statistics and suggest that about 10 to 20 percent of all crashes might be sleep or fatigue related (Horne & Reyner, 1995; Maycock, 1997; Stutts, Wilkins, Osberg & Vaughn, 2003; Stutts, Wilkins & Vaughn, 1999). It was also demonstrated in post-crash interviews that night driving, prior sleep below five hours, and the sleepiness level before the crash are major predictors of the risk of being involved in a road crash (Connor et al., 2002). In field studies (Dingus, Neale, Klauer, Petersen & Carroll, 2006; Hanowski, Wierwille & Dingus, 2003) sleepiness showed to be the major cause of self-caused crashes/near crashes. Recently, it was also shown that sleepiness may be a stronger cause of road crashes than alcohol and that they interact (Åkerstedt, Connor, Gray & Kecklund, 2008).

Countermeasures to avoid sleep related crashes could be targeted to the human and be placed on the road, in the vehicle, but also more directed against the environment in terms of fatigue management programs, regulations etc. During the last years there has been an increased interest in developing driver support systems that identify sleepiness (Dinges, 1998). These systems normally consist of sensors for measuring physiological and behavioural changes, as well as algorithms to quantify such changes and predict risk. Common measures of driver sleepiness include the standard deviation of the lateral position (O'Hanlon & Kelly, 1974; Otmani, Pebayle, Roge & Muzet, 2005), which increases when the driver becomes sleepy. The electroencephalogram (EEG) with its content of alpha band (8-12Hz) and theta band (4-8Hz) activity (Horne & Reyner, 1996, Gillberg et al., 1996), as well as the electrooculogram (EOG) are other indicators sensitive to sleepiness, which are mostly used as reference values or gold truth. The latter may involve increased duration of eye blinks (Dinges, Maislin, Brewster, Krueger & Carroll, 2005) or slow rolling eye movements (Åkerstedt et al., 1990) both used as indicators in detection systems.

Less effort has been targeted at the warning strategies, and how to provide the driver with feedback and/or a warning in a way that the sleepy driver considers the received signals and actually does something to resolve the problem. It is important, however, that the effectiveness of the feedback/warning should be considered in relation to user acceptance in order to make a real step forward. Even the most sensitive algorithm or detection system will do no good if the driver does not understand or accept the warning.

In order to maintain a high acceptance for a system a correct onset of the warning must be used. A system could be correct about the true value but this does not necessarily imply a high acceptance, since it is not sure that the driver agrees about the diagnosis. Most warning systems strive for both high acceptance and efficiency, and an evaluation of a warning system should take into account both aspects.

Most studies of sleepy driving have been carried out in driving simulators, with a relatively simple and monotonous tracking scenario, and without other vehicles on the road that might require actions (J. Horne & Reyner, 1995; Ingre, Åkerstedt, Peters, Anund & Kecklund, 2006). (Philip et al., 2005) concluded in a comparative study that sleepiness can be studied equally well in real and simulated driving conditions. The effects in terms of changes in driving behaviour are the same, except that the simulator will show

more frequent line crossings and road departures compared to the real environment. One explanation of this difference probably is the lack of complexity in the driving scenario. In a field operational test (FOT) the actual driving has a high degree of realism (Dingus et al., 2006; Hanowski et al., 2003), on the other hand it is demanding to extract the relevant data from the enormous set of data and to make sure which of the crashes and near crashes are related to sleepiness.

To summarize in most studies of sleepy drivers there is a focus on evaluating the detection or prediction of driver sleepiness or impaired driving behaviour caused by sleepiness. So far very little attention is focusing on the evaluation of the effect of feedback or warnings given to the driver. It is difficult to evaluate the effect of a given warning and eliminate confounding factors.

Different experimental settings like driving simulators, experimental vehicles, but also environments like test tracks or real roads can be used for evaluation of the effect of warnings addressed to sleepy drivers. For each of them there is a relation between the realism of the situation and the possibility of controlling the test scenario and confounding factors.

2 Aim

The aim with this work is to discuss pros and cons related to different tools and environments when evaluating the effect of warnings given to sleepy drivers, but also to give some recommendations. The evaluation will also take into account if the dependent variables are measured, observed or self reported.

3 Platforms and settings

With driving simulator we mean middle to high fidelity simulators that provide the driver with an at least somewhat genuine feeling of sitting in a real car. The environment is computer generated, and it is possible to log a host of variables. Test track studies, on-road studies and naturalistic data collection, on the other hand, are all conducted in instrumented vehicles, however, they differ both with respect to the environment and the way the study is arranged. A test track is closed to public traffic, and the experimenters have relatively free hands to adjust the environment to their needs (Anund & Hjälmdahl, 2009; Shutko, 1999; Tijerina, Parmer & Goodman, 1999). Just as a test track study, an on-road study is often quite limited in time, the driven routes are pre-determined, and there may be an experimenter in the car. The study is conducted in real traffic, however (Harbluk, Noy & Eizenman, 2002; Patten, Kircher, Östlund, Nilsson & Svenson, 2006; Philip et al., 2005; Recarte & Nunes, 2000). Naturalistic data collection is also conducted in real traffic, but no experimenter is present in the car, and the studies are usually long-term, lasting for a month or more. The drivers have free choice of route and use the vehicles for their daily lives. Typical examples of naturalistic data collection are naturalistic driving studies like the 100-car study (Dingus et al., 2006; Klauer, Dingus, Neale, Sudweeks & Ramsey, 2006; Neale et al., 2002) or FOTs (Ervin et al., 2005; LeBlanc et al., 2006).

In order to discuss pros and cons related to different tools and environments when evaluating the effect of warnings given to sleepy drivers there is a need to describe related aspects as external validity, control, data quality and implementation abilities.

4 Validity and control

A high degree of external validity and a high degree of control over the study are often difficult to reconcile with each other. Often, the more the experimenter arranges and steers, the less natural the situation becomes.

4.1 External validity – A question of generalisation

From a generic perspective external validity or ecological validity is used as a term for describing the possibility to generalise the results to be valid for situations in real life (Shadish, Cook & Campbell, 2002). To make this possible there is a need for high external validity, here described in terms of realism of the setting for the driver, taking into account aspects as obtrusive/unobtrusive instrumentation, scenarios and how well the collected data correspond to the research question asked. A FOT, using an instrumented car on real roads during normal driving, without a test leader in the car is a method with a potential to reach high external validity (FESTA Consortium, 2008). Data from driving in a simulator, on the other hand, are more difficult to use for generalisation to real life. On the other hand, the internal validity in the simulator will be high, meaning with a correct experimental setup the collected data will have a high relevance for the hypothesis tested, and the experimental control is high.

4.2 Control

In most scientific work it is of great importance to have a high degree of control of the experimental setting in order to reduce the risk for confounded results (Clark & Hawley, 2003). The control is not only related to the scenario used for driving, but also to the selection of the participants and their preparation before the study, and how they are treated during and after the study.

Control is often necessary for reliable repeatability within a reasonable time frame. With a high degree of control unusual scenarios, like a moose crossing the street, can be tested within a short time frame, but also rather more frequent scenarios, like a lead vehicle braking, can be reproduced any number of times under completely similar circumstances.

Within traffic research, control is also of importance in order to minimise the consequences of dangerous events. In an uncontrolled setting a driver who falls asleep can kill both himself and others, while in a controlled study nobody will be harmed. For ethical reasons many studies can only be conducted when a high level of control over the possible consequences is guaranteed, and even though the external validity might be reduced, the only other alternative would be not to conduct the study at all.

In the simulator a high degree of control of the driving scenarios is possible; the same road, conditions etc. will be presented to all drivers, and no unknown situations will appear. It is also possible to ensure that the participants are treated in the same way and that confounding caused by differences in the participant's preparation are minimised. Using an experimental car on a test track will increase the external validity, and depending on the test track there will still be a high degree of control. Using a real road instead of a test track will make the experiment even more ecologically valid, but the possibility to control for confounding events is much more limited. Finally, in an FOT it is not possible to control either for the driving scenario or the driver's behaviour before, under and after the driving; the only way to influence these factors is via the driver

recruiting strategy. On the other hand the results from such a study will have a high degree of ecological validity.

In summary, the simulator will have a high degree of control but a lower degree of external validity (see Figure 1). An experiment on a test track or on a road with an experimental vehicle will still provide a high degree of control of the participants, but a decrease in the control of events like interactions with other road users, animals, etc., but also of the weather and road constructions, for example. The FOT will have a low degree of control but a high degree of external validity. Regardless of the experimental setting the quality of the results is highly dependent on the data quality.

Depending on the aim with the research different experimental settings should be used. If the research question is about the prevalence of sleepiness or long term effects of a warning a FOT will be most suitable. On the other hand, if the research question is more related to increased risk or observations of changes in driver or driving behaviour a simulator study or an experiment on a test track or on-road is usually more suitable.

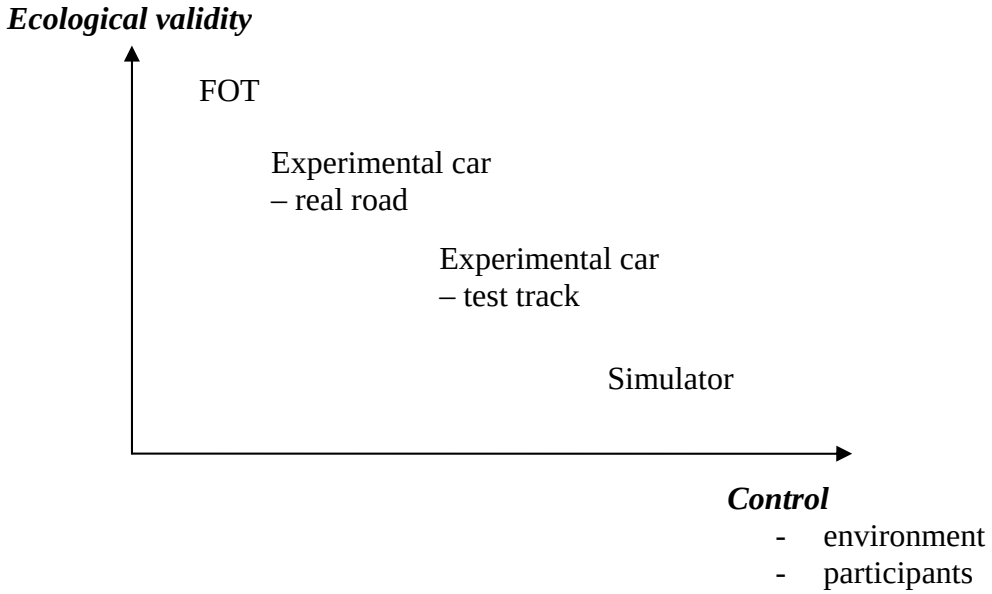

Figure 1 Ecological validity and degree of control for simulators to FOT.

Example 1
The research question is how well sleep deprived drivers can hear a warning signal depending on the pitch and loudness of the warning signal and on the noise level of the environment. The method choice will probably fall on either a driving simulator or even a laboratory setting without a simulator. The main thing is to control the ambient noise, the warning signal and the sleep deprivation level of the participant. The research question is not concerned with real traffic noise, or with ecological validity in any other way, therefore in this case the control over the situation will be maximised.

Example 2

The research question is how likely it is that an extremely sleep deprived driver would stop for a nap when within half an hour from home after a long trip in the car.

Here the research question implies that high external validity is desirable, because the likelihood of a driver's behaviour in "real life" is of interest. Theoretically it would be possible to ask drivers to drive home through the night from far away, after having been awake during daytime. The vehicle they use could be instrumented to check for their driving and resting behaviour. However, for ethical reasons it is impossible to risk the drivers' and other people's lives, therefore a method with a higher control over the consequences has to be chosen, even though some ecological validity will be sacrificed.

Example 3

The research question is how often drivers have use for a sleepiness warning system installed in the car when driving in real traffic at different times of day, taking into account for how long they have been awake, and for how long they have slept the night before.

For this question the naturalistic setting is most useful. To detect sleepiness eye blinks can be logged via a remote eye tracker without the presence of an experimenter in the car. Indicators from the blink complex can be used as sleepiness detection. One possible suggestion for answering this question would be to instrument one or several cars with eye trackers and let people with a high mileage who regularly drive at different hours of the day use the cars. Additionally, the drivers could either wear an actigraph or fill in a sleep diary to document their sleep behaviour.

5 Data

A variable is considered to be *measured* when it is logged by a sensor, which is mounted in the vehicle, the environment or on the driver. Measured data are considered to be objective in the way that they do not depend on a person's judgement. Data collected from CAN bus, radar, eye tracker, GPS receiver and the like are examples of direct measured data. Video films also belong into this category, but data reduced manually from a video film are in a grey zone between measured and observed, depending on the degree of interpretation on behalf of the reductionist. If data from different sources are combined, or if more complex than linear transformations are made on the data, the data are considered to be derived measured data. For a more detailed explanation of the different data types see also (Kircher & WP2.1 group, 2008).

Observed data are those that are recorded by an observer who is often trained for the task. This can happen both in real-time, for example when the observer is present in the car, or off-line, when the observer manually reduces a video film or other logged data. Observer-rated sleepiness is an example of observed data. *Self-reported* data are based on retro- and introspection on behalf of the test person. An example would be self-reported sleepiness. Self-reported data can also be used in order to capture the test person's opinion about acceptance and the effectiveness of a warning system. It could reflect a judgement over a long period of time, like when the test person rates the overall performance of a warning system, but it can also relate to a single warning given by the system in a certain moment.

We differentiate between driver behaviour and driving behaviour in the following way: Driving behaviour is what could be called "vehicle behaviour", it is how the vehicle is moved by the driver, both in lateral and in longitudinal direction. Measures like speed, lateral position, acceleration and related measures are all used to describe driving behaviour. Based on those measures performance indicators can be computed. They indicate how well a driver performs his task with respect to certain criteria (FESTA Consortium, 2008).

Driver behaviour on the other hand is what the driver does, but which does not necessarily influence the vehicle. Examples for measures of driver behaviour are yawning, shifting about in the seat or scratching one's face, which could be interpreted as signs of sleepiness. Physiological measures like the mean and variability of blink duration also belong to driver behaviour. Just as for driving behaviour measures, performance indicators can be computed based on driver behaviour measures.

The data described up to now are the "raw material" for the computation of so-called "performance indicators", which are necessary when comparisons should be made. In most cases driver behaviour should be compared, for example between a baseline and a treatment phase, or against a given threshold, or between different sub-populations for example. The selection of performance indicators should suit the hypothesis in question, but is also limited by the available measures and the quality of the data measured (Kircher & WP2.1 group, 2008).

Example 4

The average duration of gaze fixations and the eye blink duration of a sleep deprived and a well rested driver should be compared in real traffic. An eye tracker is used, and the participant is filmed with a high speed video camera. Due to a logging error the eye tracker only delivers data with a frequency of 10 Hz. This is barely enough to compute fixation duration, but it is completely insufficient for blink duration. Therefore two human reductionists A and B watch the high speed video frame by frame and code eye blink data manually. The inter rater reliability is computed for their coding. Then the mean duration is computed both for fixations and blinks. For the manually coded data the coding of reductionist A were used.

Within the framework of the definitions provided above, the data delivered by the eye tracker are direct measures, which are objective. The video data is also an objective direct measure. The coding of the two raters have a subjective share. Even though they are highly correlated, they do not match a hundred per cent. The project leader decided to use rater A's results, because of her higher grade of experience. This decision influences the results slightly. The mean duration of the fixations and blinks are performance indicators computed on the available data. Had the eye tracker delivered higher frequency data, the mean blink duration might have been slightly different.

The performance indicators will be compared for the two driver states in order to answer the research question.

5.1 Data quality

It is important to collect data whose quality is good enough to allow to perform the intended analyses. There are different aspects of data quality. First of all, the data need to be accurate, that is, the instrument must measure what it is intended to measure. The smaller the unsystematic variation of the data around the true value is the better. A systematic variation is only acceptable if the offset is known. It must be possible to log the data at a certain minimum frequency which depends on the analyses that will be performed. Different frequencies for sensors or ratings can be used. However, analysis of merged data will have a limit in frequency corresponding to the data with the lowest frequency. For mean speed a relatively low frequency of for example 1 Hz, as given by a GPS receiver, is often enough, but some performance indicators require higher frequencies. If necessary to make a derivation for a measures, like for example when computing acceleration from the speed signal, a higher measuring frequency is necessary compared to for example average measures. The resolution of the data must be fine-grained enough to be able to perform the desired analyses and to guarantee that effects can be found on the level required by the hypothesis. Data loss should be at acceptable levels and not systematic. Data loss that can be detected by the logger itself is generally less problematic than data loss that is undetected by the logger and might lead to wrong results if the loss is not or cannot be detected by the analyst. Last but not least it is important to time stamp the log data and keep track of the location, and to be

able to relate self-reported data to a specified time frame or location, to ensure synchronisation between the sampled measures.

> Example 5
>
> The research question is to measure whether there is a difference in the duration of eye blinks when a sleep deprived driver talks with a passenger as compared to when the driver is alone in the car. The study will be conducted on a test track in an instrumented vehicle. The choice is between a head mounted eye tracker and a remote eye tracker. While the remote eye tracker is more comfortable and easier to use, no eye tracking will be possible when the driver turns the head too far away from the cameras and straight ahead. When talking to a passenger, it is expected that the driver will frequently turn the head to the right. A systematic data loss or at least degradation in data quality is not acceptable for this research question; therefore the more cumbersome head mounted eye tracker will be used. This means an increased sacrifice of ecological validity, but guarantees data of higher quality in all relevant research situations.

5.2 Implementation

Obviously it must be possible to implement the study and extract results, not only within budget and time restrictions, but also within the restrictions posed by current technology, the participant of interest, and by ethical and legal considerations. A study requiring advanced vehicle-to-vehicle or vehicle-to-infrastructure communication, which does not yet exist in the road network will limit a study to either a simulator only, or possibly to a test track. If changes in “natural” behaviour with respect to the introduction of a sleepiness warning system should be examined, a field operational test is the only valid solution. If the goal is, however, to evaluate the design of a sleepiness warning in the moment when the driver is sleepy, it is often not feasible to wait for natural sleepiness in an FOT. Ethical and legal considerations prohibit the use of intentionally sleepy participants without ensuring their safety, like in a simulator, on a test track, or possibly in a heavily controlled on-road test.

5.3 Limitations

In order to limit the discussion to a manageable extent, a number of preconditions were assumed to be fulfilled. They are described in the following paragraphs. Additionally, definitions relevant for this paper are provided.

For the present paper the existence of a system is assumed, which is diagnosis or prediction based, with high sensitivity and specificity, having a high acceptance and high effectiveness. The human machine interface (HMI) is designed in an acceptable way, and the warnings are generally perceived as correct. The warning is based on a detection system that uses different measures as input. The warning is based on a combination of modalities with an optimal frequency and amplitude. The participants are representative. The focus of this article is on advantages and disadvantages of different methods with respect to the evaluation of the effect of such a system on driver behaviour and driving

behaviour, and not so much on the evaluation of the quality of the system functionality in itself.

There is also a difference depending on whether we are interested in looking at changes over time (repeated sequences) or occurrences of single events. It is important to decide how to measure the effect of a warning. One more critical question is how to evaluate the effects of repeated events or series of events that could appear in different order. This is not considered within this paper.

6 Evaluation

The discussion deals with three different topics; starting with the platforms, moving on to indicators describing driver behaviour and ending up with indicators describing driving behaviour.

6.1 Platforms

The platform is essential when considering external validity (see table 1). It has been shown that simulators are valuable for sleepiness experiments when the evaluation of relative changes are of interest, but they are less suitable for absolute measures (Philip et al., 2005). Most of these studies deal with changes in driving behaviour caused by sleepiness. Evaluating a warning strategy or HMI taking into account the sleepiness levels themselves is far less common. In a study about the effect of milled rumble strips on sleepy drivers, which is an environmental countermeasure, the simulator proved to have a high external validity in terms of the simulation of rumble strips (Anund, Kecklund, Vadeby, Hjälmdahl & Åkerstedt, 2008). The results showed that this type of warning made the drivers less sleepy, however the effect was short, lasting for less than 5 minutes. If this holds true also in real driving is not known and not easily tested in real driving because of safety and ethical reasons. An advantage with the simulator is the possibility to expose drivers to critical situation without danger. On the other hand the lack of real danger will contribute to a reduced degree of external validity.

There is a risk that the effect of time on task is more pronounced if a monotonous simple scenario is used. Studies have indicated that when using a more complex scenario the drivers' level of sleepiness is reduced (Richter, Marsalek, Glatz & Gundel, 2005).

The experimental vehicle used on test tracks or on roads is real driving by definition. It is still possible to control environmental factors like road type and lighting condition, but it is not possible to control weather, wild animals or other unexpected events. Both on the track and in real traffic the control of the participants' preparation is still as high as for a simulator experiment. Naturalistic driving, on the other hand, has the highest degree of external validity and the lowest degree of control. It is up to the driver to decide where to go, and what to do when. It is not known beforehand if the driver will be sleepy or not.

For data accuracy the main difference lies between simulator studies or studies in instrumented vehicles. When it comes to controllability, the decline from simulator to field is more gradual. In the simulator there are in principle no limitations for the creation of traffic situations. Obviously, there will be limits imposed by the simulator itself, but in most cases the surrounding traffic can be controlled to a much higher degree than on any of the other platforms. On test tracks, for example, it is still possible to stage quite a number of scenarios, but often it will not be possible to repeat them as accurately as it can be done in a simulator, and on test tracks environmental influences like the weather and lighting cannot be controlled as easily, even though attempts are made on high-end test tracks (VTTI, 2007). Situations that might lead to a crash can still be dangerous on a track, and necessary precautions have to be taken. As soon as the study environment is the real traffic, it becomes very difficult to stage more than basic situations. During naturalistic data collection it is impossible to influence the situations that are experienced at all, except via the participant selection. This is one of the reasons

why the data collection periods often are very long in such studies. This way it is hoped to collect enough relevant situations to allow for a meaningful data analysis.

The high precision of the data and the level of control that can be obtained from the simulator have a downside, which is the reduced external validity. It is often not clear how well the driving behaviour measured in the simulator reflects the driving behaviour in real traffic. In some studies comparisons have been made (Philip et al., 2005), and the results show a deviation in absolute levels, but not on a relative level. This is most probably simulator, scenario and environment dependent and an area for further research.

Table 1 General aspects related to the use of different platforms for evaluation of warning strategies and HMI addressed to sleepy drivers.

	Simulator	Test track	On-road experiment	FOT
General Aspects	**External validity**	**External validity**	**External validity**	**External validity**
	External validity relatively low, artificial situation, observed and safe environment.	External/ ecological validity questionable, but probably somewhat better than simulator.	External/ ecological validity better than test track, but experimenter bias possible.	High external/ ecological validity, naturalistic driving situation.
	A test leader observes from outside.	Test leader often present in vehicle.	Test leader often present in vehicle.	No test leader present in vehicle.
	Predetermined route without naturalistic purpose.	Predetermined route without naturalistic purpose.	Predetermined route without naturalistic purpose.	Route chosen by driver with naturalistic purpose.
	Computer generated road, can resemble real roads, but typically simplified environment.	Often somewhat unrealistic road environment (especially on round or oval tracks).	Real road.	Real Road.
	The sleepy driver will not stop driving when tired to reduce crash risk.	The sleepy driver might stop driving when tired to reduce crash risk.	The driver will stop driving when tired to reduce crash risk.	The driver will stop driving when tired to reduce crash risk.
	After a small number of sleepiness warnings the drivers become indifferent to further warnings due to time constraints in the study.	After a small number of sleepiness warnings the drivers become indifferent to further warnings due to time constraints in the study.	After a small number of sleepiness warnings the drivers become indifferent to further warnings due to time constraints in the study.	Warning frequency no concern, a “naturalistic” frequency can be expected.

Absolute levels of measured values are not useful, but relative levels are.	Absolute levels of measured values are not useful, but relative levels are.	Often absolute levels of measured values can be used directly.	Absolute levels of measured values can be used directly.
Obtrusive instrumentation like electrodes possible, will not reduce external validity much more.	Obtrusive instrumentation like electrodes possible, external validity might be reduced.	Obtrusive instrumentation like electrodes possible, reduction of external validity.	Obtrusive instrumentation like electrodes cannot be used.
Control	**Control**	**Control**	**Control**
High control over the driving scenario, can be programmed.	High control over the scenario, scenarios can usually be staged, some restrictions due to safety.	Reduced degree of control over the scenario, only small possibilities to stage scenarios due to safety, traffic code and other considerations.	No control over the scenario except via driver selection, impossible to stage scenarios.
High control over the road geometry can be programmed.	The road geometry is relative well known and somewhat controllable via markings and lines.	The road geometry can be investigated, but can only be influenced by route choice.	No control over the road geometry.
High control over situational factors like the weather.	Some control over situational factors like the weather, some advanced tracks have rain machines, slippery tracks, etc.	Low control over situational factors like the weather, except via driving schedule.	No control over situational factors like the weather.
High possibility to repeat the situation.	Possible to repeat the situation.	Not easily possible to repeat situations.	No control over scenario repetition.
High control over participant selection and driver state.	High control over participant	High control over participant	Medium to high control over

	selection and driver state.	selection and driver state.	participant selection, no control over driver state.
Possible to force the drivers to experience several warnings.	Possible to force the drivers to experience several warnings.	Not easily possible to force the drivers to experience several warnings.	Not possible to force the drivers to experience several warnings.
Participant safety guaranteed.	Safety issues and test track restrictions have to be considered.	Necessary to take precautions for participant safety.	Participants responsible for their own safety, no special precautions from experimenter's side.

6.2 Driver behaviour

In this section blink duration and self reported sleepiness are selected as examples for driver behaviour. Blink duration has a high validity for describing sleepiness while driving (Anund, Kecklund, Peters, Forsman & Åkerstedt, 2008; Ingre et al., 2006; Otmani, Joceline & Muzet, 2005). Blink duration could be measured by obtrusive electrooculogram (EOG) or observed by unobtrusive cameras. As a subjective measure of sleepiness the Karolinska Sleepiness Scale could be used (Åkerstedt & Gillberg, 1990). In table 2 general aspects of obtaining those performance indicators as examples of measured, observed and self-reported data are presented.

Table 2 Driver behaviour indicators related to the use of different platforms for evaluation of warning strategies and HMI for sleepy drivers.

	Simulator	Test track	On-road experiment	FOT
Driver behaviour – Measured: Blink duration with EOG	**Data quality**	**Data quality**	**Data quality**	**Data quality**
	High sampling frequency (512 Hz/256Hz).	High sampling frequency (512 Hz/256Hz).	High sampling frequency (512 Hz/256Hz).	Not possible to use EOG log equipment.
	Quality is sensor dependent.	Quality is sensor dependent.	Quality is sensor dependent.	
	No missing data caused by sun, light etc.	No missing data caused by sun, light etc.	No missing data caused by sun, light etc.	
	Implementation	**Implementation**	**Implementation**	**Implementation**
	Obtrusive sensors possible to use.	Obtrusive sensors possible to use.	Obtrusive sensors possible to use.	Not possible to use EOG log equipment, because obtrusive instrumentation cannot be used.
	EOG in simulator setting not unexpected, no “embarrassment” for participant.	EOG on test track not unexpected, no “embarassment” for participant.	Participant might feel embarrassed to wear EOG equipment in real traffic.	
	Off-line analysis needed.	Off-line analysis needed.	Off-line analysis needed.	

Driver behaviour – Observed: Blink duration with cameras	**Data quality** Medium sampling frequency (Smart Eye for example 60 Hz). No data loss caused by sun, light etc.	**Data quality** Medium sampling frequency (Smart Eye for example 60 Hz). Data loss possible due to sun, light, etc.	**Data quality** Medium sampling frequency (Smart Eye for example 60 Hz). Data loss possible due to sun, light, etc.	**Data quality** Medium sampling frequency (Smart Eye for example 60 Hz). Data loss possible due to sun, light, eye glasses, winter clothes and headgear etc. of participants (more difficult to control such factors as in other experimental settings).
	Implementation Unobtrusive sensor, installation in simulator unproblematic, frequent recalibration possible, relatively constant environment.	**Implementation** Unobtrusive sensor, installation in instrumented car relatively unproblematic, frequent recalibration possible, tracking quality can be influenced by severe temperature shifts.	**Implementation** Unobtrusive sensor, installation in instrumented car relatively unproblematic, frequent recalibration possible, tracking quality can be influenced by severe temperature shifts.	**Implementation** Unobtrusive sensor, installation in instrumented car relatively unproblematic, possible to use over longer time periods if no re-calibration necessary, tracking quality can be influenced by temperature shifts, movement of cameras, etc. More problematic if participants use their own vehicles.
Driver behaviour – Self reported: Karolinska Sleepiness Scale	**Data quality** Low frequency, average for 5 minutes. The relation to driving impairment is known.	**Data quality** Low frequency, average for 5 minutes. The relation to driving impairment is known.	**Data quality** Low frequency, average for 5 minutes. The relation to driving impairment is known.	**Data quality** Not possible to use.

Implementation	**Implementation**	**Implementation**	**Implementation**
Easy – needs training of the participants. Interaction with experimenter or other prompting device required.	Easy – needs training of the participants. Interaction with experimenter or other prompting device required.	Easy – needs training of the participants. Interaction with experimenter or other prompting device required.	Not possible to use.

Indicators of sleepiness based on measured data, like blink duration, are not possible to obtain on all types of platforms. In driving simulators, on test tracks and on-road tests it is possible, even if it most probably will cause a reduction of ecological validity especially in the latter two. One advantage with the obtrusive sensor is the high sampling frequency and a low risk for missing data. In an FOT there is no possibility to use obtrusive sensors at all. Here unobtrusive sensors like cameras are needed. The advantage is that the blink indicators can be easily extracted, on the other hand the resolution is sensor dependent and at this point this is a limitation. The sensor does not work with a frequency high enough to extract the most promising and less individual dependent ratio measures like amplitude of a blink in relation to lid closing or opening speed (Johns, Chapman, Crowley & Tucker, 2008). There are also problems with environmental disturbance as sunshine, eye glasses etc.

When evaluating the effect of a warning system the aim of the system should be decided at an early stage: a high level of acceptance or a high level of correct warnings from a physiological point of view. A system with a high degree of acceptance will most probably be used by the drivers. The Karolinska Sleepiness Scale (KSS) is a self-reported measure that can easily be used in simulator, on a test track and for on-road experiments. The measure describes the drivers' experience of sleepiness and the relation between KSS and acceptance to receive a warning is most probably high. The KSS is validated against physiological sleepiness (Åkerstedt & Gillberg, 1990). Most studies within this area focus on the situation before a warning is given. What happens afterwards is not described. Anund et al. (2008) showed that sleepy drivers hitting milled rumble strips reported a higher degree of sleepiness after a hit. Looking into other measures an increase of alertness after the hit was seen, this is a contradiction compared to the KSS. One explanation could be that due to the warning the drivers realized that they were sleepy. Another explanation could be that the frequency of the KSS reporting (once each fifth minute) did not make it possible to look at the direct effects of a given warning. KSS cannot be used at all in FOTs without severely disturbing natural behaviour. It is therefore not possible to look at the relationship between real-time experienced sleepiness and received sleepiness warnings. This has to be captured with help of for example questionnaires after driving.

6.3 Driving behaviour

In this section lane keeping quality is selected as one example performance indicator of driving behaviour. Adequate lane keeping is important for traffic safety, because in its extremes it means that the driver will either run off the road or cross into an adjacent lane. Different aspects of lane keeping quality are often used when assessing the driving performance of sleepy drivers in simulators (O'Hanlon & Kelly, 1974; Otmani, Pebayle et al., 2005; Philip et al., 2005). Milled rumble strips address bad lane keeping performance, which is helpful for sleepy drivers (Anund, Kecklund, Vadeby et al., 2008). A great many performance indicators exist that describe the quality of lane keeping, for example the standard deviation of lateral position (SDLP), the number of lane departures, time to line crossing (TLC), and so on (Otmani, Pebayle et al., 2005; Wierwille, Ellsworth, Wreggit, Fairbanks & Kim, 1994). They all have in common that they are based on the lateral position of the vehicle on the road. In table 3 general aspects of obtaining performance indicators describing the quality of lane keeping are presented for measured, observed and self-reported data.

Table 3 Driving behaviour indicators related to the use of different platforms for evaluation of warning strategies and HMI for sleepy drivers.

	Simulator	Test track	On-road experiment	Naturalistic data collection
Driving behaviour – Measured: Lane keeping performance based on lateral position	**Data quality**	**Data quality**	**Data quality**	**Data quality**
	Data quality is usually high and precise, high logging frequencies possible, high accuracy, high resolution, precise tracking info.	Data quality usually lower than in simulator, logging frequency sensor dependent, typically lower than simulator.	Data quality usually lower than in simulator, similar or somewhat lower than test track, logging frequency sensor dependent, typically lower than simulator.	Data quality and logging frequency usually lower than in the other settings, logging frequency sensor and storage dependent, data loss occurs, remote monitoring of data quality essential during trial.
	Lane position data is logged immediately and not computed, normally no data loss.	Lane position data quality depends on sensor and the quality of the road edge, which might be possible to simulate on a test track, data loss occurs.	Lane position data quality depends on sensor and the quality of the road edge, data loss occurs.	Data quality depends on sensor and the quality of the road edge.
	Implementation	**Implementation**	**Implementation**	**Implementation**
	Easily registered quantitatively, lateral position standard measure in simulator studies, easy to analyse.	Special equipment necessary to register lateral position; if LDW present, data should be on CAN, otherwise necessary to build in lane tracker or do an off-line video analysis of a camera at the wheel filming the street (see: observed).	Special equipment necessary to register lateral position; if LDW present, data should be on CAN, otherwise necessary to build in lane tracker or do an off-line video analysis of a camera at the wheel filming the street (see: observed).	Special equipment necessary to register lateral position; if LDW present, data should be on CAN, otherwise necessary to build in lane tracker or do an off-line video analysis of a camera at the wheel filming the street (see: observed).

Driving behaviour – Observed: Lane keeping performance	**Data quality** Many cameras can be used, high frequency and high resolution possible.	**Data quality** Depends on camera or observer position, if to be judged from forward facing camera filming the whole traffic scene, the accuracy is not as good as if filmed from wheel camera.	**Data quality** Depends on camera or observer position, if to be judged from forward facing camera filming the whole traffic scene, the accuracy is not as good as if filmed from wheel camera.	**Data quality** Depends on camera position, if to be judged from forward facing camera filming the whole traffic scene, the accuracy is not as good as if filmed from wheel camera.
	Implementation In simulator driving behaviour observation with respect to lane tracking quite unusual, sometimes done in sleepiness studies for warning validation purposes. Observe driving behaviour via cameras in real time. In many simulators possible in principle to use observing passenger, but rarely used. Off-line analysis possible when filmed.	**Implementation** Observer can judge lane keeping either on-line when in the car or off-line if video is recorded. If wheel camera installed an observer might note on-line whether the line was exceeded, for example. Overall driving quality can be rated with e. g. the Wiener Fahrprobe, a standardised method, which in its original format necessitates two observers in the car, lane keeping quality can be rated as part of driving quality. Off-line analyses from camera very time consuming.	**Implementation** Observer can judge lane keeping either on-line when in the car or off-line if video is recorded If wheel camera installed an observer might note on-line whether the line was exceeded, for example. Overall driving quality can be rated with e. g. the Wiener Fahrprobe, a standardised method, which in its original format necessitates two observers in the car, lane keeping quality can be rated as part of driving quality. Off-line analyses from camera very time consuming.	**Implementation** No observer in car, possible to judge quality of lane keeping off-line if video is recorded. No real time observation of lane keeping performance possible. Off-line analyses from camera very time consuming.

Driving behaviour – Self-reported: Lane keeping performance	**Data quality** No standardised method to assess self-reported lane keeping performance known. Data quality depends on instrument used.	**Data quality** No standardised method to assess self-reported lane keeping performance known. Data quality depends on instrument used.	**Data quality** No standardised method to assess self-reported lane keeping performance known. Data quality depends on instrument used.	**Data quality** No standardised method to assess self-reported lane keeping performance known. Data quality depends on instrument used.
	Implementation Obtainable both on-line while driving or off-line after the trip. Obtainable for very well defined situations if asked on-line easy to obtain via questionnaire, interview or rating scales. Prompts in simulator environment can be used for answer triggering.	**Implementation** Obtainable both on-line while driving, or off-line after trip. Obtainable for very well defined situations if asked on-line prompts in vehicle (triggered e. g. via transponders) can be used for answer triggering. If experimenter in car: verbal answer triggering possible. After the fact: video confrontation for situation specific answers (often short time between recording and video watching; confusion possible if track short and many laps driven). Obtainable via questionnaire, interview or rating scales.	**Implementation** Obtainable both on-line while driving, or off-line after trip. Obtainable for very well defined situations if asked on-line prompts in vehicle (triggered e. g. via transponders) can be used for answer triggering. If experimenter in car: verbal answer triggering possible. After the fact: video confrontation for situation specific answers (often short time between recording and video watching; confusion possible if track short and many laps driven). Obtainable via questionnaire, interview or rating scales.	**Implementation** Either rather general over a long period of time, or the participant will have to fill in questionnaires or be interviewed at certain times during the study, which might interrupt the study and remind the driver of his being observed. Triggered prompts not very common for naturalistic data collection, could be done for e. g. lane exceedances. General answers obtainable via questionnaire, interview or rating scale.

Lateral position, and thus all performance indicators based on this measure, are very accurate when obtained from a simulator. As soon as an instrumented vehicle is used, it becomes much more complicated to log this measure. For automatic detection either a painted line or a good contrast between the road edge and the adjacent area has to be present, just as for lane departure warning systems. If such a system is available on the vehicle, and if it is possible to access its data, then logging is usually not a problem. Otherwise, a custom made solution has to be found, which can either be automated based on image recognition, or manual, based on a video taken from the host vehicle, from a following vehicle or from roadside cameras. Manual data reduction is time consuming and error susceptible, which renders its use practically impossible for studies that last over a longer time period. An alternative is to analyse lateral position based on video recordings for critical situations and selected matching baseline clips only. A differentiation into “lane kept” versus “departed from lane” can be made faster than an estimation of the actual lateral position. For the evaluation of warning systems for sleepy drivers the latter is of main interest. The starting point for this article is a correctly given warning, but even so we need to identify this event for the evaluation. This is easier in a driving simulator with all possibilities to do an off-line analysis, and where the degree of control is high.

Lane tracking is not very often collected as an observed variable, especially not in driving simulators. It can be an integrated part of formalised observer-based techniques like the “Wiener Fahrprobe” (Chaloupka & Risser, 1995). Here two observers evaluate driving behaviour both based on traffic rules and more generally on driving style. Lane keeping behaviour can be one aspect of this observation.

In some cases drivers are asked to estimate their lane tracking performance, for example in relation to a baseline. In those cases it is very valuable to obtain objective measures of lane keeping, too, in order being able to make comparisons between the perceived and the measured lane keeping behaviour. Just like for impairment caused by alcohol intake sleepiness will reduce the drivers’ capability of performance (J. A. Horne & Baumber, 1991). The correctness of the judgement can depend on the drivers improve capability of judgement. There are great individual differences in drivers’ variation in lateral position (Ingre et al., 2006) and the differences between drivers are more pronounced than differences within an individual comparing driving under alert versus sleepy conditions.

7 Conclusion

Which platform is best? Obviously there is no simple and clear answer to this question. It depends on the research question asked which platform should be chosen, or different aspects of the problem should be assessed with different methods. Especially if the results then point into a common direction this approach of triangulation is very strong. A driving simulator has clear advantages when high control and repeatability are paramount. A simulator can also be used when the driver has to be put into a potentially dangerous scenario. This can be done in a simulator without any real danger to the driver. How ecologically valid the results obtained from a simulator in fact are depends very much on the fidelity of the simulator. A moving base apparatus with a realistic visual environment and vehicle model can yield quite acceptable results, especially if relative values between conditions instead of absolute values are of interest. If there is a need for obtrusive sensors for measuring physiological sleepiness (EEG/EOG/EKG) the simulator is more suitable.

If problems inherent to a simulator are to be avoided, but it is still necessary to keep as much control over the scenarios as possible, a test track setting can be used. A test track study is based on real driving and should have a higher degree of ecological validity. On the other hand the test track most often consists of an unrealistic environment. If an oval track is used the lateral position and variability cannot easily be compared with real driving. If the test track consists of curves and sections of different speed limits there could still be problems related to the sensors. It is not easy to obtain a high quality of vehicle data at low speeds and in sharp bends. If sleepy drivers are made to drive in real traffic there is a need to have a test leader in the car for safety reasons. When evaluating the effects of given warnings there is a risk that he or she will have influenced the drivers' reactions.

For long term studies of sleepiness mitigation systems FOTs are the only possible alternative. Behavioural changes and behavioural adaptation over time, be it positive or negative, can only be tested in long-term studies that are impossible in a simulator or on a track. For assessing the prevalence of drowsy driving in real traffic and in order to investigate what drivers actually do when they receive a sleepiness warning it is absolutely necessary to study their natural behaviour when they go about their daily routines. If a driver gets a recommendation to stop and rest for a while, he may very well do so while participating in a study, both to please the experimental leader and because he or she does not have anything better to do anyway. In reality, however, the driver might be under time pressure or just wants to get home, and there is no need to impress an experimental leader. Therefore, when the main research question is to investigate realistic behaviour under realistic conditions, the method of choice is an FOT or a naturalistic driving study.

On the other hand in an on-road experiment it will still be possible to capture the effect of a warning system in terms of changes in driver and driving behaviour before and after a given warning while keeping a high level of control and a low degree of confounding with other factors. It is necessary and possible to use more controlled setups during feasibility studies and for tuning warning strategies and modalities.

8 Acknowledgment

This study was supported by the national Swedish Project DROWSI.

References

Åkerstedt, T., Connor, J., Gray, A. & Kecklund, G. 2008. Predicting road crashes from a mathematical model of alertness regulation – The Sleep/Wake Predictor. *Accid Anal Prev, 40*(4), 1480–1485.

Åkerstedt, T. & Gillberg, M. 1990. Subjective and objective sleepiness in the active individual. *Int J Neurosci, 52*, 29–37.

Anund, A. & Hjälmdahl, M. (2009). *Evaluation of a sleepiness warning system – a test track study* (VTI report No. 01). Linköping.

Anund, A., Kecklund, G., Peters, B., Forsman, Å. & Åkerstedt, T. 2008. Driver impairment during night and the relation with physiological sleepiness. *Scandinavian Journal of Work Environment and Health, 34*, 142–150.

Anund, A., Kecklund, G., Vadeby, A., Hjälmdahl, M. & Åkerstedt, T. 2008. The alerting effect of hitting a rumble strip – a simulator study. *Accid Anal Prev, 40*, 1970–1976.

Chaloupka, C. & Risser, R. 1995. Don't wait for accidents - possibilities to assess risk in traffic by applying the "Wiener Fahrprobe". *Safety Science, 9*(2-3), 137–147.

Clark, P. & Hawley, K. (2003). *Philosophy of Science Today*: OXFORD UNIVERSITY PRESS.

Connor, J., Norton, R., Ameratunga, S., Robinson, E., Civil, I., Dunn, R., et al. 2002. Driver sleepiness and risk of serious injury to car occupants: population based case control study. *Br. Med. J., 324* 1125.

Dinges, D., Maislin, G., Brewster, R., Krueger, G. & Carroll, R. (2005). *Pilot test of fatigue management technologies* (No. Issue 1922): US Transportation Research Record.

Dingus, T., Klauer, S., Neale, V., Petersen, A., Lee, S. & Sudweeks, J. (2006). *The 100-car naturalistic driving study, phase II – results of the 100-car field experiment* (No. Technical Report DOT HS 810 593). Washington, DC: NHTSA.

Dingus, T., Neale, V., Klauer, S., Petersen, A. & Carroll, R. 2006 The development of a naturalistic data collection system to perform critical incident analysis: an investigation of safety and fatigue issues in long-haul trucking. *Accid Anal Prev, 38* (6), 1127–1136.

Ervin, R., Sayer, J., LeBlanc, D., Bogard, S., Mefford, M. & Hagan, M. (2005). *Automotive collision avoidance field operational test. Methodology and results* (Technical report No. UMTRI-2005-7-1; US DOT HS 809 900). Ann Arbor, MI, USA: University of Michigan Transportation Research Institute.

FESTA Consortium. (2008). *FESTA Support Action - D6.4 FESTA Handbook* (Deliverable). Brussels: European Commission.

Hanowski, R.J., Wierwille, W.W. & Dingus, T.A. 2003. An on-road study to investigate fatigue in local/short haul trucking. *Accid Anal Prev, 35*(2), 153–160.

Harbluk, J., Noy, Y. & Eizenman, M. (2002). *The impact of cognitive distraction on driver visual behaviour and vehicle control* (TP): 13889E, Transport Canada.

Horne, J. & Reyner, L. 1995. Driver sleepiness. *J Sleep Res, 4*(2), 23–29.

Horne, J.A. & Baumber, C.J. 1991. Time-of-day effects of alcohol intake on simulated driving performance in women. *Ergonomics, 11*, 1377–1383.

Ingre, M., Åkerstedt, T., Peters, B., Anund, A. & Kecklund, G. 2006. Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. *Journal of Sleep Research, 15*, 1–7.

Johns, M., Chapman, R., Crowley, K. & Tucker, A. 2008. A new method for assessing the risks of drowsiness while driving. *Somnologie, 12*, 66–74.

Kircher, K. & WP2.1 group. (2008). *A comprehensive framework of performance indicators and their interaction* (No. Deliverable of the EU Project FESTA in the 7th Frame Program – D2.1). Linköping.

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

LeBlanc, D., Sayer, J., Winkler, C., Ervin, R., Bogard, S. & Devonshire, J. (2006). *Road departure crash warning system field operational test: Methodology and results.* (Technical report No. UMTRI-2006-9-1). Ann Arbor, MI, USA: The University of Michigan Transportation Research Institute (UMTRI).

Maycock, G. 1997. Sleepiness and driving: The experience of U.K. car drivers. *Accid Anal Prev, 29*, 453–462.

Neale, V., Klauer, S., Knipling, R., Dingus, T., Holbrook, G. & Petersen, A. (2002). *The 100 car naturalistic driving study, phase I – experimental design* (Interim Report No. DOT HS 808 536). Washington, DC: NHTSA.

NTSB. (1999). *Evaluation of U.S. Department of Transportation: efforts in the 1990s to address operation fatigue* (No. Safety Report NTSB/SR-99/01). Washington, D.C.: National Transportation Safety Board.

O'Hanlon, J. & Kelly, G. 1974. A psycho-physiological evaluation of devices for preventing lane drift and run-off-road accidents. *Technical Report 1736-F, Human Factors Research Inc, Santa Barbara Research Park, Goleta, California*.

Otmani, S., Joceline, R. & Muzet, A. 2005. Sleepiness in professional drivers: Effect of age and time of day. *Accid Anal Prev, 37*, 930–937.

Otmani, S., Pebayle, T., Roge, J. & Muzet, A. 2005. Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers. . *Physiol Behav 84*, 715–724.

Patten, C., Kircher, A., Östlund, J., Nilsson, L. & Svenson, O. 2006. Driver experience and cognitive workload in different traffic environments *Accid Anal Prev, 38*(5), 887–894.

Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Åkerstedt, T., et al. 2005. Fatigue, sleepiness, and performance in simulated versus real driving conditions. *Sleep, 28(12)*, 1511–1516.

Recarte, M. & Nunes, L. 2000. Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied, 6*(1), 31–43.

Richter, S., Marsalek, K., Glatz, C. & Gundel, A. 2005. Task-dependent differences in subjective fatigue scores. *J Sleep Res, 14*(4), 393–400.

Shadish, W., Cook, T. & Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

Shutko, J. (1999). *An investigation of collision avoidance warnings on brake response times of commercial motor vehicle drivers.* (Unpublished Master Thesis). Blacksburg, Virginia, US: State University.

Stutts, J., Wilkins, J., Osberg, S. & Vaughn, B. 2003. Driver risk factors for sleep-related crashes. *Accid Anal Prev, 35,* 321–331.

Stutts, J., Wilkins, J. & Vaughn, B. (1999). *Why do people have drowsy driving crashes? Input from drivers who just did* (No. 202/638-5944). Washington, D.C.: AAA Foundation for Traffic Safety.

Tijerina, L., Parmer, E. & Goodman, M. (1999). *Individual differences and in-vehicle distraction while driving: A test track study and psychometric evaluation.* Paper presented at the 5th ITS World Congress.

VTTI. (2007). The Virginia Smart Road. Retrieved Retrieved 01-07, 2009, http://www.vtti.vt.edu/virginiasmartroad.html

Wierwille, W.W., Ellsworth, L.A., Wreggit, S.S., Fairbanks, R.J. & Kim, C.L. (1994). *Research on vehicle based driver status/performance monitoring: development, validation, and refinement of algorithms for detection of driver drowsiness.* (No. DOT HS 808).

www.vti.se
vti@vti.se

VTI är ett oberoende och internationellt framstående forskningsinstitut som arbetar med forskning och utveckling inom transportsektorn. Vi arbetar med samtliga trafikslag och kärnkompetensen finns inom områdena säkerhet, ekonomi, miljö, trafik- och transportanalys, beteende och samspel mellan människa-fordon-transportsystem samt inom vägkonstruktion, drift och underhåll. VTI är världsledande inom ett flertal områden, till exempel simulatorteknik. VTI har tjänster som sträcker sig från förstudier, oberoende kvalificerade utredningar och expertutlåtanden till projektledning samt forskning och utveckling. Vår tekniska utrustning består bland annat av körsimulatorer för väg- och järnvägstrafik, väglaboratorium, däckprovnings-anläggning, krockbanor och mycket mer. Vi kan även erbjuda ett brett utbud av kurser och seminarier inom transportområdet.

VTI is an independent, internationally outstanding research institute which is engaged on research and development in the transport sector. Our work covers all modes, and our core competence is in the fields of safety, economy, environment, traffic and transport analysis, behaviour and the man-vehicle-transport system interaction, and in road design, operation and maintenance. VTI is a world leader in several areas, for instance in simulator technology. VTI provides services ranging from preliminary studies, highlevel independent investigations and expert statements to project management, research and development. Our technical equipment includes driving simulators for road and rail traffic, a road laboratory, a tyre testing facility, crash tracks and a lot more. We can also offer a broad selection of courses and seminars in the field of transport.

vti

HUVUDKONTOR/HEAD OFFICE
LINKÖPING
POST/MAIL SE-581 95 LINKÖPING
TEL +46(0)13 20 40 00
www.vti.se

BORLÄNGE
POST/MAIL BOX 920
SE-781 27 BORLÄNGE
TEL +46 (0)243 446 860

STOCKHOLM
POST/MAIL BOX 55685
SE-102 15 STOCKHOLM
TEL +46 (0)8 555 770 20

GÖTEBORG
POST/MAIL BOX 8077
SE-402 78 GÖTEBORG
TEL +46 (0)31 750 26 00