ISBN 978-82-326-1946-7 (printed ver.)
ISBN 978-82-326-1947-4 (electronic ver.)
ISSN 1503-8181

Arne Helland

Doctoral thesis

Doctoral theses at NTNU, 2016:304

Arne Helland

Driving Simulator Validation for Drug Impairment Research

Results from a randomized, controlled study comparing simulated and test track driving under the influence of ethanol

Doctoral theses at NTNU, 2016:304

NTNU
Norwegian University of
Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Medicine
Department of Laboratory Medicine,
Children's and Women's Health

NTNU

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

Arne Helland

Driving Simulator Validation for Drug Impairment Research

Results from a randomized, controlled study comparing simulated and test track driving under the influence of ethanol

Thesis for the Degree of Philosophiae Doctor

Trondheim, October 2016

Norwegian University of Science and Technology
Faculty of Medicine
Department of Laboratory Medicine,
Children's and Women's Health

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Medicine
Department of Laboratory Medicine,
Children's and Women's Health

ISBN 978-82-326-1946-7 (printed ver.)
ISBN 978-82-326-1947-4 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2016:304

Printed by NTNU Grafisk senter

Norsk sammendrag

Kjøresimulatorer brukes blant annet til å undersøke effekten av legemidler og rusmidler på kjøreatferd og trafikkrisiko. Slik anvendelse forutsetter at testene i kjøresimulatoren virkelig måler det samme som ved kjøring på vei og er tilstrekkelig følsomme – altså at simulatortestene både har *validitet* og *sensitivitet*.

Denne avhandlingen beskriver resultatene fra en valideringsstudie hvor vi undersøkte effekten av alkohol på en rekke kjørerelaterte utfallsmål, både under virkelig kjøring på en lukket testbane og i en kjøresimulator som etterlignet forholdene på testbanen.

De fleste utfallsmålene var følsomme for alkoholeffekter i kjøresimulatoren, mens færre utfallsmål var like følsomme ved virkelig kjøring på testbanen. Kjøretøyets *grad av vingling i veibanen* (SDLP) hadde den høyeste sensitiviteten for alkoholeffekter i begge testmiljøene, og viste størst grad av samsvar mellom kjøresimulatoren og banekjøringen.

Avhandlingen gransker videre to faktorer som kompliserer gjennomføringen og tolkningen av kjøresimulatorstudier: *Simulatorsyke* oppstår fordi det ikke er samsvar mellom synsstimuli fra kjørescenariet og bevegelsesstimuli i simulatoren. *Blinding av intervensjonen* er viktig i eksperimentelle studier, men vanskelig å oppnå med alkohol. Vi fant imidlertid ingen holdepunkt for at verken simulatorsyke eller forventninger om ruspåvirkning påvirket graden av vingling i kjørebanen.

Arbeidet som er presentert i denne avhandlingen gjør det mulig å bruke kjøresimulatoren ved NTNU/SINTEF i studier av ruspåvirket kjøring. Resultatene bekrefter at grad av vingling i kjørebanen er et valid, sensitivt og robust utfallsmål i slike studier.

Kandidat: Arne Helland
Institutt: Institutt for laboratoriemedisin, kvinne- og barnesykdommer, NTNU
Hovedveileder: Lars Slørdal
Finansieringskilder: Norges Forskningsråd, St Olavs hospital og NTNU

Avhandlingen er funnet verdig til å forsvares offentlig for graden *Doctor Philosophiae* (ph.d) i klinisk medisin. Disputas finner sted i Medisinsk-Teknisk Forskningssenter på St. Olavs Hospital/NTNU fredag 28. oktober 2016 kl. 12.15.

Contents

Acknowledgments

This work has been carried out at the Department of Clinical Pharmacology at St. Olav University Hospital and the Institute of Laboratory Medicine, Children's and Women's Health at the Norwegian University of Science and Technology (NTNU), in collaboration with the SINTEF research organization in Trondheim, Norway.

My first thanks go to the study participants, who found time and motivation to take part in the study, and uncomplainingly completed the rather demanding trials both on the test track and in the simulator.

I am very grateful to my primary supervisor, Professor Lars Slørdal, for introducing me to the field of forensic toxicology and traffic medicine, for including me in the research project, and for sharing his broad and profound knowledge. His devotion, positive attitude and resolve helped me a lot during the long process that has led up to this thesis.

I would also like to thank my co-supervisor at SINTEF, Gunnar D. Jenssen, for providing insights into the technical and behavioral psychology aspects of driving simulation, which are not in my field of expertise.

The papers constituting this thesis has involved co-authors from several disciplines. I would like to extend my thanks to all of them for contributing expert knowledge within their fields. I especially want to mention Professor Jørg Mørland at the National Institute of Public Health in Oslo for sharing his expertise in the field of DUI research, Professor Stian Lydersen at NTNU for helping me understand and apply statistical methods, and Lone-Eirin Lervåg at SINTEF for answering many questions regarding the study protocol and execution.

The Department of Clinical Pharmacology has generously provided me with time for my research, and the Institute of Laboratory Medicine, Children's and Women's Health at NTNU kindly awarded me a grant to finish my work and write the thesis. I am grateful to all my colleagues at the clinical pharmacology department for providing such a stimulating, supportive and fun work environment. Special thanks to Head of department Trond Oskar Aamo, as well as Ketil A. Espnes and Andreas A. Westin for stepping in and managing much of my work obligations during my leave to write the thesis.

I would also like to thank my family and friends, who have supported me during the years it took to complete this work. In particular, I want to mention my good friends Bjørn-Thore and Aleksander, for being such constant encouraging and cheering presences in my life. Aleksander recently defended his own Ph.D. (he beat me to it!), and was able to give valuable advice during my final stage.

Finally, I am grateful to you, dear Sébastien, for caring for me, supporting me and believing in me even in my most irrational and unlovable moments. I look back at the time we spent in France while I was writing the thesis with special fondness. Thank you.

Trondheim, June 2016

Arne Helland

Abbreviations, expressions and acronyms

a priori	Given before the fact; independent of experience
ADH	Alcohol dehydrogenase
ADHD	Attention deficit hyperactivity disorder
ALDH	Aldehyde dehydrogenase
ANOVA	Analysis of variance
BAC	Blood alcohol (ethanol) concentration
BrAC	Breath alcohol (ethanol) concentration
C_{max}	Maximum drug concentration in blood after an administered dose
CNS	Central nervous system
CTI	Clinical test of impairment
CYP	Cytochrome P-450 enzyme family
DUI	Driving under the influence
GABA	Gamma-aminobutyric acid
gold standard	The best available test to which others are compared
NTNU	Norwegian University of Science and Technology
OR	Odds ratio
per se	By itself; in itself
p-value	Probability of a result equal to or 'more extreme' than that observed, when the null hypothesis is true
RCT	Randomized controlled trial
RR	Relative risk
SD	Standard deviation
SDLP	Standard deviation of lateral position
SDS	Standard deviation of speed
SINTEF	SINTEF research organization
SSQ	Simulator sickness questionnaire
SSS	Simulator sickness severity
T_{max}	Time from drug administration to maximum drug concentration in blood
VALIDAD	Validation of a driving simulator to assess drug effects (research project)
V_d	Volume of distribution
Verum	Verification (positive control) substance with known effects
WHO	World Health Organization
Widmark equation	Equation that allows the prediction of BAC from an ingested ethanol dose when sex and body weight are known
Z drugs	The benzodiazepine-like hypnotics zopiclone, zolpidem and zaleplon
ΔSDLP	Difference in SDLP from baseline values

List of papers

Paper I

Comparison of driving simulator performance with real driving after alcohol intake: A randomised, single blind, placebo-controlled, cross-over trial

Arne Helland, Gunnar D. Jenssen, Lone-Eirin Lervåg, Andreas Austgulen Westin, Terje Moen, Kristian Sakshaug, Stian Lydersen, Jørg Mørland, Lars Slørdal

Published in Accident Analysis and Prevention 2013; 53: 9–16

Paper II

Evaluation of measures of impairment in real and simulated driving: Results from a randomized, placebo-controlled study

Arne Helland, Gunnar D. Jenssen, Lone-Eirin Lervåg, Terje Moen, Thomas Engen, Stian Lydersen, Jørg Mørland, Lars Slørdal

Published in Traffic Injury Prevention 2016; 17: 245–50

Paper III

Driving simulator sickness: Impact on driving performance, influence of blood alcohol concentration, and effect of repeated simulator exposures

Arne Helland, Stian Lydersen, Lone-Eirin Lervåg, Gunnar D. Jenssen, Jørg Mørland, Lars Slørdal

Published in Accident Analysis and Prevention 2016; 94: 180–87 (online: June 17, 2016)

Paper IV

Effectiveness of ethanol blinding by use of a novel sham pill approach, and the impact of drug expectancy on lateral vehicle control in real and simulated driving

Arne Helland, Stian Lydersen, Jørg Mørland, Lars Slørdal

Submitted to the Journal of Studies on Alcohol and Drugs May 6, 2016

1 Introduction

1.1 Driving under the influence (DUI)

Driving is a complex task that requires the integrated efforts of several cognitive functions and behavior patterns to be executed safely (Mitchell 1985). Psychoactive drugs, including ethanol, illegal drugs of abuse, and medicinal drugs, may interfere with these functions and thereby impair driving ability (Ogden and Moskowitz 2004). Drug-impaired driving is a leading cause of traffic accidents, with enormous costs to individuals and society (WHO 2013). Much of what we know about the effects of drugs on driving and traffic safety come from epidemiological studies. However, epidemiological approaches cannot establish causal relationship, and are vulnerable to bias and confounding. Experimental studies are thus also needed to extend our knowledge of drug-impaired driving.

Three hierarchical levels of behavior relevant to traffic safety, with related skills that should be measured in drugged driving research, have been defined (figure 1–1) (Michon 1985, Walsh *et al.* 2008). At the control level, we find automatic action patterns (e.g. maintaining lane position) that require little or no conscious effort. The maneuvering level corresponds to controlled action patterns (e.g. maintaining distance, maintaining constant speed) that require some sort of conscious mental processing. At the top, the strategic level represents the ability to make general plans and decisions (e.g. speed choice, route planning).

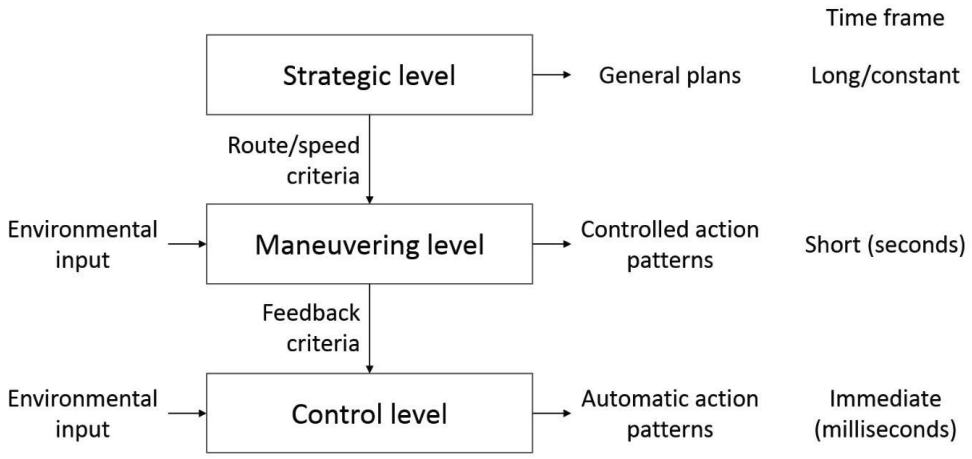

Figure 1–1. Hierarchical model of driving behavior, adapted and simplified from Michon (1985).

Another categorization, describing five cognitive domains essential to driving ability, has been put forward by Kay and Logan (2011): (1) alertness/arousal; (2) attention and processing speed; (3) reaction time/psychomotor functions; (4) sensory-perceptual functions; and (5) executive functions. In this thesis, the theoretical framework of Michon (1985) is used.

Characterization of drug effects on driving should ideally comprise all relevant behavioral levels and cognitive domains, which requires the assessment of drug effects on several outcomes, and in various settings (e.g. laboratory, driving simulators and on-road driving tests) (Walsh *et al.* 2008). Standardized measurements of cognitive functions in the laboratory have the advantage of being easily available, easy to set up and cost-effective, but direct inferences about driving impairment and traffic hazard cannot readily be made from such studies. A two-step model has been proposed, with simulator or on-road studies being carried out should the drug exhibit an impairing potential initial laboratory screening (Jongen *et al.* 2016). Driving simulators and on-road tests are more suitable to test a drug's actual impairing effects on driving than laboratory tests, since they better represent the full complexity of the driving task. However, testing facilities are scarce, and such studies may be prohibitively expensive and complex to carry out. For driving simulators, adequate validation is essential to ensure that results are applicable to real-life conditions.

Driving under the influence of ethanol and drugs is the main theme of this first chapter. Ethanol is described most thoroughly, since it is the best-characterized substance to cause drug-impairment – which was the rationale for the use of ethanol as test drug in the validation study described in this thesis.

1.1.1 Ethanol

Ethanol, commonly referred to as alcohol, is our most widely used drug of abuse. In most of the Western world, ethanol consumption is a ubiquitous and deeply ingrained part of the culture, often as the only legally available drug of abuse apart from nicotine. One thing that distinguishes ethanol from other drugs of abuse is its omnipresence, often being consumed with food and regarded as part of the diet and everyday life rather than as a psychoactive substance. In Norway, recent surveys show that more than 90 % of the adult population drinks ethanol to some degree, and the average adult consumes alcoholic beverages

corresponding to 8 liters of pure ethanol per year, equal to 17–18 g ethanol per day (Statens institutt for rusmiddelforskning 2015).

Pharmacokinetics

The pharmacokinetics of ethanol in a forensic context has recently been extensively reviewed (Jones 2011). Ethanol is a small, polar, water-soluble molecule that passes biological membranes easily. After oral ingestion, some absorption takes place in the stomach, but the rate of uptake is much higher in the duodenum. Absorption starts almost immediately after ingestion, and is nearly complete, although the time (T_{max}) from ingestion to maximum concentration in the blood (C_{max}) may vary considerably, from as low as 5–10 minutes up to 2 hours. The single most important determinant of the rate of absorption is the rate of gastric emptying, which is typically rapid when drinking on an empty stomach, and delayed when ethanol is ingested during or immediately after a meal. Some pre-systemic metabolism, commonly referred to as the first-pass effect, may occur through oxidation via alcohol dehydrogenase (ADH) enzyme present in the gastrointestinal mucosa and during the first pass through the liver. After uptake, ethanol distributes rapidly in the total body water compartment, corresponding to approx. 0.6 l/kg in women and 0.7 l/kg in men, but with considerable individual variation. Less pre-systemic metabolism and smaller volume of distribution (V_d) accounts for the higher blood alcohol concentrations (BACs) achieved in women after ingestion of similar weight-adjusted doses. On the other hand, ethanol elimination is slightly faster in women than in men. Elimination occurs primarily by oxidative metabolism via ADH in the liver to acetaldehyde, which is further metabolized to acetate by aldehyde dehydrogenase (ALDH). Up to 50 % in Asian populations lack active ALDH due to a genetic polymorphism, which leads to accumulation of acetaldehyde and the consequent occurrence of unpleasant side effects (e.g., facial flushing, hypotension, tachycardia) after ethanol intake. Ethanol is also a substrate for the hepatic cytochrome P-450 (CYP) enzyme 2E1. This pathway is usually minor, but becomes more important in heavy drinkers due to the induction of CYP2E1. This implies that ethanol induces its own metabolism, and explains the faster elimination of ethanol observed in heavy drinkers compared to moderate drinkers or teetotalers. The enzymatic breakdown of ethanol is saturated at relatively low BACs, which explains why ethanol displays zero-order kinetics at BACs above 0.1–0.2 g/l. The population average elimination rate is 0.15 g/l per hour, being slightly higher in women than in men, and

can be as low as 0.10 g/l per hour and as high as 0.25 g/l per hour in healthy individuals. Persons with chronic excessive ethanol consumption show the highest elimination rates, sometimes exceeding 0.30 g/l per hour.

Pharmacodynamics

The pharmacodynamics of ethanol has been reviewed by several authors, e.g. Naharashi *et al*. (2001), Vengeliene *et al*. (2008) and Spanagel (2009). Early research suggested that ethanol exerts its effects by stabilizing neuronal cell membranes, but this has later been shown not to be of importance except at very high BACs (Spanagel 2009). Ethanol has a complex mechanism of action in the central nervous system (CNS), affecting several neurotransmitter systems. Notably, ethanol enhances GABAergic transmission, inhibits glutamatergic transmission, and exerts a sympatholytic effect. The enhancement of GABA transmission accounts for the central nervous depressant properties of ethanol, with sedation and psychomotor retardation increasing progressively with higher BAC until stupor, coma, and eventually, death from respiratory depression. Decreased glutamatergic activity is probably related to the adverse effects of ethanol on cognition. Regular excessive drinking causes homeostatic adaptations in these systems, leading to tolerance, which reveals itself as a well-defined withdrawal syndrome upon the cessation of drinking after an extended period of frequent intake. Symptom severity depends on the degree of exposure and ranges from mild and mainly psychological symptoms (i.e., dysphoria, anxiety and sleeplessness), to serious and possibly fatal manifestations, with hallucinations and agitation as well as somatic symptoms of autonomous dysregulation (i.e., hyperthermia, hypertension, tachycardia). Complications such as alcoholic seizures and delirium may also ensue.

Another effect of ethanol, which it has in common with other addictive drugs, is the increased dopamine release in the mesolimbic system. Dopamine release is substantial already at low BACs, and may even start with just the prospect of having a drink. This is associated with the pleasurable and euphoria-inducing effects of ethanol, as well as the development of addictive behavior. The sustained activation of the brain's motivational and reward systems by prolonged, heavy drinking leads to longstanding and possibly permanent changes in brain circuitry, which may explain the persistent risk of relapse in ethanol dependent individuals even after years of abstinence (Vengeliene *et al.* 2008, Volkow *et al.* 2016).

The broad neurochemical effects of ethanol in the CNS probably account for the variable subjective effects in humans, ranging from euphoria to depression, sociability to hostility, and excitement and talkativeness to lethargy and sedation. Subjective effects of the same BAC on the ascending and the descending limb of the BAC-time curve have consistently been shown to be quite different, with stimulating effects dominating on the ascending limb and sedative effects dominating on the descending limb (Martin and Moss 1993). However, drinkers are less subjectively aware of their impairment on the descending limb, which may lead to misguided self-judgment of fitness to drive (Weafer and Fillmore 2012). This is often overlooked in studies of ethanol and driving.

Ethanol effects on driving performance and traffic safety

The broad spectrum of effects of ethanol on brain function may also explain why ethanol impairs almost every imaginable CNS function relevant to traffic safety, including perception, attention, coordination, and cognition (Mitchell 1985). There is no threshold value below which no impairment occurs; however, the sensitivity to ethanol effects varies between different measures of impairment (Ogden and Moskowitz 2004). Generally, complex tasks such as driving simulator or actual driving tests, or tests of inhibitory control or divided attention, are more sensitive to impairment from ethanol than simple tasks such as reaction time measurements or simple tracking tests (Jongen *et al.* 2016, Ogden and Moskowitz 2004). Ethanol has been shown to impair function, measured by a multitude of variables, on all behavioral levels or driving ability domains considered crucial for safe driving (Jongen *et al.* 2016, Ogden and Moskowitz 2004). Moreover, ethanol is by far the best-documented drug to induce driving impairment. For the reasons above, ethanol is considered a suitable benchmark drug or 'positive control' to assess the sensitivity of tests to detect drug impairment (Jongen *et al.* 2016, Walsh *et al.* 2008). Drug effects comparable to a BAC of 0.5 g/l are generally considered moderate, whereas drug effects comparable to a BAC of 0.8 g/l or higher are classified as severe in terms of traffic risk (Jongen *et al.* 2014).

The scientific framework of the effects of ethanol on driving performance is closely linked to two eminent scientists: the Swede Erik M. P. Widmark (1889–1945) and the American Robert F. Borkenstein (1912–2002) (Andreasson and Jones 1995, Voas 2003). They both stand out for their contributions to the scientific basis for limiting ethanol use in conjunction with driving, as well as being active proponents of legal enforcement. In 1922, Widmark introduced the

microdiffusion method of blood alcohol quantification, which allowed reliable and legally valid measurement of blood alcohol concentration and paved the way for later studies of the pharmacokinetics of ethanol. This analytical method measured BAC in mass/mass (w/w) units, i.e. mg ethanol per g whole blood, which corresponds to parts per thousand or per mille (‰). The insights from Widmark's research into the pharmacokinetics of ethanol enabled the estimation of the quantity of ingested ethanol from BAC measurement. This provided the necessary link between BAC and the level of intoxication and impairment. Norway was the first country to introduce a *per se* legal BAC limit for driving in 1936, at 0.5 ‰ w/w (corresponding to 0.5 mg/g, or ≈ 0.5 g/l). Sweden followed in 1941, with a limit of 0.8 ‰ w/w (Jones 2011).

Early adaptation of BAC limits probably accounts for the Nordic countries and Germany still using w/w units when reporting BAC in a legal context. Most other countries in Europe as well as the English-speaking world adopted BAC limits at a later stage, after the introduction of modern analytical methods such as gas chromatography. Consequently, these countries report mass/volume (w/v) units (e.g., g/dl, mg/ml or g/l). This can lead to some confusion. For most practical purposes, w/w units and w/v units may be used interchangeably (e.g., mg/g ≈ g/l). However, the density of whole blood is slightly higher than unity, at 1.055 g/ml. Hence, the BAC value is 5.5 % higher when expressed as w/v compared to w/w. For example, a BAC of 2.0 mg/g corresponds to 2.11 g/l (Jones 2011).

Although the link between ethanol inebriation and traffic risk was evident, the exact relationship between BAC and relative traffic accident risk was unknown when the Nordic countries adapted traffic impairment legislation based on BAC determination. Legislators in other countries were reluctant to introduce *per se* BAC limits in traffic without solid evidence linking the BAC limit to a quantitative measure of traffic hazard. Thus, to convict an intoxicated driver, the legal systems relied upon evidence from direct observations of the individual's incapacitation or impairment in addition to a measured BAC. The laborious and time-consuming process of obtaining a blood ethanol measurement also hindered efficient law enforcement in the area. Two achievements of Robert F. Borkenstein were crucial to overcome these hindrances.

In 1954, Borkenstein invented the Breathalyzer™, a device that measured the breath alcohol concentration (BrAC). The device was portable, easy to use and gave prompt results. This eliminated the need for trained health personnel to collect a blood sample and the wait for laboratory results, and facilitated widespread alcohol testing of drivers (Voas 2003). The BrAC/BAC ratio is approx. 1/2100 when concentrations are given in w/v units, i.e. a BrAC of 0.5 mg/l corresponds to a BAC slightly higher than 1 g/l. However, translating from BrAC into the corresponding BAC in individual cases is not recommended, as the relationship varies depending on the stage of metabolism as well as a number of other factors (Jones 2011). Most countries have adopted statutory BrAC limits that are independent of conversion to BAC to avoid this problem.

Borkenstein also conducted a large-scale case-control study, known as 'the Grand Rapids study' (Borkenstein *et al.* 1974), in which he compared the BAC of drivers involved in traffic accidents to the BAC of drivers at the same place and time of the day, but who did not have an accident. The study established the relationship between BAC and relative crash risk, and provided the scientific evidence needed to adopt *per se* BAC limits. Borkenstein's classic study was later replicated with a more refined methodology in the late 1990s ('the Long Beach/Fort Lauderdale study') (Blomberg *et al.* 2009). Borkenstein's findings were found largely to hold true for low-to-moderate BACs, whereas the relative risk (RR) estimates for BACs higher than 1.0 g/l in the Grand Rapids study were found to be underestimated. Other large, case-control studies both from America (Voas *et al.* 2012) and Germany (Krüger and Vollrath 2004) have shown similar results. The data show that traffic accident risk is a sharply rising exponential function of BAC (figure 1–2). Subgroup examination of the Long Beach/Fort Lauderdale data showed substantially higher BAC-related risks in the youngest drivers (Peck *et al.* 2008).

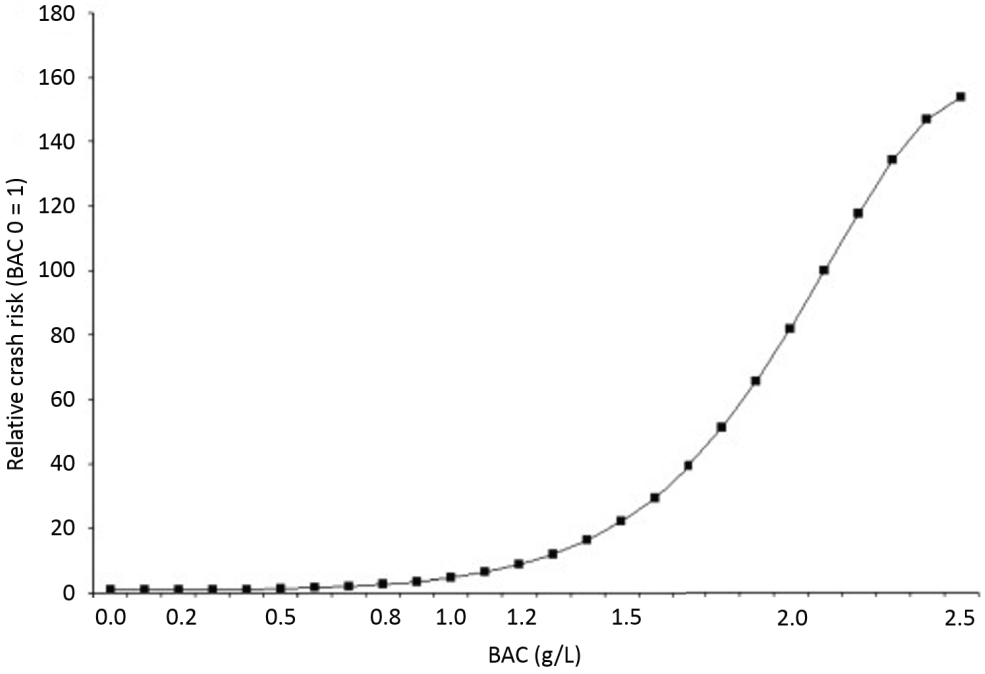

Figure 1–2. Relationship between BAC and relative crash risk. Adapted from Blomberg *et al*. (2009).

In the Long Beach/Fort Lauderdale study, a statistically significant increase in relative accident risk was observed at a BAC of 0.5 g/l, with a 38 % RR increase compared to sober driving. At a BAC of 0.2 g/l, corresponding to the legal driving limit in Norway, the RR only increased by a non-significant 3 %. Already at a BAC of 1.0 g/l, the RR increased five-fold, whereas BACs higher than 2.5 g/l were associated with a 150-fold increase in accident risk (Blomberg *et al.* 2009). One important limitation of these studies is that they do not separate the acute impairing effects of a high BAC from the chronic changes in psychomotor and cognitive abilities caused by heavy long-term alcohol consumption, which is much more likely in the cases driving with very high BACs than in the controls (Gjerde *et al.* 1986, Mørland 2000). Therefore, the very high relative accident risks associated with high BACs are likely to be a product of both the acute BAC effect and effects associated with chronic ethanol abuse.

1.1.2 Non-ethanol drugs

Any drug with somatic or psychoactive effects may potentially influence driving performance. The hazards of driving under the influence of non-ethanol drugs came to the attention much later than ethanol, but were starting to become apparent from both epidemiological and

experimental research during the 1970s. Methods for the quantification of drugs other than ethanol in blood or other biological fluids were largely unavailable until the 1950s, and were for a long time too cumbersome and inaccurate to be used for routine screening or legal purposes. By the 1970s, immunological tests enabling mass screening had been developed, and studies of the relevance of non-alcohol drugs to safe driving started to be published (Christophersen *et al.* 2016). As an example of early experimental research, a study that investigated psychomotor skills and visual functions with relevance to driving found significant impairing effects of three different benzodiazepines, and concluded that individuals should refrain from driving for a time after the ingestion of such drugs (Seppala *et al.* 1976). On the epidemiological side, one early case-control study found that drivers killed or seriously injured in car crashes were five times more likely to have been prescribed a minor tranquillizer (i.e., a benzodiazepine or related compound) than matched controls (Skegg *et al.* 1979). A multitude of potentially impairing drug classes were identified, including mainly psychoactive substances such as anxiolytics, hypnotics, stimulants, hallucinogens, cannabis, lithium and narcotic analgesics, as well as non-psychoactive drugs such as ganglionic blocking agents, insulin and sulphonylurea derivates (Seppala *et al.* 1979).

Drugs and drug effects relevant to traffic safety

Although several drugs without any direct psychoactive effects may cause impaired driving, i.e. antidiabetic drugs that may precipitate hypoglycemic attacks, or anticholinergic drugs that may disturb visual acuity through impaired accommodation (Seppala *et al.* 1979), the main concern rests on psychoactive drugs (Walsh *et al.* 2004). Their deleterious effects on driving may coarsely be divided in two main groups: stimulant effects and depressant effects. The CNS depressants, such as benzodiazepines, cannabinoids, opioids and centrally acting antihistamines, are a diverse group with different mechanisms of action. However, they all tend to reduce alertness, dull sensory perception, slow cognitive processes, lengthen reaction time, and cause somnolence. Psychomotor function, i.e. coordination and movement, is also impaired. The CNS stimulants on the other hand, such as amphetamine and cocaine, typically confer feelings of increased energy, talkativeness, restlessness and creativity. Their negative effects on traffic safety include disinhibition, impaired error monitoring, impulsivity, risk taking and aggression. After extended use of high doses, disorganized thinking, confusion and psychotic symptoms may ensue. In late phases after intense use of such drugs, the depletion

of monoamine neurotransmitters in brain circuits may in fact also lead to CNS depressant effects such as fatigue, somnolence and reduced psychomotor control (Ogden and Moskowitz 2004). Some psychoactive drugs with presumed impairing effects do not easily fit into the stimulant-depressant dichotomy, such as the hallucinogenic drugs (e.g. LSD, psilocybin).

Epidemiological (observational) studies

As drug quantification in biological samples became ever more available, efficient and accurate, systematic studies of drug use among suspected drug-impaired drivers, random drivers and the accident risk associated with several drugs and drug classes have been performed (Christophersen *et al.* 2016). Epidemiological studies are crucial to identify drugs and drug classes associated with traffic risks and to describe the use of drugs in the actual driving population and its relation to traffic accidents, injury and death – in other words, to map the real-world circumstances of driving under the influence of drugs. Purely descriptive epidemiological studies, e.g. cross-sectional studies of the presence of drugs in the blood of random drivers or drivers involved in traffic accidents, may give valuable information about the prevalence of drugged driving, but do not allow any inferences of causality (Mørland 2000). On the other hand, analytical epidemiological studies such as case-control studies or cohort studies, if well-designed and -controlled, may allow the estimation of risks associated with drugs and also indications of causal relationships, although the latter cannot be proved in such studies (Mørland 2000).

Perhaps the most central study design is the case-control study, in which 'cases', i.e. drivers having caused an accident, are compared to 'controls', i.e. comparable drivers who did not cause an accident (Gjerde *et al.* 2015, Houwing *et al.* 2009). Careful steps are taken to ensure that the cases and the controls are as closely matched as possible. For example, controls may be randomly stopped drivers at the same place, at the same time of day and year, and travelling in the same direction as the 'case' drivers who caused an accident. The odds of the 'cases' being under the influence of a drug is then divided by the odds of the 'controls' being under the influence of the same drug, to obtain the odds ratio (OR), which is the main statistic to describe risk in case-control studies. It is important to note that the odds ratio is not the same as the relative risk (RR); in fact, case-control studies do not permit the computation of relative risks. However, if the prevalence of the outcome under study is low, then the OR approaches the RR. This is the case for road traffic accidents, which means that in DUI case-

control studies, the OR is a good approximation to the relative risk associated with driving under the influence. This is beneficial, since ORs and RRs are often confused, which may lead to gross overestimation of the perceived risk when studying prevalent outcomes.

Other types of analytical epidemiological studies exist. Culpability studies are a subcategory of case-control studies, which assess the influence of alcohol and/or drug use on the likelihood of drivers being deemed to have caused an accident. For instance, a large Australian culpability study of more than 3,000 drivers killed in traffic accidents reported that those drivers under the influence of psychotropic drugs in general, and cannabis and stimulant drugs in particular, were at higher odds of having caused the accident (Drummer *et al.* 2004). Register-based studies may combine data from for instance prescription databases and traffic accident/death registers to calculate the accident risk associated with recently having filled a prescription for a certain drug. In such studies, traffic accident incidence in the exposed periods are compared with that of the unexposed periods to calculate standardized incidence ratios (SIR). As an example, a Norwegian study reported significantly higher traffic accident risks as evidenced by SIRs significantly higher than unity the first week after having filled a prescription for Z drugs (zolpidem or zopiclone), and the benzodiazepine hypnotics nitrazepam and flunitrazepam (Gustavsen *et al.* 2008).

Epidemiological as well as experimental research has demonstrated the ability to impair driving of six major drug classes: (1) Benzodiazepines and related drugs (i.e., benzodiazepine-like hypnotics (Z drugs), barbiturates); (2) cannabis; (3) opioids; (4) amphetamines and other stimulants; (5) antihistamines; and (6) antidepressants (Mørland 2000, Walsh *et al.* 2004). A recent review of epidemiological research in the field concluded that among non-alcohol drugs, the use of amphetamines confers the highest risk of traffic accident involvement. Increased risk is also well documented for benzodiazepines (incl. Z drugs), cannabis, opioids, other stimulants such as cocaine, and some antidepressants. Some other drugs such as phencyclidine (PCP) and carisoprodol have also been shown to confer substantial traffic risk (Gjerde *et al.* 2015). However, when assessing dichotomous data (presence vs. absence of drug in sample), ethanol is still the single drug that by far confers the highest accident risk. Polydrug use is generally associated with significantly higher risk than single drug use, the riskiest combination being that of ethanol + another drug (Bogstrand *et al.* 2012, Gjerde *et al.*

2011). One case-control study from Norway of 204 fatally injured drivers found the following adjusted odds ratios for ethanol, medicinal and illicit drugs and different combinations: Ethanol + drug(s) 350; ethanol alone 70; two or more illicit drugs 50; two or more medicinal drugs 17; single illicit drug 6; single medicinal drug 1.7 (Gjerde *et al.* 2011). These are crude categories and the numbers only represent one study; however, it is reasonable to assume that they convey an important insight into the relative risk differences between different types of drugs and combinations. It is vital to remember that the very high risk estimates from case-control studies are not a result of the 'pure' drug effect, but rather a product of the acute drug exposure plus all the behavioral, personality, physical and mental health factors associated with it that have not been adjusted for in the analysis (Gjerde *et al.* 2015).

For a number of reasons, it is difficult to study the dose- or concentration-dependent effects of non-alcohol drugs on driving in epidemiological studies (Gjerde *et al.* 2013). Case-control studies on non-alcohol drugs require screening and quantification of a large number of potentially impairing drugs, as well as a large number of cases, as each drug has a relatively low prevalence of detection. In addition, blood sampling for drug testing of controls – as compared to simple breath tests in ethanol studies – is necessary, which makes the recruitment of controls more difficult. Furthermore, post-mortem drug concentration changes occur to a large degree for non-alcohol drugs, making interpretation of drug concentrations in killed drivers difficult. Because of these difficulties, the relation between blood concentrations of non-alcohol drugs and crash risk is difficult to establish from epidemiological research. It is also important to remember that epidemiological studies cannot prove causal relationships, only associations, and they are vulnerable to selection bias and confounding factors (Gjerde *et al.* 2015). The group of individuals driving under the influence of drugs is a highly selected subpopulation, which is always a cause of disparity in epidemiological research. Especially the use of illegal/hard drugs is associated with multiple possible risk factors of unsafe driving, such as risk-taking personality traits, cognitive impairment or psychiatric and somatic comorbidity either preceding or being a consequence of chronic drug use (Mørland 2000). For instance, in case-control studies, a driver on methamphetamine causing a crash very likely differs from the 'matched' control not causing a crash in several important ways. Thus, what the odds ratios actually express is the overall traffic risk associated with being a user of a particular drug and all it entails, and not the

physiological drug effect *per se*. Well-designed epidemiological studies may adjust for known confounders, which permits the reporting of adjusted risk estimates, typically showing substantially lower risk than the crude ratios. Nevertheless, it is impossible to eliminate the uncertainty introduced by unmeasured or unknown confounders.

Experimental studies

In light of the limitations of the observational approach, experimental studies are crucial to investigate the impairing effects of drugs to isolate the 'pure' drug effect and establish causality as well as the quantitative relationship between drug concentrations, impaired performance and possible accident risk. In a randomized, controlled design, it is possible to isolate the effect of the factor under study (i.e., the effect of a certain blood level of a drug) on a measure with relevance to traffic safety. In the last couple of decades, many such studies have been performed (see chapter 1.1.5). The combination of evidence from epidemiological and experimental research, such as has been put together for cannabis by Ramaekers *et al*. (2004), provides the supreme basis for *per se* legislation for non-ethanol drugs (see chapter 1.1.4). However, experimental studies are also fraught with limitations. In epidemiological studies, hard endpoints with obvious relevance to traffic safety such as road traffic crashes or deaths are typically studied. For obvious reasons, such outcomes are impossible in experimental research, which necessitates the use of surrogate endpoints with supposed relevance to traffic safety. A myriad of different outcomes have been studied, but few have been rigorously validated and calibrated against a known risk quantitation, which renders much experimental research difficult to interpret. Ethical and legal concerns also limit experimental studies of drugs of abuse, and in particular of illegal drugs. In some countries, experiments involving illegal drugs are impossible to perform due to juridical issues. Either when studying medicinal drugs, or illegal drugs where such studies are allowed, ethical or safety issues restrict researchers from administering doses of similar magnitude to those commonly ingested by recreational users. For instance, it has proved difficult to reproduce the apparent high traffic risk associated with central stimulants in experimental settings, where researchers are obliged to administer low doses (Silber *et al.* 2012). Moreover, several countries do not permit on-road experiments due to safety concerns. Driving simulators may overcome some of the legal and practical concerns with experimental studies of DUI.

Therapeutic versus recreational use

The assignment of risk to certain classes of drugs is complicated by the fact that many drugs may be used both therapeutically and recreationally. Whereas therapeutic use is often characterized by relatively low and stable doses, recreational use often involves the intermittent use of higher doses, often in combination with other drugs.

Long-term use of stable doses may induce tolerance to some of the impairing effects of the drugs in question, rendering therapeutic users less impaired even with drug dosages exceeding those that would otherwise be considered impairing in non-users. This particularly concerns the opioids, but may also apply to the benzodiazepines and related compounds (Vindenes *et al.* 2012). The question whether long-term use of benzodiazepines induce tolerance to their impairing effects on driving is not settled. Combined experimental and epidemiological data suggest that partial tolerance may develop, albeit slowly and inconsistently (Verster *et al.* 2004). One study showed a markedly increased risk of hospitalization due to traffic accidents shortly after having filled a prescription for benzodiazepines, which progressively decreased (but was still elevated) the first few weeks after the prescription was filled. This could be a sign of partial tolerance developing to the impairing effects, but the study design does not permit such a conclusion. The observed risk decrease may just as well be attributed to decreased exposure (Neutel 1995). Experimental studies exist that show no or only partial tolerance developing to the psychomotor effects of benzodiazepines (Manthey *et al.* 2014, Smiley and Moskowitz 1986, Staner *et al.* 2005), and a meta-analysis showed persisting cognitive deficits in chronic benzodiazepine users (Barker *et al.* 2004). With central stimulant use, some tolerance probably develops to the subjective effects, whereas tolerance to impairing effects on driving have not been investigated in experimental studies (Strand *et al.* 2016).

The manner of use likely influences the traffic risk significantly. Recreational amphetamine users, who inject high and often repeated doses to obtain euphoria, run a high risk of traffic accident involvement, whereas therapeutic low-dose amphetamine use, for example in the treatment of ADHD or to combat somnolence and fatigue, is not associated with a high traffic risk (Gjerde *et al.* 2015), and may even be performance-enhancing (Gobbo and Louza 2014), although this is not a universal finding (Hjälmdahl *et al.* 2012). Thus, factors such as setting (therapeutic < recreational), dose (low < high), administration route (oral <

inhaled/insufflated < intravenous) and pattern of use (stable/low dose < intermittent/high dose, single drug use < multi-drug use) are highly relevant to determine the actual traffic risk.

1.1.3 Prevalence of DUI

Cross-sectional studies of drug and alcohol use, either among the general driving population (i.e., questionnaire-based or roadside surveys) or in drivers involved or killed in traffic accidents, may give a descriptive impression of the frequency of driving under the influence of drugs, as well as of changing DUI patterns over time.

Roadside surveys are complicated and costly, but more reliable than questionnaire-based surveys, as the latter are very vulnerable to recall bias, deliberate underreporting and low response rates (Christophersen *et al.* 2016). Roadside studies have been performed for many decades for ethanol, whereas other drugs have not been investigated extensively until the last couple of decades, due to the previous lack of analytical techniques to detect and quantify a large number of drugs in a small sample volume that is typically collected in roadside surveys. The introduction of saliva analysis has greatly facilitated the undertaking of large-scale roadside testing of drugs other than ethanol.

In Norway, the first roadside survey of ethanol in random drivers was performed in 1970–71 and showed that 2 % of drivers randomly stopped between 10 pm and 2 am had a BAC higher than 0.5 g/l (the legal limit at the time). This share dropped to 1 % in 1977 and to 0.27 % in 1981-2 (Christophersen *et al.* 2016). Several studies after the turn of the millennium showed a further decrease in the rate of ethanol-positive drivers. In a large, nationwide survey conducted in 2008–9, only 0.2 % of all drivers had a BAC higher than 0.2 g/l and less than 0.1 % had a BAC higher than 0.5 g/l (Gjerde *et al.* 2013). Drugs other than ethanol that could influence driving ability were included in sizeable road surveys in Norway in 2005–6, 2008–9 and 2014–5, using oral fluid as test matrix (Gjerde *et al.* 2008, Gjerde *et al.* 2013, Jamt *et al.* 2015). Medicinal drugs were detected in 3.4 %, 3.2 % and 2.1 % of drivers, respectively, whereas illicit drugs were detected in 1 %, 1.5 % and 2.1 %, respectively. The apparent reduced detection of medicinal drugs and increased detection of illicit drugs is uncertain due to differences in sampling and analytical cutoffs between studies (Christophersen *et al.* 2016). When findings of non-ethanol drugs in random drivers were restricted to blood concentrations higher than the newly-adopted *per se*-limits in Norway, the prevalence was

considerably lower (0.4 % for illegal drugs and 1.1 % for medicinal drugs, respectively), but still higher than the proportion driving with a BAC higher than the 0.2 g/l limit (Gjerde *et al.* 2011). The drugs most commonly encountered in Norwegian drivers are benzodiazepines and Z drugs among the medicinal drugs, and amphetamines and cannabis among the illicit drugs.

When drivers killed in road traffic accidents were studied, the prevalence of BAC > 0.5 g/l dropped from more than 40 % in the 1970s to below 20 % in the 2010s, whereas the findings of other drugs seem to be rather stable at around 20 % of killed drivers (Christophersen *et al.* 2016).

Studies from other countries around the world such as Australia, the US, Spain and Brazil, generally show decreasing rates of driving under the influence of ethanol through the last decades. However, the prevalence varies considerably from country to country, and is highest in Brazil, followed by the US, Spain, and at the lower end Australia, which has a prevalence similar to Norway. As for other drugs, the trends vary, from an increase in the prevalence of drugs both in random drivers and killed drivers in Australia, via stable numbers in the US, to a reduction in Spain (Christophersen *et al.* 2016).

1.1.4 Legislation

Most countries in the world have adapted laws that prohibit driving under the influence of ethanol or other drugs that impair driving ability. In principle, the legislation may be based on an 'impairment approach' or a '*per se*' (analytical) approach. With the impairment approach, the prosecution must prove that the driver was impaired or intoxicated in order to get a conviction. This means that in addition to analytical proof of drug ingestion, corroborating evidence of impairment such as a failed field sobriety test, witness observations of erratic driving, or the causation of a traffic accident, is required. In contrast, with *per se* legislation, driving with a drug concentration (most commonly in blood) above a set limit constitutes an offence by itself (*per se* literally means 'by itself'). Some countries have also implemented zero-tolerance laws, mainly for illegal substances (Jones 2005). This can be regarded as a particularly strict version of *per se* legislation, and implies that driving with any detectable amount of a drug in blood (and, in some instances, even other matrices) constitutes a crime. *Per se* legislation and, in particular, the zero tolerance approach may be problematic when it comes to therapeutically used medicinal drugs.

Per se laws for ethanol have been adapted in most of the world, although the concentration limits vary considerably (WHO 2013). In most Western countries, the current BAC limit is 0.5 g/l. Some countries, such as Norway and Sweden, have lowered the legal limit to 0.2 g/l, despite a lack of convincing evidence that such low BACs significantly increase accident risk (Blomberg *et al.* 2009). Rather, the low limits in these countries could be seen as normative political signals that drinking and driving should not be combined (Christophersen *et al.* 2016). However, prominent scientists in the field have argued that legislators are at the liberty to prohibit driving at any BAC, since even very low limits would not contradict the scientific fact that there is no lower limit to impairment (Ogden and Moskowitz 2004). Brazil and some Eastern European countries have even introduced zero (or close-to-zero) tolerance limits for BAC. On the other end of the scale, the UK and the US so far have been reluctant to lower their 0.8 g/l limits, despite evidence that lower limits save lives (WHO 2013). Even so, US drivers may still be convicted of driving under the influence with a BAC below 0.8 g/l if impairment can be proven. It is possible that the different approaches to legislation and BAC limits reflect dissimilar juridical traditions between countries, with differing views as to whether laws should be implemented as a preventive measure. In recognition of the fact that the adverse effects of ethanol on driving ability are more pronounced in young drivers, some countries have implemented zero or close-to-zero limits for drivers below a certain age (WHO 2013).

In contrast to the well-defined relationship between BAC and crash-risk, the scientific evidence linking blood concentrations of other drugs to traffic risk is much sparser. For this reason, impairment-based legislation is still the rule for non-alcohol drugs. Impairment-based laws require a lot of effort and a trained and motivated police corps to be enforced efficiently, because impairment must first be suspected to get a blood test, and then proven in court to get a conviction. Contrary to public beliefs, drug impairment is often difficult to establish by clinical observation and testing. With ethanol, several studies have shown that many individuals are declared sober after having been subjected to standardized clinical test batteries even with BACs in the area of 0.15–0.20 g/l (Ogden and Moskowitz 2004). Other drugs with less consistent concentration-effect-relationships and qualitatively different CNS effects, i.e. central simulants, hallucinogens or cannabinoids, are even harder to detect by clinical means. Nevertheless, legal retributions for drivers under the influence of non-ethanol

drugs often rely heavily on clinical findings, e.g. the standardized Clinical test of impairment (CTI) in use in Norway. The CTI has not been extensively validated for non-ethanol drugs, although it has been shown to be sensitive to the effects of common sedative-hypnotic drugs such as benzodiazepines (Bramness *et al.* 2003) and Z drugs (Gustavsen *et al.* 2009).

With the advent of portable saliva drug screening devices, roadside screening for non-alcohol drugs is starting to become widespread in some countries (Musshoff *et al.* 2014). This facilitates the implementation of routine roadside screening strategies that have so far only been employed for ethanol. Furthermore, in recent years, several European countries as well as some US states have introduced *per se* legislation for common drugs of abuse (Christophersen *et al.* 2016). Norway introduced *per se* limits for 20 drugs of abuse and medicinal drugs in 2012, with a three-tiered system of limits for graded sanctions, representing drug concentrations in blood likely to induce impairment comparable to BACs of 0.05 % and 0.12 % for 13 of the 20 substances. This system functions along with an impairment-based law that comes into use should the suspect be under the influence of multiple drugs or medicinal drugs used in accordance with a doctor's prescription (Vindenes *et al.* 2012). The exemption of therapeutic use from the *per se* legislation necessitates the evaluation of drug levels in blood against a prescribed dose, which is not always straight forward (Jones *et al.* 2007). Sweden and Finland have implemented zero-tolerance limits for illicit drugs, combined with impairment-based legislation for medicinal drugs (Holmgren *et al.* 2007, Lillsunde and Gunnar 2005). Experience from Denmark and Sweden shows that the introduction of zero-tolerance laws may dramatically increase the number of DUI-apprehended drivers (Holmgren *et al.* 2007, Steentoft *et al.* 2010).

1.1.5 Experimental research

A wide range of behavioral factors, cognitive functions and psychomotor skills is relevant to safe driving. Many measures of such functions, e.g. reaction time or tracking tasks, are comparatively easy to study under controlled laboratory conditions, and may give an indication as to the possible impairing effects of a drug on driving. However, it has been recognized that such tests do not adequately represent the full complexity of driving a vehicle, and that more externally valid driving experiments, i.e. driving simulators or on-road driving tests, are necessary to establish the true drug effect on driving (Kay and Logan 2011,

Owens and Ramaekers 2009, Walsh *et al.* 2008). Establishing representative and sensitive tests of driving impairment due to centrally-acting drugs, suitable to experimental investigation, has been a research priority for many years (Irving and Jones 1992). Several review articles summarizing experimental research on driving impairment from non-ethanol drugs have been published. Among these, a review published in 2000 (Mørland 2000) with an update in 2016 (Strand *et al.* 2016) concluded that various central depressant drugs – benzodiazepines and related compounds and cannabis in particular, but also opioids and to some extent GHB and ketamine – have been shown to confer psychomotor impairment with relevance to traffic safety. The acute effects of low to moderate doses of central stimulants do not seem to cause impaired driving (Strand *et al.* 2016).

Laboratory testing of behavior, cognitive and psychomotor skills

A multitude of tests with possible relevance to driving has been employed to test the impairing effects of ethanol and drugs. Examples include simple reaction time, vigilance tasks, tracking, tests of visual functions and tests of divided attention (Ogden and Moskowitz 2004). Experimental settings to assess effects on behaviors such as risk willingness have also been developed (McMillen *et al.* 1989). The realization that tests of simple sensory, perceptual and motor functions, such as simple reaction time or critical flicker fusion, are not sensitive to drug impairment, caused a shift to examination of more complex measures of cognitive function, such as divided attention (Ogden and Moskowitz 2004). When carefully selected, it is possible to assess measures of psychomotor function that correspond to all three levels of behavior relevant to driving. As an example, in a randomized controlled trial (RCT) of zopiclone 5 and 10 mg, ethanol and placebo, behavior was tested at the automated level (e.g. simple reaction time), maneuvering control level (e.g. choice reaction time) and strategic level (e.g. Stockings of Cambridge task). The researchers administered three different computerized tests, each with several components corresponding to different behavioral levels. The results showed a larger effect 1 hour after intake of zopiclone 10 mg than after ethanol (BAC 0.74 g/l) on the automated behavior level, and comparable effects at the controlled and strategic levels. When tests were repeated 3.5 and 6.5 hours after intake, results normalized and were not significantly different from placebo 6.5 hours after intake of any active drug condition (Gustavsen *et al.* 2011). A recent literature review of the sensitivity of driving-related laboratory tests to ethanol suggested that initial screening tests for drug

impaired driving should: (a) be standardized; (b) be sensitive to the potential impairing effects of drugs; (c) have established reliability; (d) have validity supported by theoretical models of driving behavior and (e) be calibrated by benchmark drugs and doses to ensure comparability of results from various research settings (Jongen *et al.* 2016). The review concluded that the cued go/no go task, which assesses inhibitory control, and a divided attention-test with a primary tracking task and a secondary visual search task, were consistently sensitive to the impairing effects of ethanol at low BACs.

On-road driving research

O'Hanlon and colleagues developed a standardized on-road highway driving test in the Netherlands during the 1980s (O'Hanlon *et al.* 1982, O'Hanlon 1984). This has later often been regarded as the 'gold standard' to determine driving impairment caused by drugs. The model has been employed in numerous trials in the Netherlands, investigating drugs such as ethanol, other drugs of abuse, hypnotics, anxiolytics, antidepressants and antihistamines (Verster and Roth 2011). In the on-road driving tracking task, the subjects drive a regular car for approximately 1 hour over a 100 km stretch of dual carriageway in normal traffic. A licensed driving instructor is present in the car, which is equipped with a double set of controls for the instructor to be able to intervene if necessary. Test subjects are instructed to drive with a steady lane position and at constant speed. A camera mounted on the roof of the car constantly measures the vehicle's lateral position relative to the lane markings, whereas speed is measured by magnetic induction proportional to wheel rotation.

Standard deviation of lateral position (SDLP), which is a measure of lateral control ('degree of wavering') of the vehicle (figure 1–3), is the primary outcome in the on-road test. It is a thoroughly validated measure of driving impairment. SDLP is sensitive to ethanol effects, consistently showing statistically significant, concentration-dependent increases with increasing BAC (Jongen *et al.* 2016). Furthermore, SDLP has inherent validity as a measure of vehicle control, since increased wavering must necessarily correspond to increased risk of significant lane border crossings and, ultimately, increased risk of hitting objects on the side of the road or other traffic (Verster and Roth 2011). Indirect evidence of a strong correlation between increments of SDLP and relative traffic accident risk for ethanol and diazepam has also been shown by combining SDLP results from experimental studies with risk estimates from epidemiological studies (Owens and Ramaekers 2009). Drugs that impair driving as

measured by significant SDLP increases in the on-road test include benzodiazepines and Z drugs (Leufkens *et al.* 2007, Verster *et al.* 2006), antihistamines (O'Hanlon and Ramaekers 1995), antidepressants (Ramaekers 2003) and cannabis (Bosker *et al.* 2012). By comparing SDLP increments to those seen with ethanol, it is possible to estimate the degree of driving impairment from the drug under study (Louwerens *et al.* 1987, Verster and Roth 2011).

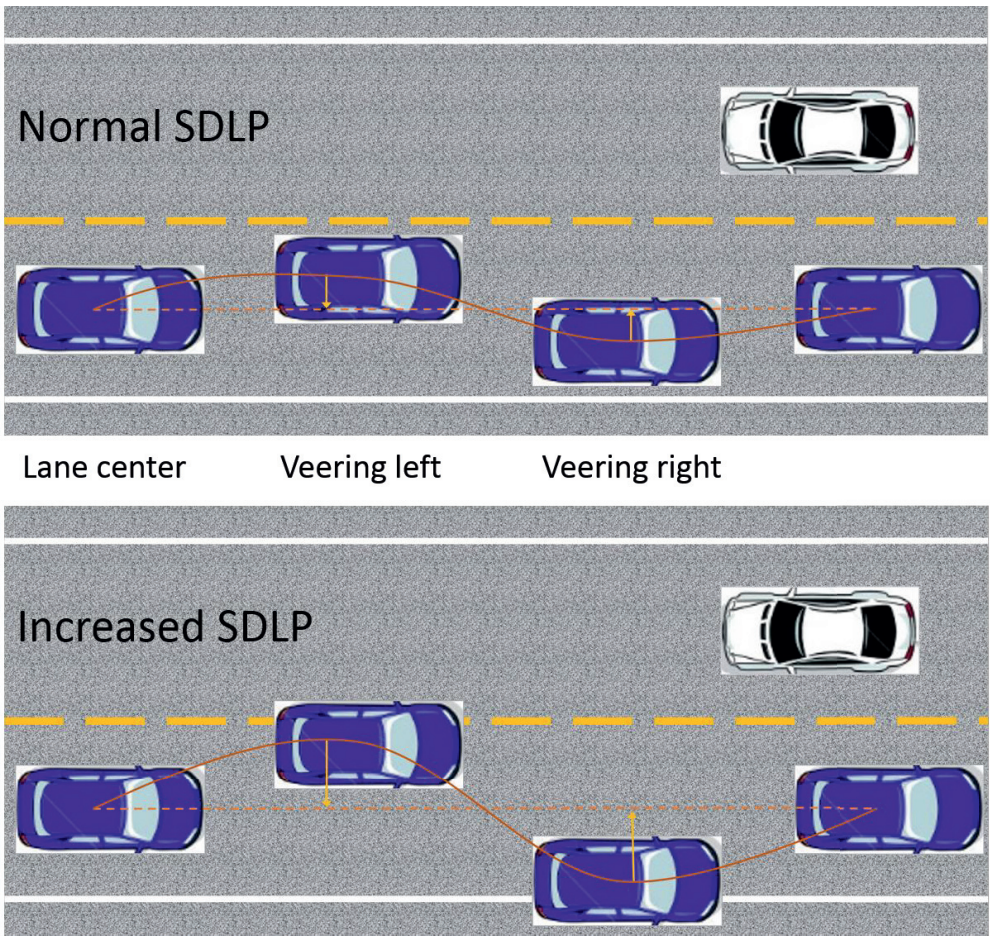

Figure 1–3. Illustration of the significance of the standard deviation of lateral position (SDLP) for lateral vehicle control. Modified from Verster and Roth (2011).

One limitation of the SDLP is that it only reflects the automatic control level, the most basic of the three levels of behaviors that are considered essential to safe driving (Michon 1985, Walsh *et al.* 2008). Furthermore, SDLP is mainly a measure of sedating and fatiguing effects, as shown also by the correlation between sleepiness and SDLP (Contardi *et al.* 2004). It is not

sensitive as a measure of the impairing effects of stimulating drugs, which may even decrease SDLP (Ramaekers *et al.* 2012). To meet these shortcomings, other driving tasks and measurements could be employed. Brookhuis *et al*. (1994) developed the car following task, in which participants are instructed to maintain a constant safety distance to a leading vehicle as it performs several accelerations and decelerations. The primary outcomes may be reaction time to the leading vehicle's change of speed, or headway distance/time (from the rear of the leading vehicle to the front of the test vehicle). This task measures the driver's ability to adjust to other traffic, and is thus a measure of the maneuvering (operational) level of driving behavior. A city driving task has also been suggested as an attempt to measure effects at the highest (strategic or tactical) level of driving behavior, such as risk-taking and reactions to sudden events (Veldstra 2014). In this task, subjects drive in an urban environment where they must undertake difficult maneuvers and interact with complex traffic. However, such driving tasks are difficult and possibly unethical to employ in real traffic because complex traffic scenarios are difficult to reproduce and potentially dangerous.

Driving simulator research

Driving simulators are designed to mimic actual driving and as such provide a realistic setting for the assessment of driving in the convenient, safe and controlled environment of the laboratory. The rationale for driving simulators in research on drug impaired driving is primarily to assess aspects of driving that are too risky to be studied in on-road experiments, such as risk-taking behavior and hazard avoidance (Jongen *et al.* 2016). However, since advanced driving simulators may provide a very good approximation of real driving, their use is by no means restricted to scenarios that cannot be assessed in on-road tests for safety reasons. Simulators may be used to measure impairment at all three main levels of driving behavior (Veldstra 2014). Driving simulators are described in detail in chapter 1.2.

Numerous simulator studies have investigated driving impairment due to various medicinal drugs as well as drugs of abuse, including cannabis (Hartman *et al.* 2015, Hartman *et al.* 2016, Lenne *et al.* 2010), benzodiazepines (Daurat *et al.* 2013), Z drugs (Staner *et al.* 2005), central stimulants (Hjälmdahl *et al.* 2012, Silber *et al.* 2012), antiepileptics (Kaussner *et al.* 2010), opioids (Lenne *et al.* 2003, Nilsen *et al.* 2011), caffeine (Mets *et al.* 2012) and antihistamines (Weiler *et al.* 2000). A few studies on drug combinations, mostly with ethanol, have also been published (Hartman *et al.* 2015, Simons *et al.* 2012, Thompson *et al.* 2010).

Possible risky states associated with drug use such as ethanol hangover (Verster *et al.* 2014) have also been shown to increase SDLP in simulated driving. There is also a growing interest in simulator studies of the effects of various disease states on driving, such as Parkinson's disease (Möller *et al.* 2002, Uc *et al.* 2009) or ADHD (Weafer *et al.* 2008), and even of disease-drug interactions, such as obstructive sleep apnea and ethanol (Vakulin *et al.* 2009).

Ethanol as positive control

Guidelines for research on drugged driving encourage the use of a reference or 'benchmark' drug to which impairment of other drugs may be compared. It is also recommended to include a placebo, as well as a 'positive control' or *verum* – a substance that is expected to elicit a response in the test that is used to prove its sensitivity (Walsh *et al.* 2008). Ethanol is suitable both for the reference drug and *verum* purposes due to its well-defined effects on real-life accident risk, as well as a multitude of laboratory and real-life tasks that measure psychomotor performance and behaviors important to safe driving (Jongen *et al.* 2016, Walsh *et al.* 2008). For the same reasons, ethanol is well suited as a study drug to validate experimental models of driving impairment. A test that presumes to measure responses relevant to driving safety should be sufficiently sensitive to be able to distinguish ethanol effects at fairly low BAC levels (Jongen *et al.* 2016). However, in studies of medicinal drugs without intoxicating effects, such as antidepressants, the appropriateness of ethanol as positive control is questionable because the subjective effects of ethanol will be very different from that of the drug under study. This may cause problems with intervention blinding, and in this setting, other drugs with established impairing effects on driving such as mirtazapine has been proposed as a more appropriate *verum* drug (Verster *et al.* 2015).

Challenges with ethanol blinding and the impact of expectancy

In studies of the effect of pharmacological interventions, a placebo condition is commonly included in the experimental design to separate drug effects from effects caused by receiving the intervention *per se*. Ethanol studies is no exception to this; however, placebo blinding is particularly difficult in studies of ethanol, especially with moderate to high doses, as the effects and cues of ethanol are easily recognizable to the participants (Sayette *et al.* 1994, Testa *et al.* 2006). Test subjects are often able to discriminate between even low alcohol doses (0.2 g/l) and placebo (Jackson *et al.* 2001). Nevertheless, placebo control may be even more important in ethanol studies than in other studies of drug effects, due to the strong

expectancy effects elicited by ethanol. Two major social-cognitive approaches exist to describe the link between ethanol ingestion and behavior: expectancy and myopia theories. The expectancy theory proposes that behaviors and impairments linked to ethanol consumption are largely results of the individual's expectations about the effects of drinking ethanol. In contrast to this, the myopia theory focuses on the pharmacological effects of ethanol in the brain, which reduce the individual's capacity to process information. This 'narrowing of the field of view' means that only the most salient information will be taken into account in decision-making. The two theories are not necessarily mutually exclusive, and there seems to be ample support for both theories from experimental studies (Moss and Albery 2009).

The 'balanced placebo design' is commonly viewed as the optimal study design to separate pharmacological and expectancy effects of ethanol. It crosses the beverage expectancy (i.e., being told that a drink contains ethanol or placebo) with the actual beverage content (i.e., getting ethanol or placebo) (Rohsenow and Marlatt 1981). Numerous studies have shown that this approach is effective in deceiving the subjects, who generally tend to believe that they have received what they have been told. However, the balanced placebo design is not well suited to within-subject testing, or to test several BAC levels, which may be desirable to establish a dose-response relationship in for instance impairment testing. The design may also constitute an ethical dilemma, since the researchers are obliged to actively deceive the subjects about not getting ethanol when they actually are.

Another approach is coined the 'alternative substance paradigm', in which participants are informed that the drink they receive may contain one out of several substances, or a combination of these. One study reported on the effectiveness of blinding using this approach, where the subjects were told that the drink may contain either ethanol, another drug or drugs, a combination, or placebo (Conrad *et al.* 2012). A special case of the alternative substance paradigm would be to administer the intervention in a 'double dummy' manner, using both a drink (placebo or ethanol) and a placebo pill.

As discussed in more detail in paper IV, an assumption of having ingested ethanol, or alternatively, other psychoactive drugs, in itself may have effects on subjective intoxication (Fillmore and Vogel-Sprott 1995, Starkey and Charlton 2014) and behaviors such as risk taking

(Burian *et al.* 2003, McMillen *et al.* 1989). Studies of expectancy effects on psychomotor measures have shown both impaired (Fillmore and Vogel-Sprott 1995) and improved (Finnigan *et al.* 1995) tracking performance in subjects expecting ethanol, as well as improved performance in a cued go/no-go task (Marczinski and Fillmore 2005).

Studies reporting ethanol expectancy effects in driving simulation tasks are few, and show somewhat diverging results (Charlton and Starkey 2015, Rimm *et al.* 1982). Expectancy effects have not been investigated for SDLP in real driving conditions. Since SDLP is an important outcome in experimental studies of drug impaired driving, it is prudent to explore whether expectancy effects may influence SDLP results, or if SDLP increments with rising BAC is mainly a consequence of the pharmacological effects of ethanol.

1.2 Driving simulation in DUI research

Driving simulators have been in use for many decades. The earliest account of an apparatus that could simulate driving dates back to 1934 (Caird and Horrey 2011), and consisted of a miniature road model transferred by a projector to a screen in front of the driver (Miles and Vincent 1934). Researchers originally developed driving simulators as a response to the growing need for research and training to combat the problem of traffic injury and death, and probably drew on the experience with flight simulators, which became available in the 1920s and were brought about by the necessity to train pilots (Caird and Horrey 2011). However, driving simulation did not become an important field of research until analog computers, electronic circuits and display technologies became more advanced and available during the 1960s. Development was motivated by the possibility to avoid the cost of field studies, as well as to achieve control over conditions and measurements, and to be able to safely present hazardous driving conditions (Allen *et al.* 2011). The earlier driving simulators were very simple, did not convey much of a feeling of realism, and could record only a very limited range of measurements. As the technological development progressed, driving simulators became increasingly advanced and realistic, and their availability improved as the technology became relatively cheaper. The first driving simulator at NTNU/SINTEF was a video-based simulator acquired in 1988. In 1999, a graphic simulator was installed (Engen 2008).

1.2.1 Types and applications

Driving simulators are often divided into three hierarchical classes according to their technological finesse and realism (Engen 2008, Kaptein *et al.* 1996):

- Low-level simulators

  These are the least advanced simulators, which are typically PC-based, and consist of a screen displaying a driving scenario and simple operating controls such as a steering wheel, brake and throttle. An advanced gaming platform equipped with steering wheel and pedal controls would be considered a low-level driving simulator. The STISIM simulator (figure 1–4) (www.stisimdrive.com) is an example of a commercially available and much used low-level driving simulator, including for DUI research (Mets *et al.* 2011).
- Mid-level simulators

  This group comprises a wide range of simulators that have in common an enhanced realism compared to the low-level simulators. They typically have the appearance of a regular car, which is placed in front of graphic screens displaying the driving scenario. The simulators can vary widely with respect to the field of view, graphic imaging techniques and features to increase realism such as vibration and sound systems. The most advanced mid-level simulators may also include a simple motion base. The NTNU/SINTEF graphic simulator may be regarded as an advanced mid-level simulator, and is described in more detail in chapter 3.1. The MUARC driving simulator (figure 1–5) located at the Monash University in Melbourne, Australia (www.monash.edu) is another example of an advanced, mid-level simulator.
- High-level simulators

  The most advanced simulators have moving platforms that can realistically reproduce motion, including gravitational forces. Examples of high-level simulators is the National Advanced Driving Simulator in Iowa, US (figure 1–6) (www.nhtsa.gov), and the VTI driving simulator at the Swedish National Road and Transport Research Institute (www.vti.se).

To put it simply, the difference between a low-level and a mid-level simulator is the appearance of the latter as a normal vehicle with original controls, whereas the full-range motion platform separates the high-level from the mid-level simulators (Engen 2008).

Figure 1–4. STISIM Drive® hardware (www.stisimdrive.com)

Figure 1–5. The Monash University driving simulator (www.monash.edu)

Figure 1–6. The NADS driving simulator (www.nhtsa.gov)

Driving simulation has many uses, of which many are unrelated to traffic medicine (e.g. road planning or vehicle design, training purposes, licensing) (Caird and Horrey 2011). Within traffic medicine, driving simulators may be used both clinically for diagnostic purposes and occupational therapy in individual patients, i.e. assessing fitness to drive (Classen and Brooks 2014), and in research (e.g. to investigate driving impairment with different diseases or disabilities, or driving impairment due to drugs and ethanol).

For the purpose of studying drug-impaired driving, simulators have many advantages over real (on-road) driving (Caird and Horrey 2011):

- They allow easy, cost-effective and simultaneous measurement of a vast array of outcomes related to driving ability in various driving environments
- They provide the opportunity to study high-risk driving scenarios (e.g., dangerous traffic situations with imminent crash risk) and drug impaired driving safely
- They allow repeatability and easy manipulation of driving conditions, e.g. daytime or nighttime driving, weather conditions, traffic conditions etc.

However, driving simulation also has drawbacks compared to real driving, of which the most cited is probably whether simulator data are representative of those measured when driving in the real world (Caird and Horrey 2011). Simulators often lack proper validation for the assessment of driving performance in general and for driving impairment due to drugs in particular. This issue is dealt with in detail in chapter 1.2.3. There are many types of driving simulators in use, equipped with driving scenarios that differ widely with respect to test duration, setting (e.g., urban or rural), measurements that are recorded, and degree of realism. The lack of standardization of simulator testing of drug impaired driving makes it difficult to compare results across studies and to draw firm conclusions. It has also been pointed out that driving simulators are good at assessing driving performance, i.e. what the driver *can* do, but they do not necessarily reflect driver behavior, i.e. what the driver *will* do (Caird and Horrey 2011).

1.2.2 Measurements and tasks

One of the advantages with driving simulators is the relative ease of obtaining a wide range of measurements in a variety of different driving tasks, which together can describe most aspects of the complex behavior that constitutes driving. At the same time, the seemingly infinite possible combinations of measurements and driving tasks may be problematic, since this hinders standardization and comparability across studies and between different driving simulators. It may also lead researchers to measure 'anything measurable', without a proper rationale that what is measured actually is relevant to safe driving. In other words, a driving task with a certain outcome should be properly validated before being used to characterize drug impaired driving.

Measurements in driving simulator studies commonly include the following:

- Longitudinal control: Speed measures
  Speed is obviously related to traffic accident risk and severity. In simulator studies, average speed can be measured throughout the driving task, as well as during subsections (e.g., curves and straight sections) or specific maneuvers. Speed variability, expressed as the standard deviation of speed (SDS), is a measure of the driver's ability to maintain a constant speed. It is evident that the interpretation of both average speed and SDS is highly task-dependent. For instance, in the Dutch on-road driving task, the subjects are

instructed to keep a stable speed at 95 km/h, and the driving conditions do not require much speed adjustments. Under these circumstances, the measurement of average speed is mainly a control of the test subject's compliance with the driving task, whereas the SDS is seen as a performance measure that is not necessarily linked to accident risk (Verster and Roth 2014). In other test scenarios, with more challenging driving conditions and no explicit speed instructions, an increase in average speed may be seen as a marker of risky driving behavior and impaired self-assessment (Fillmore *et al.* 2008, Zhang *et al.* 2014), whereas SDS may be difficult to interpret since for instance a winding driving scenario in itself may dictate variations in speed (Helland *et al.* 2016).

- Headway: Maintaining distance

  The ability to maintain a steady and safe distance to the car in front is important to traffic safety, and may also be seen as a measure of longitudinal control. Some simulator studies employ a car-following task, in which the headway distance is used as an outcome (e.g. Hartman *et al.* 2016), or the percent of time tailgating the car in front, i.e. keeping a time headway of less than 1 second (Kenntner-Mabiala *et al.* 2015).

- Lateral control: SDLP, departures out of lane, etc.

  The lateral control of a vehicle, i.e. the ability to keep it stable within the confines of the road, has obvious relevance to traffic safety. If the deviation from the optimal trajectory becomes pronounced, the risk of driving off the road or into adjacent or oncoming traffic increases (Verster and Roth 2011). As described earlier, the standard deviation of lateral position (SDLP) is a thoroughly validated and much used outcome in on-road studies of drug-impaired driving, and is also increasingly being used in simulator studies (e.g. Hartman *et al.* 2015, Verster *et al.* 2015). Number or frequency of lane departures is an alternative measure of lateral control that could arguably have more direct relevance to traffic risk, since increased wavering within the assigned lane does not necessarily increase accident risk. However, measurement of excursions out of lane has been shown to be much less sensitive to drug impairment than SDLP (Verster and Roth 2014).

- Steering wheel and pedal use

  The frequency and intensity with which the test subjects manipulate the steering wheel and pedal controls are readily available measurements in driving simulators. They could

be regarded as possible proxies of lateral control (steering) and longitudinal control (braking/accelerating), respectively.

- Eye movements and physiological parameters

  Some simulators are equipped to register eye movements of the test subjects, which can be used to measure the range of gaze or if the subjects actually register relevant information, such as road signs. Physiological parameters such as skin conductance or heart rate may also be registered as measures of stress or mental workload.

- Reactions to sudden incidents, potential crash events, complex traffic

  The most obvious advantage of driving simulators over real-world driving studies is perhaps the opportunity to experimentally reproduce high-risk scenarios, which would be unethical, if not outright dangerous, in on-road studies. For example, reaction time to sudden incidents, such as the time from another car undertaking a dangerous maneuver to the activation of the brake pedal, could be measured. Number of collisions, as an outcome with obvious relevance to traffic safety, could also be used as an outcome measure in a challenging driving scenario. However, many such outcomes have been shown to be relatively insensitive to drug impairment (Berthelon and Gineyt 2014, Veldstra *et al.* 2012). Kenntner-Mabiala *et al*. (2015) suggested the combination of several parameters of unsafe driving, as judged by a professional evaluator, into a global variable as a strategy to improve sensitivity.

- Dual attention tasks

  Data from experimental studies show that the ability to divide attention between two tasks is very sensitive to ethanol effects. Obviously, this skill is relevant to safe driving, and can be exemplified by for instance the distraction of responding to a phone call or soothing a crying baby while driving. There are many ways to test the ability of divided attention. For example, one study employed a dual task requiring the subjects to drive the car in a simulated highway scenario while at the same time identifying numbers that appeared on the screen as either even or odd. Outcomes of both the primary task (driving; SDLP) and the secondary task (number identification; reaction time) were found to be impaired (Freydier *et al.* 2014). One weakness of such tasks is that they are artificial, and the transferability to actual driving may be questionable.

Although the possibilities of creating scenarios and tasks seem endless, those suitable for DUI testing in driving simulators may be categorized in three main principal driving tasks (Kenntner-Mabiala *et al.* 2015, Owens and Ramaekers 2009, Veldstra *et al.* 2012):

- Highway/rural road driving test (O'Hanlon 1984)
  In this task, subjects drive for an extended period in a highway scenario or, alternatively, in a rural road scenario like the one used in our validation study. They may be instructed to keep constant speed and lateral lane position, as in the Dutch on-road test, or they may receive more 'naturalistic' instructions to keep in lane and within speed limits or safe speeds as dictated by the road curvature and traffic conditions, and otherwise drive according to their habits. This task is best suited to investigate effects on automated driving behavior and the primary outcome will often be a measure of lateral control such as SDLP, but a range of other measures at various levels of driving behavior (e.g. speed and speed variability, steering wheel and pedal use, reactions to sudden incidents) may also be measured.
- Car following task (Brookhuis *et al.* 1994)
  In this task, participants are instructed to maintain a constant safety distance to a leading vehicle as it performs several accelerations and decelerations. The primary outcomes may be reaction time to the leading vehicle's change of speed, or headway distance/time (from the rear of the leading vehicle to the front of the test vehicle). The amount of time spent travelling at an unsafe headway distance may also be reported. This task measures the driver's ability to adjust behavior to other traffic, and is thus a measure of the operational level of driving behavior.
- City driving test (Veldstra 2014)
  In this task, subjects drive in an urban environment where they must undertake difficult maneuvers and interact with complex traffic. This is perhaps a type of driving task particularly suited to driving simulators, since complex and potentially hazardous traffic scenarios may be safely enacted and exactly reproduced. The city driving test allows testing outcomes on the tactical (strategic) level of driving behavior, such as risk-taking and reactions to sudden events.

1.2.3 Validation

According to the Oxford online dictionary (2016), validation is "the action of checking or proving the validity or accuracy of something", whereas validity is defined as "the quality of being logically or factually sound". Validation of a scientific method, such as driving simulation to test drug impairment, is thus to make sure that the experimental results are factually sound.

There are several different concepts of validity that refer to different aspects of the experimental method (Engen 2008, Kaptein *et al.* 1996):

- *Face validity* refers to the test subjects' intuitive perception that the test actually measures what it purports to measure. For instance, an advanced driving simulator, with a high-resolution, realistic driving scenario that provides a naturalistic impression of driving, conveys a high face validity for measuring driving ability, whereas a laboratory test of psychomotor performance (e.g. a divided attention task involving tracking and reaction time) has a low face validity in this context. The results obtained in an experiment with low face validity are not necessarily invalid, but the low face validity may negatively affect the test subjects' motivation.
- *Construct validity* describes the applicability of the test for measuring the theoretical construct it has been designed to measure (i.e., in a study of drug effects on driving, does the test actually measure 'driving ability' and not just some other drug effect that may not be relevant to safe driving?).
- *Statistical conclusion validity* includes the use of appropriate sampling procedures, adequate statistical tests, and reliable measurement procedures. For example, inadequate sampling may lead to statistically underpowered studies and, consequently, a risk of type II error (concluding with no difference when a difference actually exists), whereas performing multiple statistical tests ('data fishing') increases the risk of type I error (concluding with a non-existing difference due to chance alone).
- *Internal validity* refers to the ability of the test to reproduce predictable results (i.e. effects on the dependent variable) that reflect changes in the variable that is being studied (i.e. the independent variable), when conditions are otherwise unchanged. The prerequisites for internal validity are: 1) a temporal relationship (cause precedes effect);

2) co-variability (cause and effect are related); and 3) probable causation (no other plausible explanations for the observed covariation exist). In experimental research, the 'gold standard' to achieve high internal validity is the randomized controlled trial.

- *External validity* relates to the generalizability of the results of a test. There most important factors that influence external validity are: 1) the representativeness of the sample (is it possible to generalize from results obtained with 'young, healthy, male firefighters'?) and 2) the representativeness of the setting (e.g., in driving experiments: is daytime simulated highway driving generalizable to all driving situations in real life?). In other words, external validity concerns the generalizability across populations and situations. Researchers generally take steps to increase external validity by choosing a representative sample of the population that is under study (i.e., the general driving population in driving studies), and an experimental setting that closely resembles or is considered representative of the real-life situation.

The validity of driving simulation studies is most often discussed in terms of the realism and generalizability of the simulation, i.e. external validity with extra emphasis on *ecological validity*, which concerns the resemblance of the test conditions to 'real life' conditions. External and ecological validity are often confused with one another, but are not identical. Ecological validity may rather be seen as a particular aspect of the broader term external validity. For example, the Dutch on-road test is said to have high ecological validity, since the subjects drive a regular car on a public road among other traffic, performing a naturalistic driving task (Verster and Roth 2011). However, for the test to be externally valid, this also requires that the study subjects are representative of the population in question, and the performed task is representative of the driving situation under study.

The 'gold standard' to assess overall validity of simulator research is to compare the simulated driving task to a similar real driving task (Blaauw 1982, Shechtman *et al.* 2009). This requires both physical validation (also called vehicle response validation or intrinsic validation, or referred to as the simulator's fidelity) and behavioral validation (also called driver response validation, predictive validation or sometimes, confusingly, external validation). The physical validation part investigates how well the simulator dynamics and visual reproductions replicate the vehicle and driving scenario that is being simulated. The behavioral validation

assesses the rapport between driver behavior in simulated and 'real-world' driving. Not subjecting the simulator to such validation may leave doubt as to whether the subjects' performance in the simulated scenario is transferable to performance in real driving situations. Behavioral validity is specific for the particular type of simulator, scenario, test and population used in the validation assessment, and will not necessarily be transferable to other driving simulators, scenarios, tests, or populations. However, it has been argued that the accumulated evidence of validity from a range of different driving simulators, scenarios and tasks adds weight to the validity of driving simulators in general (Godley *et al.* 2002).

Behavioral validity is absolute if the simulator and the real driving test invoke the same effect to the same extent (e.g., no statistical difference in SDLP values between the simulated and the real driving scenarios). Relative behavioral validity implies a trend of change in the same direction both in the simulator and in the real driving environment, but of different magnitudes (Kaptein *et al.* 1996, Shechtman *et al.* 2009). In previous research validating mid-level driving simulators in a range of different driving scenarios and tasks, relative validity has often been established, whereas evidence of absolute validity is rare (Engen 2008, Godley *et al.* 2002, Kaptein *et al.* 1996, Shechtman *et al.* 2009). However, for a simulator to be a useful research tool, absolute validity is not essential, but relative validity is necessary (Törnros 1998). This is because the research question usually pertains the difference between control and intervention conditions, and not absolute numerical measurements (Godley *et al.* 2002).

Since validity is specific, it is important to validate the simulator for its intended use (Kaptein *et al.* 1996). In research of driving impairment due to drugs, including ethanol, this necessitates the assessment of drugged driving in the simulator as compared to drugged behavior during real driving (Owens and Ramaekers 2009). Without such validation, results from experimental research on drugged driving in the simulator are difficult to interpret. If, for instance, a drug with suspected impairing effects on driving is tested in a simulator study, and no impairing effect is found, the lack of validation will render such a result uninterpretable. The results may indicate a lack of impairing effects of the drug, but they may just as well be due to inadequate sensitivity of the simulator test. As a case in point for the necessity of task-specific validation, the lane-change task, originally developed and validated as a reliable detector of driver distraction, was shown not to be sensitive to ethanol

impairment (Huemer and Vollrath 2010). Thus, negative results in such a task would not exclude the possibility of relevant impairment of driving skills. Positive findings may be equally difficult to interpret, as a certain impairing effect measured in the simulator is impossible to relate to real world effects unless validation has been undertaken. Hence, a validation study should be performed before the simulator is used to study drugs with unknown effects on driving, using a *verum* or 'positive control' drug, i.e. a drug that is known to cause driving impairment in the real world. Ethanol is often used for such purposes because of its well-characterized effects on driving. In addition, it is a legal drug to which a majority of the population expose themselves regularly, which reduces ethical concerns. Another important reason for choosing ethanol as a probe drug for validation of drug impaired driving is the known quantification of traffic risk as a function of BAC. This 'benchmark drug' quality of ethanol in some cases may even enable the use of BAC as a 'translation factor' between a certain measure of driving impairment and traffic risk (Jongen *et al.* 2014).

Surprisingly few validation studies using this approach have been published, with the important disclaimer that there may exist a large volume of 'gray' simulator validation literature in the form of unpublished internal reports or conference proceedings etc. (Caird and Horrey 2011) that is not available for review. However, in a study published as far back as 1988, the effects of different BACs on driving performance were compared in a driving simulator and on a closed course (Gawron and Ranney 1988). The performance of twelve subjects in a driving simulator was compared to the performance of six other subjects driving a real car on a closed test course at three different BACs (0.00 %, 0.07 % and 0.12 %). The researchers reported generally increased standard deviations of measures of longitudinal and lateral vehicle control as BAC increased, notably SDLP on the closed course. However, the study had several weaknesses: The sample size was small, different subjects partook at the two test arenas, and the measurements at the closed course and in the simulator did not overlap (e.g., SDLP was not measured in the simulator). Obviously, the simulator used was also of a much more primitive type than the advanced simulators of today. More recently, several research groups have performed calibration studies with ethanol to validate driving simulators (Kenntner-Mabiala *et al.* 2015, Mets *et al.* 2011, Veldstra *et al.* 2012). These studies will be discussed further in chapter 5.

Other psychoactive substances with known effects on measures of driving may be used as benchmark drugs in validation. One recent study compared dose-related cannabis effects on simulated and on-road driving. Twenty-four healthy subjects undertook a highway driving and a car following task, both in a driving simulator and on-road, after administration of either placebo, 10 mg or 20 mg of the oral medicinal cannabinoid preparation dronabinol. SDLP showed similar significant, dose-related effects between on-road and simulated driving, although variability in driving performance in the simulator was higher. Results from the car following were however not comparable between on-road and simulated driving (Veldstra *et al.* 2015). Another recent study tested the effects on driving the morning after evening intake of zopiclone 7.5 mg and placebo in a low-level simulator (MiniSim), and found significant increases in SDLP in the zopiclone condition, of similar magnitude to that found in previously published on-road studies (Simen *et al.* 2015). This proves the sensitivity to zopiclone for the simulated driving task, but external validity cannot be claimed from such a study, since the correlation to real-life driving has not been proven by direct comparison.

The first two papers in this thesis present the results from the validation of a driving simulator scenario that was intended for future studies to assess drugs with possible impairing effects on driving. The validation study compared simulated driving with real driving on a test track across three BAC levels. Paper I concerns the main outcome variable, SDLP, whereas paper II describes results from secondary outcomes.

1.2.4 Simulator sickness

The phenomenon of simulator sickness is a major obstacle to the use of driving simulators. The syndrome resembles motion sickness, and comprises several symptoms: headache, cold sweats, mouth dryness, drowsiness, disorientation, vertigo, dizziness, nausea and vomiting. Different theories of causation exist (Brooks *et al.* 2010). The 'sensory conflict theory' is the most widely accepted explanation, stating that simulator sickness stems from an incompatibility of sensory input, with conflicting signals from the visual system, perceiving motion, and the vestibular system, perceiving little or no motion. According to the alternative 'postural instability theory', the driving simulator is a novel environment to which the test subject must adapt in order to maintain postural stability. This is analogous to 'getting one's

sea legs' in a boat, and may also explain the phenomenon that subjects experiencing simulator sickness often feel sick long after they have left the driving simulator.

The degree and prevalence of simulator sickness varies between studies, but is often sufficient to cause significant dropout, which reduces statistical power and may introduce bias in the study population and confound results (Brooks *et al.* 2010, Classen *et al.* 2011). Although research in the field has identified some factors that predict simulator sickness (Classen *et al.* 2011, Kawano *et al.* 2012, Milleville-Pennel and Charron 2015), as well as techniques to alleviate it (Domeyer *et al.* 2013, Galvez-Garcia *et al.* 2015), the occurrence of simulator sickness is omnipresent. The possibility that it may influence outcomes in experimental studies of driving, such as SDLP, has received little attention, although effects on psychomotor control have previously been described (Cobb *et al.* 1999). Hence, there is a risk that simulator sickness may act as a confounder, and decrease the validity of the results. Little is also known about the interplay between study design (e.g., repeated measurements), interventions (e.g. ethanol) and the degree of simulator sickness. Given these shortcomings in the existing literature, which are further elaborated in paper III, it is important to investigate the factors that may influence the degree of simulator sickness, as well as its possible impact on the main study outcomes.

2 Aims

2.1 The VALIDAD research project

VALIDAD (Validation of a driving simulator to assess drug effects) has been a joint research project between the SINTEF research organization, the Norwegian University of Science and Technology (NTNU) and St. Olav University Hospital, all based in Trondheim, Norway. Its purpose was to validate the advanced mid-level NTNU/SINTEF driving simulator for use in studies of drug impaired driving, and to investigate the impairing effects of certain psychoactive substances (benzodiazepines, Z drugs, codeine) in common use. Validation of the driving simulator test scenario by comparing simulator driving with real driving on a test track under the influence of ethanol in a within-subject design represented a new approach to experimental DUI research. It could have the potential to eliminate validity concerns of simulated driving while at the same time avoiding economical, ethical and safety concerns associated with on-road studies.

The project received funding from the Research Council of Norway. A large number of personnel were involved in the studies, including technicians, driving instructors, biochemical analysts, researchers with various backgrounds (medical doctors, behavioral psychologists, engineers), and even a professional bartender to help mix drinks that most effectively could conceal the ethanol content to aid blinding. It was also necessary to obtain permissions from several authorities (ethical committee for human research, road and traffic authorities, and police) in order to perform the research.

Two pilot studies and one full validation study have been completed. They were all performed with healthy volunteers, with ethanol as test substance, both in simulated and real (closed test course) driving. Validation was performed to ensure adequate sensitivity of the simulator test scenario to relevant driving impairment, as well as to confirm the external validity of results from simulator testing. The pilot studies included two and eight test subjects, respectively, and provided useful experience in conducting studies of drug impairment in the simulator, as well as crucial information concerning the suitability of test scenarios to detect drug impairment and the sensitivity of different outcome variables. Results from the second pilot study with eight test subject have been described in an internal report at the SINTEF research organization (Sakshaug 2008).

The original plan was to move on to studies on drugs after the second pilot study, which was supposed to provide validation of a driving scenario with sufficient sensitivity to investigate drug effects on driving performance. However, the scenario and outcomes investigated in the pilots lacked adequate sensitivity to ethanol effects, and barely showed significant results. The experience from the pilot studies prompted us to develop a more monotonous, nighttime driving scenario of longer duration, and to test it in another validation study with more participants, as described in chapter 3.2.

So far, the VALIDAD project has provided us with a validated test scenario that is suitable for the investigation of drug-impaired driving. Unfortunately, due to the unexpectedly demanding validation process and a lack of funding, as well as organizational issues, the second part of the project (i.e. planned testing of medicinal drugs) has not been completed.

2.2 Aims of the thesis

This thesis is based on the main validation study within the VALIDAD project described in chapter 2.1. It comprises the validation of a driving simulator test scenario for use in studies of drug impaired driving (paper I) and the comparison and evaluation of various driving-related outcomes in the simulator (paper II), as well as aspects of simulator sickness (paper III) and ethanol blinding (paper IV) that are of concern in such studies.

The aims were as follows:

- to develop a driving simulator test scenario with sufficient sensitivity to measure relevant drug-impairing effects, defined as a statistically significant, dose-dependent increase in SDLP over a BAC range of 0–0.9 g/l (paper I)
- to validate the test scenario against real driving at a test track to ensure external validity of the results (papers I and II)
- to explore the dose-dependent effects of ethanol on measurements that represent different behavioral levels relevant to driving, and compare their effect sizes in simulated and real driving as well as against that of SDLP (paper II)
- to explore the possible influence of simulator sickness on several measures of impaired driving, including SDLP (paper III)
- to investigate the effect of BAC and repeated exposures to the simulator on the reported degree of simulator sickness (paper III)

- to describe the efficacy of a novel approach to blinding of the ethanol intervention in placebo-controlled studies by use of a sham sedative pill (paper IV)
- to explore the possible effects of subjective drug expectancy on SDLP (paper IV)

3 Material and methods

The methodology used has been described in each paper, with emphasis on those aspects most relevant to each paper's research questions. In this chapter, the material and methods of the validation study are described in more detail. A certain degree of overlap with the papers was difficult to avoid.

3.1 The NTNU/SINTEF driving simulator

Description

The NTNU/SINTEF graphic driving simulator was acquired from Autosim A/S in 1998 and has been in operation since 1999. The Norwegian university of science and technology (NTNU) and the SINTEF research organization jointly own the simulator, which is located at a SINTEF research facility in Trondheim, Norway. Since its inauguration, the simulator has been in frequent use, both for training and research purposes, and has been subject to constant development and improvement. It has been used for research within the fields of driving performance, driver assistance systems, traffic regulation and road design, information technology, training methods, and traffic medicine. The simulator may be equipped with either an ordinary car or a lorry cabin. It originally had no motion base, only a vibration system, but was later upgraded with a simple three-axis motion system, after which it could be regarded as an advanced mid-level driving simulator (Engen 2008).

In the validation experiment, the driving environment consisted of a Renault Scenic 1997 model with automatic transmission and original controls. The simulator was equipped with a three-axis moving platform, a vibration mechanism in the chassis, as well as a motored steering wheel to provide force feedback. A four-channel hi-fi sound system provided the driver with a realistic sound experience. The driving scenario was depicted by a total of five projectors onto three screens in front of the car and two screens behind the car, each screen measuring 3.1 m in width by 2.4 m in height. In total, the screens provided a 180° horizontal field of view and a 47° vertical field of view to the front, and a 90° horizontal field of view and a 47° vertical field of view to the back. The resolution was 1400 x 1050 pixels for the center front projector, and 1024 x 768 pixels for the other projectors. Inside and outside mirrors synchronously reflected the driving scenario. The visualization program was run on PCs with a Windows operating system. Immediately adjacent to the room housing the simulator was an

operator area from which the simulator and test subjects could be observed directly, as well as via cameras placed in the car cabin. There was also a spacious and comfortably furnished waiting area for test subjects, including toilet facilities.

Figure 3–1. Appearance of the driving simulator

The simulator software allowed the manipulation of several variables, enabling the construction of different driving scenarios, and consisted mainly of three modules:

- A terrain database contained information about the road network, traffic rules, road surface and visual context (environment) used.
- A model database contained information about vehicle characteristics and appearance, as well as physical properties of other objects, such as other vehicles, traffic signs, traffic cones etc.
- A scenario database described the actions of interacting vehicles, pedestrians etc.

Additional modules allowed the manipulation of light conditions (from daylight to darkness), time of year, and sight conditions (clear visibility, fog, rain).

Numerous measurements may be recorded while using the simulator. Among the most important are position within the driving scenario, speed, lateral position within lane, distance

to other vehicles, brake and speed pedal pressures, steering wheel movement and position. The simulator also has equipment for video recording of test subjects while driving and eye-tracking equipment. The measurements used in the validation study are described further in chapter 3.2.5.

Previous validation

The NTNU/SINTEF simulator has been subject to validation studies previously. Moe (1995) examined test participant's view of the realism of the ordinary car version of the driving simulator. Seventeen professional and 20 non-professional drivers took part in the validation, which was conducted by use of a questionnaire encompassing the physical, operational and psychological experience when driving the simulator. The non-professional drivers rated the physical realism (i.e. appearance) of the simulator as high, whereas operational (i.e. operation and handling of the car) and psychological (i.e. perceived) realism were rated as medium, with an average score of four on a scale from one (not realistic) to seven (very realistic).

Engen (2008) studied the validity of using the driving simulator as a research tool for reaction time. Thirty-one test subjects with previous experience in the driving simulator were told to drive as they would normally do. Measurements of reaction time to various incidents in simulated driving were found to be comparable to measurements of reaction times to changing traffic lights in intersections in real driving, as well as reaction times described in the literature. In conclusion, the simulator was found to have external validity in this respect.

Engen (2008) also studied the validity of speed and lateral placement measurements in the simulator, related to effects of road width and design of midline markings. Twenty-nine test subjects took part in the evaluation in the driving simulator, where they drove three different road stretches with different road width and midline markings. Their results were compared to real-life measurements of speed and lateral placement of ordinary vehicles circulating on existing road stretches with similar characteristics. The results showed that mean speed and mean lateral position within lane were of the same magnitude in the simulated drive as in real driving, and co-varied in a similar way according to the manipulations of road characteristics. However, real world observations showed a larger variability. This difference was explained by the increased stochastic variability of real world driving, which is necessarily more influenced

by confounding variables than the more simplified and controlled laboratory environment of the driving simulator.

In a car-following validation study, 14 participants each undertook four different driving simulator scenarios in which they were forced to drive behind a slower moving vehicle (no overtaking opportunities) over a long stretch. The time gap (time between the passing of the rear of the slower vehicle to the passing of the front of the test vehicle) was measured, and compared to car-following situations in an on-road study with four participants who drove an instrumented vehicle. The results showed that the time gap between vehicles in simulated driving was less than half that of on-road driving. Test subjects thus seem to accept a much shorter safety distance in the simulator than in on-road driving. The difference was explained by the lack of real danger in the simulator, as well as a much more predictable and overly simplistic driving environment (e.g. no other traffic, constant speed of the vehicle in front) (Engen 2008).

3.2 The validation experiment

3.2.1 Trial design

In light of the main aims of the VALIDAD research project (chapter 2), the validation study was designed to:

- reflect a naturalistic driving situation in which a disproportionately large share of DUI-related accidents occur
- as far as possible employ identical driving scenarios for simulated and 'real' driving
- characterize and validate measurements of drug impaired driving by use of ethanol as a validation and calibration substance
- as far as possible eliminate learning effects, expectancy effects and the impact of confounding factors by use of familiarization, randomization and blinding procedures

In accordance with this, the experiment was designed as a randomized, placebo-controlled, single blind, crossover study. Each test subject undertook a total of six driving tests – three in the simulator and three on a closed test track with a length of approx. 1.4 km (figure 3–2).

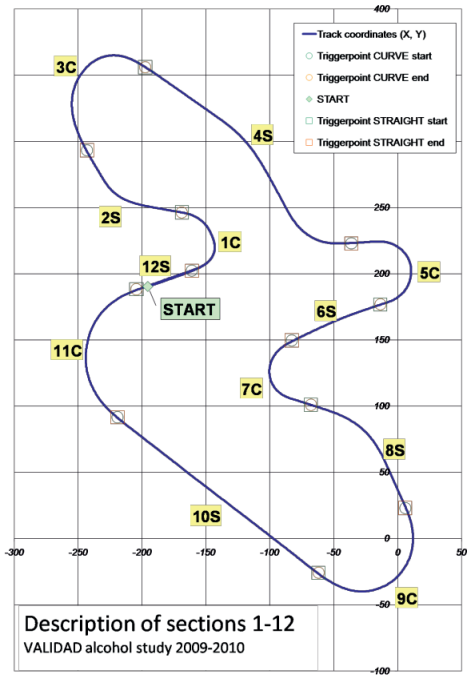

Figure 3–2. Schematic layout of the driving scenario

Three interventions (placebo, low BAC ≈ 0.5 g/l and high BAC ≈ 0.9 g/l) were crossed with two 1-hour driving tasks: driving an instrumented vehicle on a closed test track, and driving in the simulator with a driving scenario that was modeled to mimic the test track as closely as possible. The test track resembled a narrow Norwegian rural road, around 5.5 m wide with authentic midline and side markings. The order of the interventions was randomized by use of a counterbalanced multiple-condition design. Ethanol doses necessary to obtain the desired BACs were calculated based on the participant's body weight and the Widmark equation (Jones 2011). As an additional confounder to enhance blinding, the study subjects were administered a placebo pill, which they were told may or may not contain a sedative drug, with the drink. Only the necessary personnel knew the details about which substances were administered and at which dosage. Venous blood samples for BAC measurement were drawn immediately before and after each driving session. See figure 1 in paper I for an outline of the trial design.

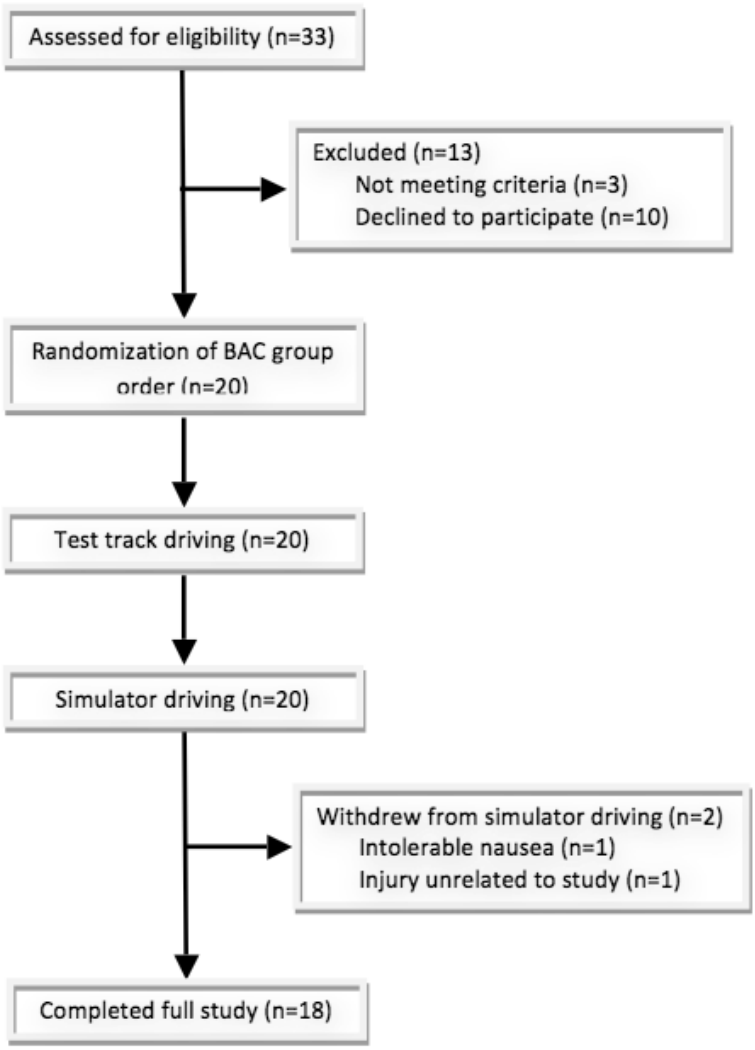

Figure 3–3. Study flow chart.

3.2.2 Test subjects

Participants were recruited through medical students' organizations, student and employee networks at the Norwegian University of Science and Technology, and the employee website of the SINTEF research organization. We chose to recruit a rather narrow demographic group to minimize variability in driving experience, ethanol tolerance and metabolism. The teratogenic risk of ethanol was an additional reason to exclude women. Thus, twenty healthy, Caucasian, male volunteers aged 25–35 years (mean 28.7 years) who had been in possession of a driver's license for at least 5 years (mean 10.6 years), were included in the study. They

were all recreational users of alcohol, and as a group drove slightly more and had a higher educational level than the general population (see tables 3–1 and 3–2 for details).

Table 3–1. Educational level of included test subjects compared to the general Norwegian driving population

Highest education accomplished	Included subjects	General adult population[1]
Primary education (compulsory)	5 %	29 %
Secondary education (high school, incl. vocational ed.)	25 %	45 %
Higher education (university/college), bachelor degree	30 %	17 %
Higher education (university/college), master degree	40 %	8 %

[1] Population's level of education 2009. Statistics Norway (www.ssb.no)

Table 3–2. Automobile driving distance per year among included test subjects compared to the general Norwegian driving population

Driving distance last year	Included subjects	General driving population[1]
0–5 000 km	10 %	13 %
5 000–10 000 km	15 %	22 %
10 000–15 000 km	25 %	28 %
15 000–20 000 km	25 %	20 %
20 000 km and above	25 %	18 %

[1] Survey accomplished in connection with "Evaluating penalty point endorsements of driving licences" (Stene TS, Moe D and Sakshaug K, SINTEF Report A4448, 2007)

Other exclusion criteria were previous or present drug or alcohol abuse, history of DUI or aggressive reactions to alcohol, intolerance to blood sampling, daily intake of any prescribed drug, or high likelihood of simulator sickness as assessed with a modified version of the Apfel risk score for postoperative vomiting (Apfel *et al.* 1998). Written information about simulator sickness was given prior to inclusion, and repeated orally at each driving session. Before final inclusion, prospective participants underwent a screening trial of 20 minutes' duration in the simulator to exclude persons with excessive simulator sickness and familiarize them with the simulator to reduce learning effects. All participants received a gift certificate worth NOK 1000 (approx. USD 150) upon completion of the study.

Ethical considerations

Each participant underwent a screening for eligibility, received written and oral information about the study and provided a written consent to participate. Ethical concerns with the study lay mainly in the dangers of real driving under the influence of ethanol, as well as the administration of an intoxicating substance. We took every step to exclude drivers with concomitant drug use, previous drug addiction, aggressive reactions to ethanol or a history of

DUI. All test subjects were recreational drinkers with a known reaction pattern to ethanol. The highest dose of ethanol administered, designated to result in a BAC of approx. 0.9 g/l, is comparable to doses that are regularly consumed by a large part of the population, and may thus not be considered dangerous *per se*. During test track driving, the instrumented vehicle had a double set of controls, and experienced driving instructors were present in the passenger seat at all times to intervene if necessary. The participants were informed that they should consider themselves to be under the influence and not attempt to drive a vehicle, operate machinery or engage in any kind of risky activities until the day after the test (including placebo sessions). All subjects were under continuous observation at the test sites. A physician was present on the site at all times during test drives. After each driving test, study personnel drove them directly home after they were deemed fit.

It was necessary to store personal data (including biological material) for a period of time in connection with the study. The biological samples were marked with a participant identification number, linked to participant's names via a confidential list that was available only to the necessary study personnel. Biological samples were destroyed after analyses had been performed and their validity ensured. All results from the study are presented without identification of the participants. The experiment was executed in accordance with relevant national and international regulations and standards for good clinical practice (GCP). The study was approved by the Regional Ethics Committee, and was registered as a clinical trial in the ClinicalTrials.gov database (identifier: NCT00967421). Permission to carry out the test track driving was granted from the local police.

3.2.3 Test track driving

The test track part of the experiment was completed during a frost-free period of six weeks in the autumn. All drives commenced after nightfall, between 20:00 and 01:00 hrs. The test track circuit was 1.37 km long, laid out in hilly terrain, with both gentle and sharper curves. The track was hard-surfaced (grey asphalt), with lanes measuring 2.75 m in width in each direction, and had midline and side markings similar to standard Norwegian road marking. Surprise obstacles ($1m^3$ foam rubber cubes) were placed in two locations on two occasions, one at the beginning and one towards the end of each driving trip, and were to be avoided by the test subjects. Traffic lights present in two locations each turned red on one occasion

during each trip. The participants were instructed to keep in the middle of the lane, adjust speed according to the driving conditions and otherwise drive as they would normally do.

The participants drove an instrumented vehicle (Volvo V70 2.4s) with automatic transmission (figure 3–4). The measurements available with the instrumented vehicle had been selected with the possibility in mind to validate the driving simulator through parallel studies of driving behavior in 'real world' and in the simulator (Engen 2008).

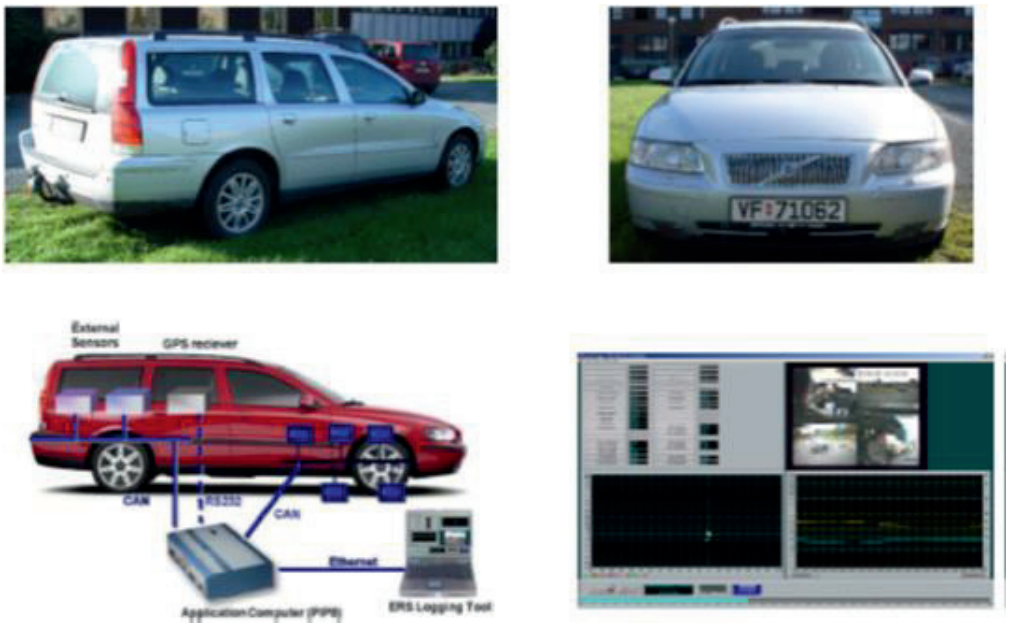

Figure 3–4. Instrumented vehicle in use at the closed test course. From Engen (2008).

To enable continuous recording of lateral position in the road lane, the test car was equipped with an infrared wide-angle camera fixed to the roof of the car, pointing at a downward angle to the rear of the car (figure 3–5). The data were stored in a database and analyzed in a program for photo analysis (Open Source Computer Vision Library). A filtering algorithm (Hough transformation) that is able to identify areas with contrasting colors was used to identify roadside markings. The standard deviation of lateral position was calculated from the measured distance from the left side of the vehicle frame to the closest line marking (usually the road midline). The car also featured other equipment for recording the location of the car on the test circuit (global positioning system; GPS), speed, pedal use and steering wheel movements.

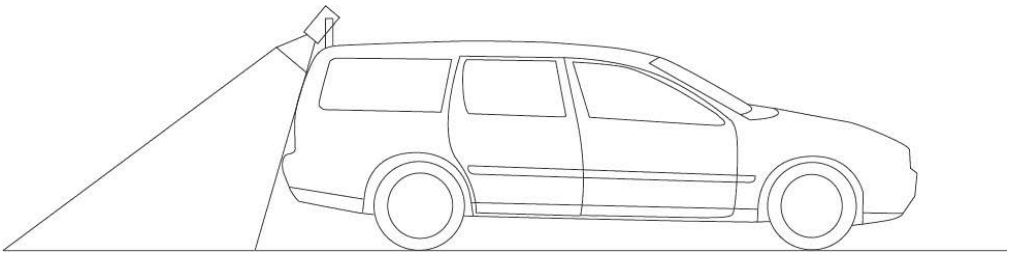

Figure 3–5. Instrumented vehicle with roof-mounted camera to record lateral position during test track driving.

3.2.4 Simulator driving

The driving simulator tests took place in late autumn after the test track part of the experiment had been completed. Test sessions were done at the same times during the evening and night as on the test track, to eliminate differences in circadian influences. The simulator scenario included two incidents (a car abruptly entering the road and a pedestrian crossing the road in front of the driver) that occurred one time each towards the end of the driving session, but was otherwise identical to the test track. See figure 2 in paper I for a comparison of the driver's visual impression in simulated and test track driving. The simulator is described further in chapter 3.1. Data on lateral position, speed, pedal use and steering wheel movements over the entire duration of the test sessions were extracted directly from the simulator computer and recorded with a frequency of 20 Hz.

3.2.5 Measurements

Driving outcomes

We considered the standard deviation of lateral position (SDLP) to be the most appropriate and relevant predefined primary outcome to validate and calibrate the driving simulator scenario for testing of drug impaired driving. It is a thoroughly validated measure of the degree of lane wavering as a marker of drug-impaired driving, and has been shown to correlate with BAC in a dose dependent manner in on-road driving tests (chapter 1.1.6).

SDLP primarily represents the automated control behavioral level of driving, the other two being the maneuvering control level and the executive planning (strategic) level. We therefore also included several possible alternative measures of drug impaired driving. Even though we classified reactions to sudden traffic incidents or observance of traffic signs and rules to represent the automated behavior level, these are complex outcomes that some researchers consider to represent higher driving behaviors (Berthelon and Gineyt 2014). In

any case, these outcomes may complete the picture of driving impairment at the automated behavior level together with SDLP. Steering wheel measures might also be relevant as a proxy to SDLP. The outcome measures included in the validation study are presented and classified according to the driving behavior level that they represent in table 3–3.

Table 3–3. Driving outcomes in the validation study.

Driving outcome	Level of driving behavior (Michon 1985, Walsh *et al.* 2008)
SDLP Observance of traffic lights[1] Driving behavior at incidents[1,2] Steering wheel measures[3]	Automated control level (automatic action patterns)
Brake/accelerator use intensity SD of speed	Maneuvering control level (controlled action patterns)
Average speed	Strategic level (general plans)

[1] Complex outcomes that may reflect functions also on higher behavior levels
[2] Foam rubber cubes (simulator and test track); car and pedestrian (simulator)
[2] Movement speed; movement per distance driven; reversal frequency; reversals per distance driven

Questionnaires

The participants completed an initial questionnaire upon inclusion in the study, which mainly concerned background demographics such as age, education, time in possession of a driver's license, driving experience, and alcohol/drug habits. During the trial, they filled in questionnaires before and after each drive, with items mainly covering their feelings of intoxication, mastery of the driving task, safety, sleepiness, degree of simulator sickness, alertness, whether they thought the drink had contained ethanol, and whether they thought the pill had contained a sedative drug. On the test track, driving instructors were asked to rate the test subjects' degree of intoxication and driving performance after each drive. Driving instructors were also blinded regarding the intervention. Many items asked the subjects to rate their subjective experiences on numerical scales from zero (typically "not at all") to 10 (typically "very much"). An example of such an item is shown in figure 3–6. In statistical analyses, we considered the data from such numerical scales as continuous variables.

C8 Hvor påvirket (av rusmidler) følte du deg under kjøringen?										
Ikke påvirket										Veldig påvirket
0	1	2	3	4	5	6	7	8	9	10

Figure 3–6. Example of numerical scale item used in the pre- and post-drive questionnaires.

BAC measurement

Venous blood was drawn by cubital venipuncture into EDTA vacutainers immediately before and after each drive. The blood was stored at -18°C until all driving tests were completed (max. 3 months), after which they were all analyzed in one batch. Blood ethanol concentrations were quantified using a headspace gas chromatography-mass spectrometry (GC-MS) method. In brief, 200 µL blood was mixed with 50 µL internal standard (d6-ethanol). Samples were left for 30 minutes to achieve equilibrium before the gas fraction was aspirated into an Agilent HP 6890-5973 GC-MS system (Agilent, Palo Alto, CA). Separation was performed on a J&W Scientific 123-9134 DB-ALC1 (30 m x 1.2 mm) column with a helium mobile phase and a run time of 0.90 min. Ethanol was monitored at m/z 31 and the internal standard at m/z 33. The level of quantification (LOQ) was 2 mmol/l (approx. 0.09 g/l). Between-day coefficient of variation (CV) calculated from quality control samples was 4.5 % at 5 mmol/l (0.22 g/l) and 1.8 % at 50 mmol/l (2.2 g/l).

3.2.6 Statistical analyses

Sample size

We conducted an *a priori* estimation of the sample size required to obtain significant results at the significance level (α) of 0.05 and power (1-β) of 0.95, utilizing the SDLP data from a test with five male subjects who drove for 1 hour in the driving simulator, with the same test scenario that was later used in the main study.

In order to obtain this, we performed paired *t*-tests of the differences in mean effects of BAC 0–BAC 0.5, BAC 0.5–BAC 0.9, and BAC 0–BAC 0.9 on SDLP in straight sections, with different N to see what was the lowest N possible in order to obtain significant results ($p < 0.05$). To calculate SD with N different from five, the following formula was used:

$$SD_n = \sqrt{\frac{5}{n}} \times SD_5$$

The sample size estimation showed that N = 11 was found to be the lowest possible in order to have significant results for the differences in mean effects of BAC 0–BAC 0.5 and BAC 0–BAC 0.9. Although the analyses showed 11 subjects would be sufficient to detect significant differences in BAC level influence on SDLP, we decided to aim for no less than 15 subjects completing the study, to allow for the considerable uncertainty in the assumptions of effect

size and standard deviation. On top of that, we had to consider the possibility of dropouts, for instance due to simulator sickness. Consequently, we invited 20 drivers to the study.

Methods to test hypotheses and relationships between variables

In the four articles published from the study, multiple hypotheses and relationships have been tested. Table 3–4 provides an overview of different dependent and independent variables and the statistical methods employed. In all the analyses in this thesis, two-sided *p*-values < 0.05 were considered significant. The analyses were performed in SPSS 18–21 and Stata 12.

Table 3–4. Overview of independent and dependent variables and statistical methods used.

	Independent variables	Dependent variables	Statistical methods
Paper I	BAC No. of test exposures Curve or tangent driving Quartile of trip duration	SDLP	Linear mixed model
	Test arena	BAC within each BAC group	Paired sample *t*-test
	BAC	Subjective rating of driving performance Driving instructor's rating of driving performance	Linear mixed model
Paper II	BAC	Steering wheel - reversal frequency - reversal per distance - movement per distance - movement speed Average speed SD of speed Accelerator use Brake use	Linear mixed model
	BAC group	Driving through red light (number of occurrences within each BAC group)	Paired sample *t*-test[1]

[1] Statistical method inaccurately reported in paper II. See chapter 9 (Errata) for details.

	Independent variables	Dependent variables	Statistical methods
Paper III	Simulator sickness score	SDLP Steering wheel - reversal frequency - reversal per distance - movement per distance - movement speed Average speed SD of speed Accelerator use Brake use	Linear mixed model
	BAC No. of simulator exposures	Simulator sickness score (log-transformed)	Linear mixed model
Paper IV	Perceived drug intake (yes/no)	SDLP	Independent sample *t*-test
	Subjective intoxication (BAC-adjusted)	SDLP	Linear mixed model
	Ethanol dose group	Assumed ingested ethanol dose Subjective intoxication	Linear mixed model
	Ethanol dose group	Perceived drink/pill content	Logistic mixed model
	Perceived drink/pill content (ethanol yes/no, sedative yes/no)	Subjective intoxication (BAC-adjusted)	Linear mixed model

Linear mixed models

A mixed model is a statistical model that includes both fixed and random factors/effects (i.e. mixed). A fixed effect is a variable of interest with an assumed effect on the dependent variable that would be repeated if the experiment were to be replicated. A random effect is not a variable under study, but rather represents a random selection from a larger population. Treatment or intervention levels are normally fixed effects, whereas the subjects typically correspond to a random effect. Mixed model analysis is a valid, flexible and easily interpretable analytical approach to data sets with several grouped or clustered observations of the same, continuous variable, for instance when there are repeated measurements within the same subject. Such observations are not independent, an important issue that the mixed model approach can accommodate, as opposed to ordinary ANOVA, which requires independent observations. Repeated measures ANOVA, on the other hand, handles correlations within subjects, but still has some disadvantages compared to a mixed model (McCulloch 2005). Notably, repeated measures ANOVA uses only subjects with complete

cases in the analysis, and the results are unbiased only if data are missing completely at random (MCAR). Mixed models, on the other hand, also includes information from subjects with partially missing data, and results are unbiased under the less restrictive missing at random (MAR) assumption. Another advantage of mixed models is that they handle uneven spacing between measurements (Seltman 2015), e.g., in our study, varying individual BACs within each intended BAC group. For all these reasons, the mixed model approach was considered the most appropriate for our data, since we dealt with repeated within-subject observations, missing data points, and highly variable individual BAC values despite our efforts to tailor the ethanol dose to each individual.

In a within subject, repeated measurement design analyzed with a linear mixed model, a linear regression equation is calculated for each subject, which is then treated as a single independent observation along with the other subjects' regression equations. The subject is a random factor, drawn 'at random' from the broader general population, to which it is the intention to generalize the results. The random factor is thus assumed to be uncorrelated to the independent variable(s) in the model, as opposed to fixed factors/effects (e.g., BAC level), which are expected to correlate with the independent variable(s). In this thesis, linear mixed models were used for the main analyses of BAC effects on driving outcomes in paper I and II. In paper III, we used linear mixed models to investigate the relationship between simulator sickness score and driving outcomes, as well as the relationship between BAC, number of simulator exposures and the degree of simulator sickness. In paper IV, linear mixed models were used for the analyses of subjective intoxication effects on SDLP, the relation of ethanol dose group to assumptions of ethanol dose and subjective intoxication, and the relationship between perceived drug intake and the degree of subjective intoxication. When analyzing the relationship between ethanol dose group and perceived drug intake, logistic mixed models were used, since the dependent variable is categorical (Rosner 2006).

The general criteria for using linear regression analyses are the following:

1. All observations must be independent
   In the study, for each subject, data were logged very frequently while driving. However, those are not independent observations. Therefore, in the analysis the mean value of a dependent variable for one subject and BAC level is regarded as a single observation.

2. The distribution of the residuals must be normal for all values of the independent variables

   To test the assumption of normality, analysis of standardized residuals is often used. If the underlying distribution of the dependent variable is normal, the standardized residuals should also be normal. Visual inspection of histograms and Q-Q plots of the standardized residuals may be used to determine if the assumption of normality is met. A histogram of the measured values of the independent variable should comply with the shape of a classic Gauss curve. In a Q-Q plot, the observed values of the independent variable along the x-axis are plotted against the value expected from normality along the y-axis. If the assumption of normality is true, the plotted values should fall along an x = y straight line. Analyses of kurtosis and skewness may also be used. Fortunately, linear regression analyses will give reasonable answers in spite of some deviation from normality. If the data are not normally distributed, one way to handle this can be to logarithmically transform the data, which was done in paper III for the dependent variable Simulator sickness score. The same tests for normality may then be applied to the log transformed data to see if the requirement is met.
3. The variance of the dependent variable must be the same for all values of an independent variable

   To test this assumption, standardized residuals may be plotted against the predicted value of the independent variable in what is called a residuals plot. If the assumption of equal variance is met, the residuals should be clustered symmetrically around the x-axis, with values mostly in the lower single digits and no trends of increase with higher predicted values or any other clear patterns.
4. The relationship between the independent and a dependent variable is assumed to be linear

   If the standardized residuals in the residuals plot described above are symmetrically scattered around the x-axis, this is a strong indicator of a linear relationship.

Statistical tests to compare means

Paired-sample *t*-test is used to compare means in two dependent samples, e.g. two measurements in the same subjects at two different time points. In this thesis, it has been

used for instance to test if there were significant differences in BAC between the test-track and simulator driving tests.

Independent-sample *t*-test is used to compare means in two independent samples, e.g. the same measurement in two independent groups of subjects. In this thesis, independent-sample *t*-test has been used to test whether drug expectant subjects had significantly different SDLP values compared to subjects without drug expectancy.

The central criterion for the use of *t*-tests is the assumption that the underlying distribution of the random variable is normal. For the standard independent sample *t*-test, equal variance in the two samples is also assumed (Rosner 2006).

4 Results

4.1 Summary of paper I

Comparison of driving simulator performance with real driving after alcohol intake: A randomised, single blind, placebo-controlled, cross-over trial

Paper I presents the SDLP results from the validation and calibration of a driving simulator scenario intended for studies on drugs that possibly cause driving impairment. The validation study compared simulated driving with real (test track) driving across three BAC levels.

The results showed statistically significant, dose-dependent increases of SDLP as a function of BAC in both the simulated and the real driving scenario. Baseline SDLP and the rate of increase with BAC were higher and showed a larger inter-individual variability in the simulator than on the test track. In the simulator, the BAC-related SDLP increase was significant already after 15 minutes' driving. In the simulator, SDLP values were higher in curved sections than in straight sections, whereas no such difference was found with real driving.

4.2 Summary of paper II

Evaluation of measures of impairment in real and simulated driving: Results from a randomized, placebo-controlled study

Paper II expands the validation and calibration of the driving simulator scenario to include other potential measures of driving impairment covering a broader range of driving behavior, such as speed and steering measures, pedal use and driving behavior at incidents. Effect sizes and relative sensitivity in the simulator vs. the test track were compared with those of SDLP.

Ethanol intake increased several steering measurements, average speed, standard deviation (SD) of speed, and pedal use frequency during simulated driving. With real driving on the test track, only some steering measurements and SD of speed increased significantly with BAC. Reaction to unexpected incidents and observance of red traffic lights were adversely affected by ethanol in the simulator but not on the test track. In general, BAC-related changes in the measured variables were smaller and less significant during real driving on the test track than during simulated driving. Whereas SDLP showed a relatively large effect size that was similar in simulated and real driving, all other measures demonstrated smaller effect sizes, with less pronounced BAC effects on the test track than in the simulator.

4.3 Summary of paper III

Driving simulator sickness: Impact on driving performance, influence of blood alcohol concentration, and effect of repeated simulator exposures

Paper III explores the possible impact of simulator sickness severity (SSS) on SDLP and other measures of driving. The possible influence of BAC or repeated exposures to the simulator on the severity of simulator sickness is also explored.

The mean SSS score across all drives were 2.5 on the 0–10 scale used, with a distribution that was highly skewed to the left. There was no evidence for an impact of SSS on SDLP. The only driving measures found to be statistically related to the simulator sickness severity were average speed and steering wheel reversal frequency, which both decreased with higher SSS. The impact of SSS on these measures seemed to be less pronounced at higher BACs. We found a statistically significant, negative relation between BAC and SSS, whereas the negative relation between the number of previous exposures to the simulator and SSS was not statistically significant.

4.4 Summary of paper IV

Effectiveness of ethanol blinding by use of a novel sham pill approach, and the impact of drug expectancy on lateral vehicle control in real and simulated driving

Paper IV reports the effectiveness of a novel approach to enhance blinding in experimental studies with ethanol by use of a sham pill. We also explored the possible impact of subjective feeling of intoxication and drug expectancy, i.e. perceiving to having received a drug when in fact driving sober, on SDLP.

Only 44 % of the placebo interventions were correctly identified (mostly due to misidentification of the placebo pill), as compared to 87 % of the alcohol interventions in the low-dose ethanol group and 79 % in the high-dose ethanol group, respectively. Under alcohol conditions, only 5 % misidentified their drink as not containing ethanol. No indication of an effect of drug expectancy or subjective intoxication on SDLP was found, either in the simulator or at the test course.

5 Discussion

5.1 Appraisal of the main findings

Driving simulators is a relevant approach to experimental DUI research, because laboratory testing on one hand may not accurately predict traffic risk in real life, whereas on-road studies on the other hand are complicated and expensive. Validity and sensitivity assessments are essential to ensure that driving simulators yield reliable and significant data.

Many validation studies of simulator driving against real driving with various driving tasks and outcomes have been published, often establishing relative validity (Mullen *et al.* 2011). The sensitivity of various driving simulation tasks and outcomes to ethanol effects have also been shown by others (Jongen *et al.* 2016). The unique feature of the work presented in this thesis is that we combined the validation of simulated driving against real driving with the calibration to ethanol. A high degree of similarity between the two test environments would increase the confidence that the driving simulator test gives valid and significant results at the levels of drug impairment that are relevant to traffic safety.

The findings show that the simulator test is sensitive to ethanol effects in the BAC range up to 0.9 g/l. Sensitivity to ethanol effects was shown for outcomes reflecting automated behavior (SDLP, steering wheel measures, observance of traffic lights), controlled behavior (SD of speed, frequency of brake/accelerator use) and executive planning behavior (average speed). However, only SDLP, steering wheel movement/speed, and SD of speed showed significant BAC-related effects in real driving with a similar driving scenario. Of these, SDLP showed a high degree of similarity between the two test sites, whereas SD of speed and steering wheel measurements showed considerably lower effect sizes in real driving. The results confirm the sensitivity and validity of SDLP as a primary outcome to study drug impaired driving, also in simulator studies. We did not succeed in identifying other valid and sensitive measures that could have extended the spectrum of driving behavior levels tested in the same scenario.

The thesis also explores two major factors complicating the execution of such a validation study, namely simulator sickness and ethanol blinding. Simulator sickness significantly decreased average speed and steering wheel reversal frequency, especially in the placebo group, which seems to cancel the apparent BAC effect on these measurements. This indicates

that for these measures, simulator sickness may lead to false conclusions about BAC-related effects if not accounted for. Simulator sickness did not significantly influence SDLP or any of the other driving outcomes. Simulator sickness seemed to be less pronounced at higher BACs. Our method of ethanol blinding, with the use of a placebo pill as an additional confounder, was effective in concealing the placebo condition but ineffective in concealing the ethanol conditions. We did not find any evidence of drug expectancy effects on SDLP.

There are important limitations to our findings. Some are common to all the papers in this thesis and discussed jointly in chapter 5.2. Bias and dropout due to simulator sickness, as well as the possibility of simulator sickness influencing driving measures directly, are discussed in chapter 5.1.3. Limitations due to imperfect blinding and the possibility of expectancy effects are dealt with in chapter 5.1.4.

5.1.1 Paper I: Validation of the simulator scenario with SDLP as primary outcome

The primary purpose of the validation study was to validate a simulator scenario for research on driving impairment due to drugs. We chose SDLP as the primary outcome of interest because of its well-established sensitivity to the impairing effects of ethanol and other CNS depressant drugs as well as its relevance to traffic safety. In on-road studies, the relationship between BAC and SDLP has been established since a study by Louwerens *et al*. (1987), the results of which have been used as a comparator in numerous on-road studies ever since (Verster and Roth 2011). To investigate the sensitivity of SDLP, we assessed the statistical significance of BAC-related changes, as well as the effect size compared to baseline. To assess external validity, we compared the aggregated results from simulated driving to those of real driving. A statistically significant BAC-related increase of SDLP in both simulated and test track driving would indicate relative validity of the outcome (Kaptein *et al.* 1996), but this alone would be a rather lax criterion of relative validity. Ideally, the relative effects (i.e., the slope of the regression line divided by the intercept value) of simulated and test track driving should be comparable as well (Blaauw 1982). Absolute validity would require numerical equivalence, i.e. that the absolute values (intercept and slope) should be of comparable magnitude. We also considered individual SDLP data to get an impression of individual variation in the comparability of simulator and test track results.

We found a positive, statistically significant dose-response relationship between BAC and SDLP in both simulator and test track driving. These results support the sensitivity and relative validity of SDLP for the assessment of drug impaired driving in the simulator scenario. The relative BAC effects in the two test arenas were comparable, with that of test track driving amounting to 78 % of that of simulated driving. The results did not meet criteria for absolute validity, as the numerical values for both intercept and slope differed between simulated and real driving. The SDLP values were higher at baseline, increased more (in absolute terms) with BAC and showed higher inter-individual variability in the simulator than on the test track.

Our findings of relative validity are consistent with other studies that have validated measures of lateral position by comparing simulated and real driving, although not in a DUI setting (Blaauw 1982, Blana and Golias 2002, Reed and Green 1999, Törnros 1998). In our study, the baseline (i.e., sober driving) value of SDLP vas 32 % higher in the simulator than on the test track (29.4 cm vs. 22.3 cm, respectively). This discrepancy is nevertheless much smaller than that found between simulated and real driving in earlier studies (i.e. Blaauw (1982): 88 % higher in inexperienced drivers and 46 % higher in experienced drivers; Blana and Golias (2002): 97 % higher in curved sections and 64 % higher in straight sections; Reed and Green (1999): more than twice the value of on-road driving). The higher degree of similarity between simulated and real driving in our study may be attributed to a more advanced and realistic simulator. The author of the first published validation study of lateral vehicle control noted that the lack of absolute validity appeared to be due to diminished perception of lateral translations, i.e. an absence of kinesthetic information, in a fixed-base simulator (Blaauw 1982). A later study showed greater lateral variation in a fixed-based simulator compared to an advanced moving-base simulator in the same scenario (Engström *et al.* 2005). It seems reasonable to assume that the more realistic the simulator is in terms of kinesthetic feedback, the higher the similarity of lateral vehicle control.

Interestingly, Blaauw (1982) in his study noted that the simulator was more sensitive than the real driving test to discriminate between experienced and inexperienced drivers. This could be related to our results, where it seems that the simulator more efficiently separates ethanol-impaired drivers from unimpaired drivers, as seen by the much steeper slope of the regression line showing SDLP as a function of BAC in figure 4, paper I. However, what is also

evident from the same figure is that the variability is much higher in the simulator. Increased variability may impede the simulator's ability to predict driving impairment. To further characterize this, we looked at individual data to see if any discernible patterns would emerge. In figure 5, paper I, the curves describe the individual relationships between BAC and SDLP in both simulated and real driving. There was a high degree of intra-individual similarity of SDLP results in the two test conditions in most participants. However, in a few, results did not correspond well, such as for instance subject 15, who attained a very high SDLP value in the simulator in the high BAC condition. It was our experience during the study execution that the participant's attitudes to the test situation varied, as did their reactions to ethanol, especially the high ethanol condition. Whereas most participants adhered to instructions, some individuals clearly experienced pronounced disinhibiting effects at the highest BAC dose. In the test track driving, the aspect of real danger as well as the presence of a driving instructor in the car probably had a restraining effect, causing the subjects to adhere to protocol. In the simulator, we believe that the disinhibiting effects of ethanol, coupled with the lack of real danger, caused some subjects to depart from the agreed experimental preconditions and instead behave as if the simulator were a video game. Thus, some of the observed BAC-related increase in SDLP, and much of the increased variance, may not be due to reduced ability, but rather due to a lack of effort in some subjects. This obviously renders their results invalid. However, it is difficult to define objective criteria to exclude invalid results due to 'attitude problems'. Furthermore, in a simulator validation study, we believe that the data should be presented 'warts and all', precisely to show all the possible challenges that the use of driving simulators poses. Thus, we chose to include the results from all completed drives in the analyses.

To be of use in the characterization of drug-impaired driving, a simulator test must fulfil two essential criteria: it must be sensitive to the effects of a 'benchmark drug' (i.e., ethanol) with known impairing effects on driving, and it must be externally valid. As for the sensitivity issue, a recent review reported significant increase from baseline of SDLP in simulated driving in 33 % of studies with BAC ≤ 0.3 g/l, 61 % of studies with BAC 0.31–0.6 g/l, and 96 % of studies with BAC > 0.6 g/l (Jongen *et al.* 2016). The review also reported that SDLP is a less sensitive parameter in simulated than on-road driving, since fewer studies found significant SDLP increases at medium BACs. Our findings at first glance seem to contradict this, since we found

both absolute and relative BAC-related increase in SDLP to be higher in the simulator. However, as already discussed, variability is much higher in the simulator, and this may influence sensitivity because differences of group means become less statistically significant. Thus, sensitivity may be lower in the simulator even though effect size is higher, as these are not synonymous.

As for the validity issue, relative but not absolute validity was achieved for SDLP in our scenario. Absolute simulator validity may be an unrealistic aim, given the large underlying variability in driver performance and the difficulty of fully reproducing the real driving experience in a simulated setting (Caird and Horrey 2011). Luckily, whereas relative validity is a prerequisite, absolute validity is seldom required (Törnros 1998). In research of driving impairment due to drugs, SDLP is a surrogate measure of the outcome of interest, i.e. traffic risk. As such, increased SDLP is a measure of reduced psychomotor control, which is one of several BAC-related impairments that may increase traffic risk (figure 5–1), and should not be interpreted as a direct measure of traffic risk neither in simulated nor on-road driving.

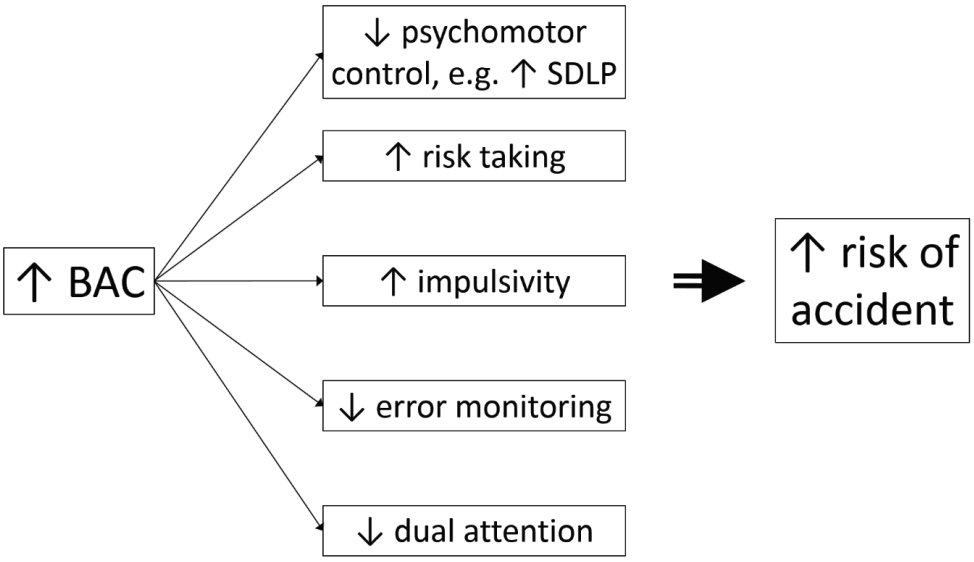

Figure 5–1- Some effects of ethanol on skills and behaviors relevant to driving and traffic accident risk

What is needed to interpret SDLP findings in terms of degree of impairment, is a 'translation factor' between SDLP results obtained in a certain test scenario, and an estimate of the corresponding relative accident risk. BAC may constitute such a translation factor, since the

relationship between BAC and accident risk is well known. Since *per se* legislation concerning BAC is in place in most countries, it is also of interest to compare the impairment of other drugs to that of ethanol. The calibration of the simulator scenario with three BAC levels means that the regression equation from the validation study may be used to quantify impairment from other drugs in terms of 'ethanol equivalents'. In an attempt to harmonize DUI legislation in Norway concerning ethanol and non-ethanol drugs, a similar approach has already been applied to construct *per se* legal limits of a range of common non-ethanol drugs equal to BACs of 0.2 g/l, 0.5 g/l and 1.2 g/l (Vindenes *et al.* 2012).

Another approach would be to use comparisons with the impairment seen at certain BAC intervals (e.g. BAC < 0.5 g/l, 0.5–0.8 g/l or > 0.8 g/l) to classify drugs as being presumably safe, or likely to cause minor or major impairment (Kenntner-Mabiala *et al.* 2015). This could be fruitful if the goal is to determine whether a certain drug is safe at therapeutic doses, and could be used to guide advice to patients. However, for legal and legislative purposes, such an approach would not make sense, as the impairing effects of most non-ethanol drugs are dose-dependent. Thus, psychoactive drugs are generally not 'safe' or 'unsafe' *per se* – but they may be more or less safe according to dose or, more accurately, blood concentrations.

There are of course many uncertainties with an attempt to 'translate' experimental results to ethanol equivalents or relative accident risks. First, there can be only indirect evidence of a relationship between SDLP and actual traffic risk, since a direct relationship is not possible to prove experimentally. Second, the assumption that there are similar relationships between SDLP and traffic risks for different drugs and drug classes, or that the driving impairment from any drug should somehow be comparable to the driving impairment from ethanol, is of course an oversimplification. The mechanism of action of ethanol is very complex, giving rise to both stimulating and sedative effects as well as reduced error monitoring, whereas non-ethanol drugs often have more specific effects. Thus, even the comparison between ethanol and other CNS depressant drugs is not straightforward. As for CNS stimulants or other drug classes, the utility of SDLP as a marker of driving impairment is dubious, and attempts at comparing effects with the effects of ethanol may be ill advised.

Our SDLP findings in the simulator are consistent with those of other recent simulator validation studies using ethanol as a benchmark drug (Berthelon and Gineyt 2014, Kenntner-

Mabiala *et al.* 2015, Mets *et al.* 2011, Veldstra *et al.* 2012), which also reported significant concentration-dependent effects on SDLP. However, the absolute SDLP values at baseline and the magnitude of BAC-related increase of SDLP varies between studies. Veldstra *et al*. (2012) found a placebo mean SDLP of 15.8 cm and an increase from baseline (ΔSDLP) of 4.1 cm in the BAC 0.8 g/l group. Kenntner-Mabiala *et al.* (2015) found slightly higher values, depending on type of route, with the highest baseline value (21 cm) recorded during driving on a winding, forested route. We found higher SDLP at baseline (29.4 cm) and a larger BAC-related increase, corresponding to a ΔSDLP of 10.6 cm calculated at a BAC of 0.8 g/l. Our results are more comparable to those of Mets *et al*. (2011) and Bertheleon and Gineyt (2014), who reported baseline SDLP and ΔSDLP at BAC 0.8 g/l of 28.0 cm and 5.8 cm, and 26.4 cm and 5.5 cm, respectively. However, due to differences in simulator design and scenario choice, it is probably not realistic to obtain corresponding absolute values of baseline SDLP or ΔSDLP between studies. This underlines the necessity of validation of the specific simulator and driving scenario in use.

The SDLP values obtained with real driving on the test track were lower than in the simulator, as discussed previously. SDLP at baseline was 22.3 cm, whereas the calculated ΔSDLP at a BAC of 0.8 g/l was 6.1 cm. As a comparison, aggregated data from several placebo-controlled on-road studies show that the mean SDLP at the sober state in monotonous highway driving is 18.8 cm (Verster and Roth 2011) , whereas a reference study reported ΔSDLP at a BAC of 0.8 g/l of 4.3 cm (Louwerens *et al.* 1987). The higher values in our study probably reflects the more challenging driving scenario.

SDLP is a sensitive and thoroughly validated measure of lateral vehicle control, reflecting tracking ability and vigilance, which are of vital importance to safe driving and should never be neglected when assessing drug impaired driving (Owens and Ramaekers 2009). However, it only represents the automated control level, which is not the only behavioral level of concern when characterizing drug effects on driving. Arguably, the particular advantage of simulators may lie just with the possibility of investigating higher levels of driving behavior. We therefore explored the sensitivity of alternative outcomes representing a broader range of behaviors in paper II.

5.1.2 Paper II: Validation of alternative driving outcomes in the simulator

After having established the sensitivity and relative validity of SDLP as the primary measure of impaired driving due to drugs in the simulator, we similarly investigated the sensitivity and validity of other potential measures of impaired driving, representing a broader range of driving behavior than SDLP alone. To assess the sensitivity of the outcomes, we considered the statistical significance of BAC-related changes, as well as their effect sizes, measured as the relative effect of a BAC of 1 g/l compared to baseline. To assess external validity, we considered the statistical significance of the BAC-related changes in the same outcomes on the test track, and compared the relative effect sizes of all outcome measures in the simulator to those of real driving. A statistically significant BAC-related increase in both simulated and test track driving would indicate relative validity of the outcome, although a large difference in relative effect sizes between the simulated and the real driving scenarios would cast doubt on its validity. Thus, to be able to establish relative validity, the relative effect sizes of simulated and test track driving should be comparable. Absolute validity would require that the absolute values (intercept and slope) should be of comparable magnitude as well. The data from these assessments can be found in table 2, paper II.

The assessment of sensitivity showed that outcomes representing all three main behavioral levels of driving (Michon 1985, Walsh *et al.* 2008) were statistically related to BAC in the simulator, including measures of automatic behavior (quantitative steering measures), measures of controlled behavior (SD of speed, brake/accelerator use frequency) and measures at the strategic level (average speed). However, the relative BAC effects for many of the outcomes were small, as shown in table 2 in the paper. For instance, the relative increase in average speed at a BAC of 1 g/l compared to baseline level was only 13 %, and that of steering wheel reversals per distance driven a mere 2.7 %. None of the outcomes tested had a relative effect equal to that of SDLP (45 %).

While all BAC effect estimates were positive also on the test track, only steering wheel movement speed, steering wheel movement per distance driven, and SD of speed, increased significantly with BAC. Furthermore, the relative BAC effects on the test track were even smaller than in the simulator. The relative BAC effect on the test track divided by the relative BAC effect in the simulator may quantify the external validity of the outcome measure. A

value close to unity indicates an equal BAC effect in the simulator and on the test track; whereas a low value signifies a less pronounced BAC effect on the test track than in the simulator. For the outcomes validated in paper II, these values were all rather low, and ranged from 0.10 to 0.48. By comparison, SDLP showed a much greater similarity between simulated and test track driving, with a value of 0.78.

At the potential crash incidents in the simulator (objects to be avoided, car and pedestrian entering the road), increasing BAC caused significant increases in average speed, standard deviation of speed, and steering wheel movement speed. However, no significant changes in these measurements were observed at similar incidents on the test track. No collisions occurred. The observance of red traffic lights decreased with increasing BAC in the simulator, whereas no such effect occurred on the test track.

A discussion of the validity of driving performance outcomes should not only be limited to quantitative results and statistical significance. In addition, the outcomes must be interpreted in the context of the driving scenario and the instructions given to the study subjects:

- Average speed reflects the strategic level in our rather challenging driving scenario, with a narrow, curvy road laid out in hilly terrain, since the subjects were free to choose their preferred speed. In this setting, increased average speed is a relevant measure of risk willingness and ability of self-assessment (Fillmore *et al.* 2008, Zhang *et al.* 2014). Average speed has also been shown by others to be positively correlated with BAC in simulator scenarios resembling ours (Zhang *et al.* 2014), as well as in city driving scenarios (Veldstra *et al.* 2012). However, we found a low rapport between simulator and test track results, and no significant increase with real driving. One plausible explanation would be that in the simulator setting, the lack of real danger combined with the disinhibition and impulsivity caused by ethanol leads to higher speeds, whereas in real driving, the awareness of the real danger of speeding and the presence of a driving instructor in the car opposes such effects. It could be argued that safe driving would normally dictate lower speeds at high BACs to compensate for driving impairment. With this in mind, even stable speed may be seen as a sign of impaired self-assessment due to ethanol (Mitchell 1985). As discussed in paper III (chapter 5.1.3), we later found that simulator sickness decreases average speed, and at the same time, there is an inverse relationship between

BAC and simulator sickness. Thus, it seems that the observed increase in average speed at higher BACs may be mediated by less simulator sickness, and not the BAC increase *per se*. Taken together, these shortcomings cast doubt on average speed as a valid measure of ethanol driving impairment in our simulator scenario.

- SD of speed has been reported as a measure of longitudinal control with relevance to traffic safety, e.g. by Zhang *et al*. (2014), but the interpretation is highly dependent on the type of scenario and instructions to the drivers. The hypothesis behind including it as a measure of drug impaired driving is that ethanol impairs the timing of accelerations and decelerations, which will increase the frequency and intensity of speed changes. In the Dutch on-road test, SD of speed is included as a secondary outcome, but has not been conclusively linked to increased accident risk (Verster and Roth 2014). The experimental conditions (narrow and winding test road and lack of instructions to maintain constant speed) may limit the use of SD of speed as a variable in our model. Arguably, a low SD of speed is not necessarily a sign of adequate speed adjustment in a scenario with many curves. However, in our experiment, SD of speed shows a clear dose-response relationship with BAC in both test arenas, regardless of whether it was measured during driving in curves, straight sections or overall. Thus, SD of speed may still be a variable of interest in the present context.
- Brake and accelerator use are measurements of longitudinal vehicle control. Although we found highly significant correlations with BAC and rather large relative effect sizes in the simulator, brake pedal use was much less frequent and not significantly related to BAC in real driving. The reason for this may be that a lack of sensory feedback in the simulated setting makes speed adjustments difficult. The large discrepancy suggests that recordings of brake use in the simulator will not even remotely reflect brake use in real driving, which renders this variable unsuited to characterize impaired driving in the real world. Accelerator use was not available in test track driving; therefore, the external validity of this variable cannot be properly assessed. However, considering the theory of speed adjustment difficulties in simulated driving, the same conclusion probably holds true for accelerator use as for brake use. External validity is thus dubious for these measurements in our scenario.

- Steering wheel measures would logically be expected to show similar patterns to SDLP, but this was not the case. We found small relative effect sizes, especially in test track driving. The driving simulator was developed and internally validated to be representative of the properties of a Renault Scenic, whereas the instrumented vehicle was a Volvo V70. Differences in steering wheel backlash, or insufficient fidelity of the equipment recording steering wheel movements in the instrumented vehicle, could explain the discrepancies between simulated and real driving to some degree. Another explanation may be that steady steering is probably more difficult in the simulator, due to lack of haptic feedback and inertia that is present with real driving. The significant BAC-related increase in steering wheel reversal frequency found in the simulator may be a result of less simulator sickness at higher BACs, and not higher BACs *per se*, as discussed in detail in paper III.
- Unexpected incidents in the form of objects in the road to be avoided by the study subjects, as well as a car and a pedestrian entering the road in the simulator scenario, served as potential crash events. The lack of collisions obviously implies that, although clearly a measure with relevance to traffic safety, the occurrence of crashes is too insensitive at low to moderate BACs to be a primary measure of interest. One could argue that this could be resolved by constructing more challenging potential crash situations. In fact, this possibility is often cited as one of the main advantages of simulated driving (Kenntner-Mabiala *et al.* 2015). However, the inclusion of very high-risk situations, which would only be ethically feasible in the simulator and not on the test track, would work against the aim of validating a realistic driving scenario. The inclusion of adrenaline-provoking events would also counteract the fatigue-provoking features of the scenario, which could decrease the sensitivity of outcomes that respond to sedative effects of drugs, such as SDLP. Significant increases in average speed, SD of speed, and steering wheel movement speed at the incidents were observed with higher BACs in the simulator, but these were not reproduced on the test track. This indicates that ethanol does indeed change some parameters of driving at potential crash events in the simulator, but it is very difficult to interpret the relevance of these changes to traffic safety. For instance, increases in SD of speed may rather be seen as an adequate coping strategy. Other investigators have also concluded that variables requiring reactions to sudden events are insensitive to alcohol effects (Berthelon and Gineyt 2014, Veldstra *et al.* 2012). One

explanation for insignificant findings at potential crash events may be a high variability of intra-individual performance due to learning effects, which may overshadow BAC effects, even in a randomized study design. Another possible explanation is the existence of multiple alternative avoidance strategies to sudden events (Kenntner-Mabiala *et al.* 2015, Veldstra *et al.* 2012).

- Observance of red traffic light decreased significantly with higher BAC in the simulator, but not on the test track. Limitations in graphic contrast and luminosity in the simulated scenario to some degree may explain why this measure was more sensitive in simulated than real driving. In addition, the observance of traffic lights is an all-or-nothing response that is deeply ingrained in drivers, for which the threshold of significant effect may be too high to be a sensitive measure at low to moderate BACs.

In summary, compared to SDLP, all the alternative outcomes showed lower relative effects of BAC, and there is a rather large discrepancy between the results in the simulator and on the test track for most outcomes. The sensitivity to BAC effects were highest for measures at the automated level of driving behavior. This is in line with previous research, which has also shown that complex driving behaviors are less sensitive to low-to-moderate BACs (Jongen *et al.* 2016), probably because these behaviors are flexible and allow several different strategies to compensate for intoxicating effects (Veldstra *et al.* 2012). SD of speed, steering wheel movements per distance driven and steering wheel movement speed increased significantly with BAC in both test arenas. This may be taken as evidence of relative validity for these measures; however, the relative BAC effects were much lower with real driving than with simulated driving, which casts doubt as to the external validity of these outcomes as well. External validity should not be understood as an all-or-nothing phenomenon, and we think that the comparison of effect sizes between simulated and real driving is a useful quantitation of external validity. The steering wheel measures essentially reflect the same automated behavior as SDLP; thus, we see no advantage in supplementing SDLP with these measures, which are bot less sensitive and less valid compared to SDLP. SD of speed on the other hand is a measure of longitudinal control as opposed to lateral control for SDLP, and reflects maneuvering control behavior to a higher degree than SDLP, which mainly reflects the automated behavior level.

In conclusion, with the cautious exception of SD of speed, we did not succeed in establishing valid and sensitive outcomes that reflect higher levels of driving behavior. The results support the use of SDLP as a primary outcome and SD of speed as a secondary outcome in our driving scenario, which is similar to the choice of outcomes in the Dutch on-road driving test (Verster and Roth 2011). We did not find convincing evidence for the use of any of the other outcomes that we explored in our scenario. Complex behaviors allow multiple compensatory strategies, rendering single parameters insensitive. Precisely because of this, aggregate parameters that count the total number of errors has been proposed as a solution to this sensitivity problem at the higher behavior levels of driving (Kenntner-Mabiala *et al.* 2015, Shechtman *et al.* 2009). Kenntner-Mabiala *et al*. (2015) found an aggregated outcome of driving errors to be the only outcome that significantly discriminated between all the BAC levels tested in their recent validation study. In this study, they validated the driving scenario that they had previously developed to show differential effects on driving impairment of the antiepileptics carbamazepine and oxcarbazepine (Kaussner *et al.* 2010). The authors stated that their next step would be to compare performance in the driving simulator with that of real driving on a test course, to establish the representativeness of their approach further (Kenntner-Mabiala *et al.* 2015). This is similar to our strategy for validation (although conducted in a two-step fashion in two separate studies instead of testing both sensitivity and external validity in one study), and also in line with the only other example of comparison of ethanol effects between simulated and closed course driving that I have been able to identify (Gawron and Ranney 1988). The latter study did not employ directly comparable scenarios or outcomes between the simulated and the closed course settings.

5.1.3 Paper III: Simulator sickness: Influencing factors and consequences

The purpose of this paper was to investigate the possible influence of simulator sickness on the driving performance outcomes in the study, as well as the possible relationship between BAC, repeated exposures to the simulator, and the severity of simulator sickness. The importance of this lies in the recognition of simulator sickness as a possible confounder, as well as a source of possible bias, in driving simulator studies in general and in the present simulator model in particular.

We found no indications of simulator sickness influencing the primary driving performance outcome SDLP, nor did we find evidence of an interaction effect between simulator sickness scores and BAC. This affirms the confidence in SDLP as a robust measure of driving impairment due to drugs in simulator studies, and is an important contribution, since little has been published in this field previously. Our data confirm the findings in another study showing no effect of simulator sickness on lateral vehicle control (Muttray *et al.* 2013).

The only outcomes that were statistically related to the degree of simulator sickness were steering wheel reversal frequency and average speed (table 1, paper III). We also found a significant interaction with BAC in the case of steering wheel reversal frequency, whereas in the case of average speed, this interaction was not statistically significant. When we explored the differential effects of simulator sickness on steering wheel reversal frequency at different BAC levels, the effects of simulator sickness seemed to be mitigated at higher BACs (figure 3, paper III). In paper II (Helland *et al.* 2016), we reported that both steering wheel reversal frequency and average speed increased significantly with BAC. However, it seems that when simulator sickness is introduced as an explanatory factor in the model, the BAC effect is cancelled. The data suggest that these two outcomes are not sensitive measures of driving impairment in our scenario, but rather reflect aspects of driving style that are particularly sensitive to simulator sickness. Subjects affected by simulator sickness may seek to alleviate their symptoms by adopting a more cautious driving style, with lower speed and more guarded steering. At the same time, there is an inverse relationship between BAC and simulator sickness in the present study. Hence, we believe that the apparent BAC-related increases in average speed and steering wheel reversals described in paper II may actually be a consequence of less simulator sickness at higher BACs. The finding that the effects of simulator sickness on speed and steering wheel reversals are less pronounced at higher BACs supports this notion. In more formal terms, the interpretation of our findings may be that SSS acts as a mediator of the BAC effects on these outcomes, while at the same time being moderated by BAC (figure 5–2) (Muller *et al.* 2005). This interpretation also explains why the apparent BAC-related increases in these two outcomes were observed only in the simulator and not during test track driving.

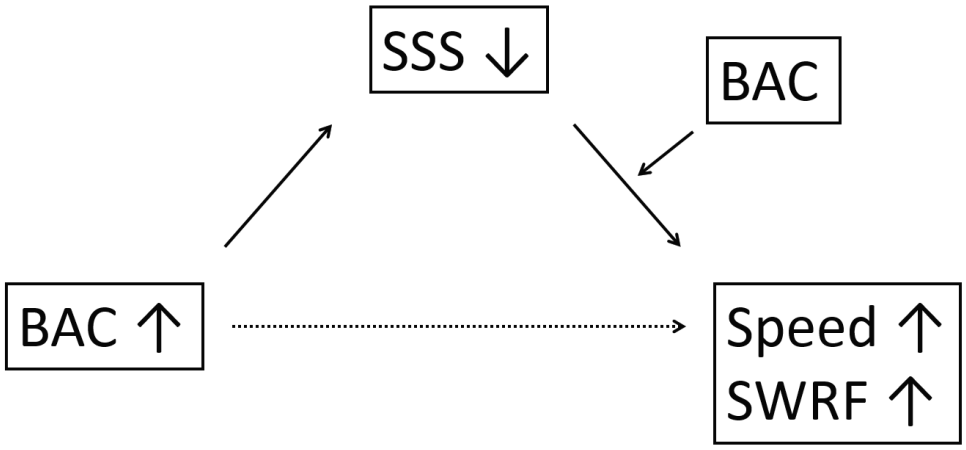

Figure 5–2. Model showing the role of simulator sickness severity (SSS) as a mediator of the apparent BAC effect on the outcomes average speed and steering wheel reversal frequency (SWRF). In addition, BAC may moderate the effects of SSS on the outcomes, as shown by the differential effects of SSS at the three BAC levels tested (see figure 3, paper III).

Ethanol inebriation seems to alleviate simulator sickness, the effect amounting to 1.6 points at a BAC of 1 g/l on the zero to 10 ordinal scale that we used to measure simulator sickness (figure 4, paper III). To our knowledge, this has not been shown before. The mechanism behind this finding is not clear, and at first glance, the finding may seem paradoxical, as many associate ethanol with nauseating effects. However, at low to moderate BACs, nausea is not a common ethanol effect. Ethanol has multiple mechanisms of action in the CNS, affecting many neurochemical circuits and systems (Spanagel 2009), but is classified as a CNS depressant, and has the effect of dulling sensory input. In the context of simulator driving, one may speculate that ethanol, in accordance with ethanol myopia theory (Moss and Albery 2009), may suppress sensory input to the extent that the subjects do not experience the sensory conflict by which simulator sickness most likely occurs (Brooks *et al.* 2010). This effect of ethanol may be an important contribution to the differential effects of simulator sickness at different BAC levels on certain driving parameters, as discussed above. It may also introduce bias by causing lower dropout rates due to simulator sickness with higher BAC.

Although statistically insignificant, the numerical results may also be consistent with habituation to the sickness-inducing effects with repeated exposures, as has been shown by others (Domeyer *et al.* 2013, Kennedy *et al.* 2000). The exclusion of those most vulnerable to simulator sickness, both through general exclusion criteria (i.e., age, sex) and pre-study

screening for risk factors and actual simulator sickness during familiarization, may possibly explain why this relationship was not significant in our study. For instance, elderly subjects, excluded from our study, are both more vulnerable to simulator sickness and habituate more slowly to the simulator environment than young subjects (Kawano *et al.* 2012).

The most important weakness of this study was that we did not use a standardized tool for measurement of simulator sickness, i.e. the Simulator Sickness Questionnaire (SSQ), which is the 'gold standard' to quantify the severity of simulator sickness (Kennedy *et al.* 1993). By using an overly simplistic measurement tool (ordinal scale from 0–10 where the subjects were simply asked, "to which extent did you experience simulator sickness" at the end of each drive), there is a risk that we underestimated the true extent of simulator sickness in our study. There is also a risk that the subjects interpreted the question differently, although it is reasonable to believe that each subject interpreted the question similarly between different sessions, which is more important in a within subject design. Lastly, the use of a simple ordinal scale instead of a multi-item combination variable caused our data to become highly skewed. In the statistical analyses, this was handled by log transforming the simulator sickness score when it acted as a dependent variable. The SSQ is very extensive, featuring 16 items, which makes it difficult to fit into the busy schedule of a randomized controlled trial. Also, the questionnaire has been criticized to include some items that are not specific to simulator sickness (Muttray *et al.* 2013) – for instance, fatigue and difficulty concentrating may just as well be symptoms of ethanol inebriation or simply an effect of boredom and tiredness from exposure to a long-lasting simulator scenario. Hence, not using the SSQ may have had its advantages, and may have led the subjects to focus on the core symptom of simulator sickness, i.e. queasiness/nausea. Other recent studies of simulator sickness have also used a simplified ordinal scale similar to ours (Bridgeman *et al.* 2014).

Apart from this, the most important limitations of the study included a rather small sample size and lack of power estimates, which means that there is a risk of false negative findings. In light of the exploratory nature of our study and its shortcomings with regard to sampling bias, power and measuring technique of simulator sickness severity, our findings should be regarded as tentative. See also chapter 5.2 for a discussion of general limitations of the study.

5.1.4 Paper IV: Challenges related to ethanol blinding

Blinding of the intervention, either from the point of view of the subject (single-blind) or from the point of view of both subject and investigator (double-blind), is often seen as an integral and essential part of the design of an experimental study. The purpose is to balance expectancy effects across the intervention groups, so that the observed differences between groups will only be due to the 'true' intervention effect. This is standard procedure in RCTs investigating drug effects, and is no less important in ethanol studies, as the subjects' familiarity with ethanol means that expecting alcohol is likely to produce particularly strong effects. At the same time, this very familiarity makes ethanol blinding particularly difficult, as the sensory cues and psychoactive effects of ethanol even at low BACs enable the subjects to identify the ethanol intervention. It is a paradox that while blinding is seen as essential and proven to be of major consequence to the results in experimental trials, the success of the blinding procedure is seldom reported or, if reported, often ignored. When the direction of pharmacological and expectancy effects of a drug is the same, the consequences of unsuccessful blinding will be an exaggeration of the perceived 'pure' pharmacological effect of the intervention, since the intervention group receives both expectancy and pharmacological effects, whereas the placebo group receives no effect (and is thus more comparable to a non-intervention group). In the case of compensatory alcohol expectancy effects opposing the pharmacological effects of the drug, which have been shown for instance in simple tracking tasks (Finnigan *et al.* 1995), unsuccessful blinding would have the opposite consequence (i.e., underestimation of the 'pure' pharmacological effect).

What is the measure of a well-functioning blinding procedure? A blinding procedure after which half the subjects believed having received placebo and the other half believed having received active treatment, with no correlation between the belief and the actual intervention received, would no doubt be considered successful. However, such a result is very difficult to obtain in a study with ethanol. Successful blinding may be achieved in the placebo group or with very low doses of ethanol, but as ethanol doses increase, few subjects are duped by the blinding procedure (Conrad *et al.* 2012, Martin and Sayette 1993). In addition, those in the placebo group who believe they have received ethanol, tend to give considerably lower estimates of the ethanol dose than those actually receiving ethanol (Testa *et al.* 2006). Nevertheless, successful blinding may not depend on the utopian goal that the subjects'

ability to guess the intervention should be equal to chance. The aim of blinding is to balance expectancy effects across the placebo and intervention conditions, so that the difference represents the 'true' intervention effect. Since the concealment of ethanol has proven to be difficult as the doses increase, an alternative approach would be to increase the expectancy in the placebo group as much as possible by introducing 'false' cues, while at the same time decrease expectancy as much as possible in the intervention groups by taking steps to conceal the ethanol content. The measure of success would then be equal percentages reporting the belief of having received active intervention across the intervention groups. This proportion need not necessarily be 50 %. In addition, the participants' estimates of their ingested ethanol dose should not be too dissimilar across the intervention groups.

As for the provision of 'false' cues in the placebo group, researchers commonly smear ethanol on the rim of the container or dribble a small amount on top of the drink to deliver sensory cues of ethanol (Charlton and Starkey 2015, Conrad *et al.* 2012). Several behavioral manipulations may also be used, such as creating a 'bar atmosphere', using sealed bottles with familiar alcohol trademarks, etc. In some experimental designs, notably the 'balanced placebo design', participants are also actively deceived in the 'expect alcohol/receive placebo' condition (Rohsenow and Marlatt 1981).

Ways to conceal ethanol content in the active intervention groups include sensory confounders such as the addition of spicy ingredients (e.g. Tabasco sauce) to the drink (Mørland *et al.* 1974) or make the subjects use strong mouthwash to numb the senses of taste and smell prior to drinking (Rohsenow and Marlatt 1981). Providing a sham pill as an alternative explanation to internal cues of ethanol inebriation has also been advocated to increase credibility of the anti-placebo condition (i.e., 'receive ethanol/expect placebo') of the balanced placebo-design (Epps *et al.* 1998). The familiar psychoactive and somatic effects of ethanol may be less palpable if the ethanol doses are kept as low as the experimental purpose allows, and administered over a long period of time to avoid rapidly rising BAC.

In our study, we used a set of manipulations to balance expectancies. The placebo drinks were spiked with vodka essence to mimic vodka taste. All drinks were diluted with rather large amounts of fruit juices, cooled with ice, served in lidded containers and were to be sipped through a straw to avoid obvious sensory ethanol cues. The drinks were administered

over a long period of time to avoid pronounced 'high' effects. In addition to such techniques, which are not unique to our study, we also attempted to confuse the participants by administering a placebo pill along with each drink. The subjects were told that the drink may or may not contain ethanol, and that the pill may or may not contain a sedative drug. Thus, their expectations were that they could receive either ethanol, a sedative, both, or none. This may be considered a version of the design coined the 'Alternative substance paradigm' by Conrad *et al*. (2012), where the expectancy of an alternative drug to ethanol is introduced although ethanol is the only drug that is actually being administered. Our design is unique in that we used a sham pill to enhance the 'alternative drug' expectancy, whereas in the methodology used by Conrad *et al*., subjects are simply told that the drink they receive may contain the alternative drug. Our hypothesis was that the use of a placebo pill would increase placebo response (i.e., expectancy), especially in the placebo group.

In line with our hypothesis, the sham pill indeed seemed to increase drug expectancy in this group, as the subjects believed they had received a sedative in 41 % of all placebo trials, whereas they believed they had received ethanol in only 15 % of the trials (table 1, paper IV). Thus, the majority of subjects believed they were under the influence of ethanol or a sedative when they in fact were sober. This is a slightly higher proportion of drug expectancy in the placebo group than in the study of the 'Alternative substance paradigm' by Conrad *et al*. (2012), and may be taken as an indication that the inclusion of a sham pill may increase expectancy and thus contribute to balancing expectancy effects across intervention groups.

On the other hand, few subjects misidentified the beverage content with our design, and only a very few subjects believed having received a combination of substances. Thus, the blinding procedure with the use of a sham pill worked well in the sober condition, whereas few were deceived by this approach in the alcohol conditions. Also, the belief of having ingested a sedative pill was associated with less subjective intoxication than the belief of having received ethanol. After adjusting for actual ethanol intake, the alcohol expectancy effect amounted to 2.6 points on a 0–10 intoxication scale, whereas the expectancy of having received a sedative pill was associated with a 1.5-point gain on the same scale. The equivalence of 'ethanol expectancy' and 'sedative expectancy' is thus questionable. A larger effect of expecting

alcohol than expecting a sedative is reasonable, as the participants' experience with ethanol would be much more extensive than their experience with sedative drugs.

Nevertheless, under the assumption that 'ethanol expectancy' and 'sedative expectancy' may to some degree express the same phenomenon, we investigated the effects of overall 'drug expectancy' on lateral vehicle control, both in simulated and real driving. In the placebo group, close to half of the participants fell into either the expectancy or the non-expectancy group. Since the effects of ethanol and sedative expectancy are not equal in terms of the subjective intoxication that they produce, we also explored the relationship between subjective intoxication score and SDLP, after adjusting for the effect of actual ethanol intake. We found no evidence of an impact on SDLP in either of the analyses (see tables 2 and 3, paper IV). A lack of expectancy effects on simulated driving was also shown in an early simulator study (Rimm *et al.* 1982), whereas the results of a more recent simulator study suggested possible expectancy effects on SDLP (Charlton and Starkey 2015) (see paper IV for a more detailed discussion). Our results could in fact be compatible with a compensatory effect of drug expectancy in simulated driving, since the effect estimates on SDLP of both the perceived drug intake and the subjective intoxication, although not statistically significant, were both negative. Compensatory effects of ethanol expectancy have been shown previously for other measures of ethanol effects (Finnigan *et al.* 1995, Marczinski and Fillmore 2005). The relatively small sample size and lack of specific power calculations preclude any strong conclusions from our study, but the negative SDLP effect estimates at least do not point towards an impairing effect of expectancy on lateral vehicle control.

5.2 Methodological considerations, strengths, limitations and weaknesses

Very few published studies have attempted to calibrate several aspects of driving impairment in a driving simulator to different BAC levels, and even fewer have validated the research by comparing the results to real driving, one early and notable exception being Gawron and Ranney (1988). One recent study by a Dutch group used this approach with cannabis as the active drug (Veldstra *et al.* 2015), and another research group in Germany has stated its intention to proceed with a validation against closed course driving after having calibrated their driving simulator scenario to several BAC levels (Kenntner-Mabiala *et al.* 2015). This dual

approach is the main strength of our validation study. We have also delved into two important challenging features of such studies, namely simulator sickness and aspects of ethanol blinding and expectancy effects, both areas in which the scientific literature is limited.

The strengths of our study include a strong theoretical and empirical basis for the scenario that was used, as we took the existing theoretical framework (Michon 1985), guidelines for research on drugged driving (Walsh *et al.* 2008) as well as own experience from pilot studies into consideration when planning our driving scenario. As prominent researchers in the DUI field have noted, many studies lack proof and theoretical rationale that their test battery actually measures driving skills relevant to traffic safety (Owens and Ramaekers 2009). However, our opportunities were limited, as we were restricted to use one specific closed course with one instrumented vehicle as our external validity control. This meant that we could not merely replicate for instance the Dutch highway driving or car following scenarios, which have already been calibrated and validated elsewhere and as such would have been a much easier approach. Instead, we sought to extend the validity of SDLP as a marker of drugged driving in monotonous highway scenarios to scenarios that are more relevant to the setting of many serious traffic accidents in Norway, namely nighttime driving on narrow and winding rural roads. This fitted the layout of the closed test course that was available to us well. The properties of the closed course were carefully replicated in the driving simulator scenario to ensure that the validation against real driving would be as realistic as possible.

On the other hand, our scenario did not include optimal driving tasks for assessing higher levels of driving behavior, such as a car-following task or complex driving situations. First, the technical equipment both in the instrumented vehicle and in the simulator restricted the range of measurements that were available to us. Second, from our pilot studies, our experience was that complex scenarios did not yield significant results. Thus, we devised a long and monotonous driving scenario to maximize the sensitivity of measures of automated driving behavior, which is the level of driving behavior that has consistently been shown to be most sensitive to ethanol effects, while at the same time keeping some measurements to cover all three main levels of driving behavior. This approach was partly successful in that we managed to get significant BAC effects on SDLP and established relative validity for this

outcome in our scenario. Regrettably, most other outcomes, with a cautious exception for SD of speed, did not prove sufficiently sensitive and/or valid in our scenario.

We recruited a rather narrow sample of young, experienced male drivers who are not representative of the general driving population. On the other hand, they do represent a group that is involved in a disproportionately high percentage of traffic accidents, and drives more than the average population. Other studies have shown that the lateral vehicle control of poor/inexperienced drivers as well as female drivers are more affected by ethanol (Harrison and Fillmore 2005, Miller *et al.* 2009). This means that in this respect, the ethanol-induced SDLP increase measured in our study most likely underestimates the increases to be expected in the general driving population. We received much critique from the referees to our published papers regarding the sample, especially our exclusion of female participants. We recognize the major drawback that the exclusion of women makes the study less generalizable to half of the driving population. Our main reason for excluding females was the ethanol intervention, which would necessitate interviews and administration of pregnancy tests before each session, which we felt would be too intrusive. The ethical committee that approved our study accepted this argument. However, we later realized that most researchers in the field do not share this view. Another reason to exclude women (as well as non-Caucasians, very young or older drivers) was to recruit a homogenous sample to avoid excessive variability in the data that could lessen the sensitivity of the outcomes that we were trying to validate. We are aware that this approach at the same time rendered the results less generalizable to the driving population at large.

In our validation study, which is based on the comparison of simulated driving to 'real' driving, an underlying assumption is that test track driving is a close approximation to real driving. However, this assumption has not been proven. Even though the test track resembled common driving conditions on rural Norwegian roads, it is probably not possible to eliminate the feeling of an artificial situation when driving on a closed test track. The presence of a driving instructor for safety reasons may also have conveyed an artificial restraining effect, although they were instructed to be completely passive and not interfere with the driver unless necessary for safety reasons. This latter limitation to the realism of the comparator scenario would have been the same in an on-road test, and is thus no different from the

established ‘gold standard’ in DUI experimental research. This serves to underline that when undertaking experimental DUI studies, it is practically impossible to create a fully realistic scenario.

As discussed in chapter 5.1.1, there were some individual outliers in the simulator, with highly discrepant SDLP values compared to test track driving. Whereas most subjects seemed to keep to the instructions and tried as best they could to drive responsibly like they would normally do, some were obviously more influenced by ethanol, especially at the highest BAC level, and tended to depart from the instructions and act like rally drivers – without the skills. We suppose that such reactions stem from individual personality traits combined with individual variances in the response to ethanol. Obviously, these subjects’ results contributed greatly to the large variability in performance in the simulator. This is a threat to the sensitivity of simulator testing, because larger variability leads to less significant results. As some individuals may have a strong conditioned disinhibition response to ethanol ingestion, it is probably difficult to prevent the occurrence of such behaviors in certain subjects. In theory, one could exclude prospective study subjects whose behaviors are not ‘real-worldly’ in the simulator on the basis of a test drive – under ethanol – both in the simulator and on the test track. In practice, this would be too costly and cumbersome. One could also question whether this would be a sensible approach in a validation study, which is undertaken precisely to investigate the rapport between simulated and real driving. However, one function of a validation study may also be to recruit and screen individuals to a ‘bank’ of approved test subject for future studies. In future studies of driving impairment due to other drugs, it would be sensible to exclude subjects who previously showed very deviant behavior in the simulator.

Four participants were tested per day, and were allowed to mingle in a living room-like area next to the simulator room while they consumed their drinks and waited to drive. This setting may have contributed to disinhibited behavior in some subjects, since the ambience tended to become a bit ‘party-like’ on some nights. Isolating the subjects may have avoided this. A couple of the subjects, notably subject 15 identified in figure 5, paper I, reported to have a special interest for cars and driving. It is not unexpected that such individuals may be drawn to volunteer for a study like ours. This is another factor that could have contributed to the

rather hazardous driving of some of the subjects, which could possibly have been avoided by a more thorough screening process.

We chose to use a linear mixed model for the main statistical analyses in the papers. This methodology is well suited to a within-subject design with several, non-independent observations for each subject. This is a strength of this work compared to many other studies in the field, which have often used the more 'traditional' approach of analysis of variance (ANOVA), e.g. Kenntner-Mabiala *et al*. (2015). Using ANOVA may not be the best choice in such study designs, since it is vulnerable to non-normality of the dependent variables and is not very flexible in handling missing data. It also requires non-identical values of the independent variable (i.e., BAC) to be clustered into groups for analysis, which may be inaccurate and waste statistical power (McCulloch 2005, Seltman 2015).

Another strength of our validation study is that we measured actual BACs by drawing blood instead of estimating BACs from breath alcohol (BrAC) measurements. Although BAC and BrAC are closely correlated on the population level, the relationship may vary considerably between individuals (Jones 2011). Thus, measuring BrAC in study settings is not a very accurate measure of the actual, individual BAC. In studies using breath analysis to determine ethanol levels, actual BAC may vary considerably within the apparently homogenous BrAC groups. Some studies for instance apply repeated breath analyses to determine the appropriate time for testing, when the subject's BrAC has reached the desired level. The uncertainty in predicting BAC from BrAC on an individual level may in fact lead to a considerable variability within each intervention group with this approach. We instead measured actual BAC in blood samples drawn immediately before and after each driving session, and calculated the mean of the two measurements as the best estimate of the mean BAC during driving. The statistical method that we used considers each individual's actual BAC level instead of grouping different individuals with slightly varying BAC into one BAC group. This way, the statistical method maximizes the use of the information that is carried by every data point in the data set.

Due to technical problems, 10 of 60 drives on the test track, in 10 different subjects, did not yield valid SDLP readings. This occurred at random and did not systematically happen in, say, only one intervention group. With a 'traditional' statistical approach such as ANOVA, the

missing data would mean that the results from half of the subjects would have to be discarded. This is another strength of the mixed models approach, which allows the inclusion of subjects with some data missing, as long as the missing data occurred at random (Seltman 2015).

Throughout the four papers resulting from the validation study, a high number of inferences have been tested. Multiple inferences generally increase the chance of false-positive findings. The pre-defined primary outcome of our validation study was SDLP, so the concerns of multiple inferences do not extend to the SDLP results. Many of the inferences tested, for instance the possible effects of simulator sickness on the driving outcomes, are to be considered exploratory in nature, as there was no formal hypothesis testing. Indeed, for many of the inferences, hypotheses would have been inappropriate, as experimental data were simply lacking. Much of the results presented in the papers should be regarded as hypothesis generating; to be confirmed in future studies.

A weakness of our study design is that a clear hypothesis and analysis strategy for the main purpose of the study, namely validation and calibration of the simulator scenario, were imperfect at the outset. After the experimental phase of the study had been finished, we were somewhat overwhelmed by the enormous amount of data collected and the task of making sense out of it. Apparently, this is a rather common experience for driving simulator 'first-timers' (Caird and Horrey 2011). Optimal planning of a driving simulator study for DUI research, especially when it involves ethanol, requires specific experience, which we did not possess at the time. In light of this, despite its imperfections, the successful validation of the simulator scenario may not be that bad a result.

5.3 Ethical concerns

The main ethical dilemma of the validation study was the risk of exposing intoxicated subjects to real driving on a rather demanding test course, with the small but real possibility that an accident may be serious or even fatal. We took every step to ensure the safety of the participants, most importantly by the presence of experienced driving instructors with access to vehicle controls that could intervene if necessary. Both the ethical committee and the local police approved the study. Nevertheless, we all let out a sigh of relief when the last

participant had been safely returned home from the test course. We recorded no incidents or even potentially hazardous situations throughout the study.

Apart from this, the ethical committee was very concerned that the high BAC condition could render the participants intoxicated and possibly pose a danger to themselves or others. The committee first wanted us to book the participants into a hotel to let them sleep it off before we returned them to their loved ones. We opposed this, as we saw it as a counter-productive, inconvenient and very costly safety measure. The highest BAC attained during the study was 1.26 g/l. Such a BAC is not unheard of in social settings, and all the subjects had been social drinkers for many years. We thus did not see the BAC levels *per se* as problematic. The ethical committee finally agreed to this. The participants were returned directly to their homes by study personnel, and only after they had been assessed as fit to go home.

Researchers often point out the possibility of exploring hazardous driving scenarios, and even collision events, as one of the main advantages of driving simulators over on-road driving tests. However, it has been pointed out that although there may be no physical harm, the potential for psychological trauma with virtual collisions is largely unknown (Caird and Horrey 2011). Since we recorded no collisions throughout the study, this potential ethical problem did not affect us.

6 Conclusions

The findings in this thesis show that multiple driving simulator outcomes, representing several behavior levels of driving, are sensitive to ethanol effects. However, relative validity, as compared to real (closed test course) driving, was only confirmed for SDLP, which is a measure of lateral vehicle control at the automated behavior level. The results support the choice of SDLP as a primary outcome to study drug impaired driving in our simulator scenario. Since SDLP increase was significant already after 15 minutes' driving, the test duration could be shortened in future studies. This most likely would also reduce the severity of simulator sickness.

Simulator sickness seems to modify driving style, mediating decreased average speed and steering wheel reversal frequency, but higher BAC seems to ameliorate simulator sickness and cancel the effects on these driving measures. This may lead to false conclusions about BAC-related effects if simulator sickness is not taken into account. Simulator sickness did not significantly influence SDLP.

Our method of ethanol blinding, with the use of a placebo pill as an additional confounder, was effective in concealing the placebo condition but ineffective in concealing the ethanol conditions. We did not find any evidence of drug expectancy effects on SDLP.

The work in this thesis makes it possible to utilize the NTNU/SINTEF driving simulator in future studies of driving impairment due to CNS depressant drugs. Although validation is essentially specific to the simulator and task used, this work also contributes to the overall understanding of driving simulator research on drug impaired driving.

7 Future perspectives

The purpose of the validation study was to develop a valid tool to investigate drug effects on driving. There are many holes in our knowledge about drugs and driving. In the area of DUI legislation, there is a great need for research that connects concentrations of medicinal and recreational drugs in blood to quantitative measures of driving impairment and comparable BAC levels. Another area of research is to establish the effect of newly developed drugs on driving. Regulatory authorities such as the FDA increasingly demand studies to establish the traffic safety of new drugs or drug formulations (Farkas *et al.* 2013). There is also a new drug plague ravaging Europe called 'new psychoactive substances' (NPS), with an ever-changing plethora of novel drugs being ordered on the Internet and shipped directly to the customers – a phenomenon that started in the 2000s and has grown ever since. Many of these drugs only make fleeting appearances before they disappear again, whereas others seem to stick around. Very little is known about their influence on driving, which is a problem to DUI legislators and enforcers. Lastly, there is also a lack of knowledge regarding long-established medicinal drugs such as anti-epileptics (e.g. pregabalin) and opioids (e.g. tramadol).

As mentioned previously, several of our findings should be considered preliminary and hypothesis generating, and should preferably be confirmed in better-designed studies. This includes the findings that simulator sickness and drug expectancy do not seem to influence SDLP, as it is important to characterize the factors that may or may not influence this much-used outcome in experimental studies of driving impairment.

Many experts on DUI research agree that the final judgment about a drug's effect on driving should rest upon both epidemiological and experimental data. As for the latter, experimental data should preferably cover all the important levels or domains of driving behavior. However, it has proven difficult to establish valid and sensitive tests of impairment at the higher driving behavior levels. Unfortunately, our efforts did not meet this demand, and the quest continues. In this regard, the reports from the German group in Wurzburg of a composite endpoint that incorporates a range of driving behaviors and is sensitive to rather low BACs (Kenntner-Mabiala et al. 2015), and their plans to validate this approach against real driving, seem promising.

References

Allen RW, Rosenthal TJ, Cook ML. A short history of driving simulation. In: Fisher DL, ed. Handbook of driving simulation for engineering, medicine, and psychology. Vol 1. Boca Raton, Florida, U.S.: CRC Press; 2011.

Andreasson R, Jones AW. Erik M.P. Widmark (1889-1945): Swedish pioneer in forensic alcohol toxicology. *Forensic Sci Int.* 1995;72:1-14.

Apfel CC, Greim CA, Haubitz I, Goepfert C, Usadel J, Sefrin P, Roewer N. A risk score to predict the probability of postoperative vomiting in adults. *Acta Anaesthesiol Scand.* 1998;42:495-501.

Barker MJ, Greenwood KM, Jackson M, Crowe SF. Persistence of cognitive effects after withdrawal from long-term benzodiazepine use: a meta-analysis. *Arch Clin Neuropsychol.* 2004;19:437-54.

Berthelon C, Gineyt G. Effects of alcohol on automated and controlled driving performances. *Psychopharmacology (Berl).* 2014;231:2087-95.

Blaauw GJ. Driving Experience and Task Demands in Simulator and Instrumented Car - a Validation-Study. *Human Factors.* 1982;24:473-86.

Blana E, Golias J. Differences between vehicle lateral displacement on the road and in a fixed-base simulator. *Human Factors.* 2002;44:303-13.

Blomberg RD, Peck RC, Moskowitz H, Burns M, Fiorentino D. The Long Beach/Fort Lauderdale relative risk study. *J Safety Res.* 2009;40:285-92.

Bogstrand ST, Gjerde H, Normann PT, Rossow I, Ekeberg O. Alcohol, psychoactive substances and non-fatal road traffic accidents--a case-control study. *BMC Public Health.* 2012;12:734.

Borkenstein RF, Crowther RF, Shumante RP, Ziel WB, Zylman R. Die Rolle des alkoholisierten Fahrers bei Verkehrsunfällen ("The Grand Rapids study"). *Blutalkohol.* 1974;11:1-7.

Bosker WM, Kuypers KP, Theunissen EL, Surinx A, Blankespoor RJ, Skopp G, Jeffery WK, Walls HC, van Leeuwen CJ, Ramaekers JG. Medicinal Delta(9) -tetrahydrocannabinol (dronabinol) impairs on-the-road driving performance of occasional and heavy cannabis users but is not detected in Standard Field Sobriety Tests. *Addiction.* 2012;107:1837-44.

Bramness JG, Skurtveit S, Mørland J. Testing for benzodiazepine inebriation--relationship between benzodiazepine concentration and simple clinical tests for impairment in a sample of drugged drivers. *Eur J Clin Pharmacol.* 2003;59:593-601.

Bridgeman B, Blaesi S, Campusano R. Optical correction reduces simulator sickness in a driving environment. *Hum Factors.* 2014;56:1472-81.

Brookhuis K, Dewaard D, Mulder B. Measuring Driving Performance by Car-Following in Traffic. *Ergonomics.* 1994;37:427-34.

Brooks JO, Goodenough RR, Crisler MC, Klein ND, Alley RL, Koon BL, Logan WC, Jr., Ogle JH, Tyrrell RA, Wills RF. Simulator sickness during driving simulation studies. *Accid Anal Prev.* 2010;42:788-96.

Burian SE, Hensberry R, Liguori A. Differential effects of alcohol and alcohol expectancy on risk-taking during simulated driving. *Hum Psychopharmacol.* 2003;18:175-84.

Caird JK, Horrey WJ. Twelve practical and useful questions about driving simulation. In: Fisher DL, ed. Handbook of driving simulation for engineering, medicine, and psychology. Vol 1. Boca Raton, Florida, U.S.: CRC Press; 2011.

Charlton SG, Starkey NJ. Driving while drinking: performance impairments resulting from social drinking. *Accid Anal Prev.* 2015;74:210-7.

Christophersen AS, Mørland J, Stewart K, Gjerde H. International trends in alcohol and drug use among vehicle drivers. *Forensic Sci Rev.* 2016;28:37-66.

Classen S, Bewernitz M, Shechtman O. Driving simulator sickness: an evidence-based review of the literature. *Am J Occup Ther.* 2011;65:179-88.

Classen S, Brooks J. Driving simulators for occupational therapy screening, assessment, and intervention. *Occup Ther Health Care.* 2014;28:154-62.

Cobb SVG, Nichols S, Ramsey A, Wilson JR. Virtual reality-induced symptoms and effects (VRISE). *Presence-Teleoperators and Virtual Environments.* 1999;8:169-86.

Conrad M, McNamara P, King A. Alternative substance paradigm: effectiveness of beverage blinding and effects on acute alcohol responses. *Exp Clin Psychopharmacol.* 2012;20:382-9.

Contardi S, Pizza F, Sancisi E, Mondini S, Cirignotta F. Reliability of a driving simulation task for evaluation of sleepiness. *Brain Res Bull.* 2004;63:427-31.

Daurat A, Sagaspe P, Motak L, Taillard J, Bayssac L, Huet N, Authie C, Mestre D, Philip P. Lorazepam impairs highway driving performance more than heavy alcohol consumption. *Accid Anal Prev.* 2013;60:31-4.

Domeyer JE, Cassavaugh ND, Backs RW. The use of adaptation to reduce simulator sickness in driving assessment and research. *Accid Anal Prev.* 2013;53:127-32.

Drummer OH, Gerostamoulos J, Batziris H, Chu M, Caplehorn J, Robertson MD, Swann P. The involvement of drugs in drivers of motor vehicles killed in Australian road traffic crashes. *Accid Anal Prev.* 2004;36:239-48.

Engen T. Use and validation of driving simulators. Thesis. Department of Civil and Transport Engineering, Norwegian University of Science and Technology; 2008.

Engström J, Johansson E, Östlund J. Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F.* 2005;8:97-120.

Epps J, Monk C, Savage S, Marlatt GA. Improving credibility of instructions in the balanced placebo design: a misattribution manipulation. *Addict Behav.* 1998;23:427-35.

Farkas RH, Unger EF, Temple R. Zolpidem and driving impairment--identifying persons at risk. *N Engl J Med.* 2013;369:689-91.

Fillmore MT, Vogel-Sprott M. Expectancies about alcohol-induced motor impairment predict individual differences in responses to alcohol and placebo. *J Stud Alcohol.* 1995;56:90-8.

Fillmore MT, Blackburn JS, Harrison EL. Acute disinhibiting effects of alcohol as a factor in risky driving behavior. *Drug Alcohol Depend.* 2008;95:97-106.

Finnigan F, Hammersley R, Millar K. The effects of expectancy and alcohol on cognitive-motor performance. *Addiction.* 1995;90:661-72.

Freydier C, Berthelon C, Bastien-Toniazzo M, Gineyt G. Divided attention in young drivers under the influence of alcohol. *J Safety Res.* 2014;49:13-8.

Galvez-Garcia G, Hay M, Gabaude C. Alleviating simulator sickness with galvanic cutaneous stimulation. *Hum Factors.* 2015;57:649-57.

Gawron VJ, Ranney TA. The effects of alcohol dosing on driving performance on a closed course and in a driving simulator. *Ergonomics.* 1988;31:1219-44.

Gjerde H, Sakshaug J, Mørland J. Heavy drinking among Norwegian male drunken drivers: a study of gamma-glutamyltransferase. *Alcohol Clin Exp Res.* 1986;10:209-12.

Gjerde H, Normann PT, Pettersen BS, Assum T, Aldrin M, Johansen U, Kristoffersen L, Øiestad EL, Christophersen AS, Mørland J. Prevalence of alcohol and drugs among Norwegian motor vehicle drivers: a roadside survey. *Accid Anal Prev.* 2008;40:1765-72.

Gjerde H, Normann PT, Christophersen AS, Mørland J. Prevalence of driving with blood drug concentrations above proposed new legal limits in Norway: estimations based on drug concentrations in oral fluid. *Forensic Sci Int.* 2011;210:221-7.

Gjerde H, Normann PT, Christophersen AS, Samuelsen SO, Mørland J. Alcohol, psychoactive drugs and fatal road traffic accidents in Norway: a case-control study. *Accid Anal Prev.* 2011;43:1197-203.

Gjerde H, Christophersen AS, Normann PT, Assum T, Øiestad EL, Mørland J. Norwegian roadside survey of alcohol and drug use by drivers (2008-2009). *Traffic Inj Prev.* 2013;14:443-52.

Gjerde H, Strand MC, Morland J. Driving Under the Influence of Non-Alcohol Drugs--An Update Part I: Epidemiological Studies. *Forensic Sci Rev.* 2015;27:89-113.

Gobbo MA, Louza MR. Influence of stimulant and non-stimulant drug treatment on driving performance in patients with attention deficit hyperactivity disorder: a systematic review. *Eur Neuropsychopharmacol.* 2014;24:1425-43.

Godley ST, Triggs TJ, Fildes BN. Driving simulator validation for speed research. *Accid Anal Prev.* 2002;34:589-600.

Gustavsen I, Bramness JG, Skurtveit S, Engeland A, Neutel I, Mørland J. Road traffic accident risk related to prescriptions of the hypnotics zopiclone, zolpidem, flunitrazepam and nitrazepam. *Sleep Med.* 2008;9:818-22.

Gustavsen I, Al-Sammurraie M, Mørland J, Bramness JG. Impairment related to blood drug concentrations of zopiclone and zolpidem compared to alcohol in apprehended drivers. *Accid Anal Prev.* 2009;41:462-6.

Gustavsen I, Hjelmeland K, Bernard JP, Mørland J. Psychomotor performance after intake of zopiclone compared with intake of ethanol: a randomized, controlled, double-blinded trial. *J Clin Psychopharmacol.* 2011;31:481-8.

Harrison EL, Fillmore MT. Are bad drivers more impaired by alcohol? Sober driving precision predicts impairment from alcohol in a simulated driving task. *Accid Anal Prev.* 2005;37:882-9.

Hartman RL, Brown TL, Milavetz G, Spurgin A, Pierce RS, Gorelick DA, Gaffney G, Huestis MA. Cannabis effects on driving lateral control with and without alcohol. *Drug Alcohol Depend.* 2015;154:25-37.

Hartman RL, Brown TL, Milavetz G, Spurgin A, Pierce RS, Gorelick DA, Gaffney G, Huestis MA. Cannabis effects on driving longitudinal control with and without alcohol. *J Appl Toxicol.* 2016.

Helland A, Jenssen GD, Lervåg LE, Moen T, Sakshaug K, Lydersen S, Mørland J, Slørdal L. Evaluation of measures of impairment in real and simulated driving: Results from a randomized, placebo-controlled study. *Traffic Inj Prev.* 2016;17:245-50.

Hjälmdahl M, Vadeby A, Forsman Å, Fors C, Ceder G, Woxler P, Kronstrand R. Effects of *d*-amphetamine on simulated driving performance before and after sleep deprivation. *Psychopharmacology (Berl).* 2012;222:401-11.

Holmgren A, Holmgren P, Kugelberg FC, Jones AW, Ahlner J. Predominance of illicit drugs and poly-drug use among drug-impaired drivers in Sweden. *Traffic Inj Prev.* 2007;8:361-7.

Houwing S, Mathijssen R, Brookhuis KA. Case-control studies. In: Verster JC, Pandi-Perumal SR, Ramaekers JG, de Gier JJ, eds. *Drugs, driving and traffic safety*. Basel, Switzerland: Birkhäuser Verlag AG; 2009:107-20.

Huemer AK, Vollrath M. Alcohol-related impairment in the Lane Change Task. *Accid Anal Prev.* 2010;42:1983-8.

Irving A, Jones W. Methods for testing impairment of driving due to drugs. *Eur J Clin Pharmacol.* 1992;43:61-6.

Jackson A, Stephens DN, Duka T. A low dose alcohol drug discrimination in social drinkers: relationship with subjective effects. *Psychopharmacology (Berl).* 2001;157:411-20.

Jamt REG, Nilsson G, Bogstrand ST, Eliassen E, Normann PT, Gjerde H. Roadside survey on alcohol and drug use among drivers in the arctic county of Finnmark,Norway. 53rd meeting of The international association of forensic toxicologists; August 30 - September 4, 2015; Florence, Italy.

Jones AW. Driving under the influence of drugs in Sweden with zero concentration limits in blood for controlled substances. *Traffic Inj Prev.* 2005;6:317-22.

Jones AW, Holmgren A, Kugelberg FC. Concentrations of scheduled prescription drugs in blood of impaired drivers: considerations for interpreting the results. *Ther Drug Monit.* 2007;29:248-60.

Jones AW. Pharmacokinetics of Ethanol - Issues of Forensic Importance. *Forensic Sci Rev.* 2011;23:91-136.

Jongen S, Vuurman E, Ramaekers J, Vermeeren A. Alcohol calibration of tests measuring skills related to car driving. *Psychopharmacology (Berl).* 2014;231:2435-47.

Jongen S, Vuurman EF, Ramaekers JG, Vermeeren A. The sensitivity of laboratory tests assessing driving related skills to dose-related impairment of alcohol: A literature review. *Accid Anal Prev.* 2016;89:31-48.

Kaptein NA, Theeuwes J, van der Horst R. Driving simulator validity: Some considerations. *Transportation Research Record.* 1996;1550:30-6.

Kaussner Y, Kenntner-Mabiala R, Hoffmann S, Klatt J, Tracik F, Kruger HP. Effects of oxcarbazepine and carbamazepine on driving ability: a double-blind, randomized crossover trial with healthy volunteers. *Psychopharmacology (Berl).* 2010;210:53-63.

Kawano N, Iwamoto K, Ebe K, Aleksic B, Noda A, Umegaki H, Kuzuya M, Iidaka T, Ozaki N. Slower adaptation to driving simulator and simulator sickness in older adults. *Aging Clin Exp Res.* 2012;24:285-9.

Kay GG, Logan BK. Drugged Driving Expert Panel report: A consensus protocol for assessing the potential of drugs to impair driving. National Highway Traffic Safety Administration; 2011.

Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire (SSQ): a new method for quantifying simulator sickness. *Int J Aviat Psychol.* 1993;3:203-20.

Kennedy RS, Stanney KM, Dunlap WP. Duration and exposure to virtual environments: Sickness curves during and across sessions. *Presence-Teleoperators and Virtual Environments.* 2000;9:463-72.

Kenntner-Mabiala R, Kaussner Y, Jagiellowicz-Kaufmann M, Hoffmann S, Kruger HP. Driving performance under alcohol in simulated representative driving tasks: an alcohol calibration study for impairments related to medicinal drugs. *J Clin Psychopharmacol.* 2015;35:134-42.

Krüger HP, Vollrath M. The alcohol-related accident risk in Germany: procedure, methods and results. *Accid Anal Prev.* 2004;36:125-33.

Lenne MG, Dietze P, Rumbold GR, Redman JR, Triggs TJ. The effects of the opioid pharmacotherapies methadone, LAAM and buprenorphine, alone and in combination with alcohol, on simulated driving. *Drug Alcohol Depend.* 2003;72:271-8.

Lenne MG, Dietze PM, Triggs TJ, Walmsley S, Murphy B, Redman JR. The effects of cannabis and alcohol on simulated arterial driving: Influences of driving experience and task demand. *Accid Anal Prev.* 2010;42:859-66.

Leufkens TR, Vermeeren A, Smink BE, van Ruitenbeek P, Ramaekers JG. Cognitive, psychomotor and actual driving performance in healthy volunteers after immediate and extended release formulations of alprazolam 1 mg. *Psychopharmacology (Berl).* 2007;191:951-9.

Lillsunde P, Gunnar T. Drugs and driving: the Finnish perspective. *Bull Narc.* 2005;57:213-29.

Louwerens JW, Gloerich ABM, de Vries G, Brookhuis KA, O'Hanlon JF. The relationship between drivers' blood alcohol concentration (BAC) and actual driving performance during high speed travel. In: Noordzij PC, Rozbach R, eds. *Alcohol, drugs and traffic safety.* Amsterdam: Excerpta Medica; 1987:183-92.

Manthey L, van Loenen-Frosch F, Giltay EJ, van Veen T, Glashouwer K, Penninx BW, Zitman FG. High dose benzodiazepines prolong reaction times in chronic users who have major depressive and/or anxiety disorders. *Br J Clin Pharmacol.* 2014;77:571-7.

Marczinski CA, Fillmore MT. Compensating for alcohol-induced impairment of control: effects on inhibition and activation of behavior. *Psychopharmacology (Berl).* 2005;181:337-46.

Martin CS, Moss HB. Measurement of acute tolerance to alcohol in human subjects. *Alcohol Clin Exp Res.* 1993;17:211-6.

Martin CS, Sayette MA. Experimental design in alcohol administration research: limitations and alternatives in the manipulation of dosage-set. *J Stud Alcohol.* 1993;54:750-61.

McCulloch CE. Repeated measures ANOVA, R.I.P.? *Chance.* 2005;18:29-33.

McMillen DL, Smith SM, Wells-Parker E. The effects of alcohol, expectancy, and sensation seeking on driving risk taking. *Addict Behav.* 1989;14:477-83.

Mets M, Baas D, van Boven I, Olivier B, Verster J. Effects of coffee on driving performance during prolonged simulated highway driving. *Psychopharmacology (Berl).* 2012;222:337-42.

Mets MA, Kuipers E, de Senerpont Domis LM, Leenders M, Olivier B, Verster JC. Effects of alcohol on highway driving in the STISIM driving simulator. *Hum Psychopharmacol.* 2011;26:434-9.

Michon JA. A critical view of driver behavior models: what do we know, what should we do? In: Evans L, Schwing RC, eds. *Human behavior and traffic safety*. New York: Plenum Press; 1985:485-520.

Miles GH, Vincent DF. The Institute's tests for motor drivers. *The Human Factor.* 1934;VIII:245-57.

Miller MA, Weafer J, Fillmore MT. Gender differences in alcohol impairment of simulated driving performance and driving-related skills. *Alcohol Alcohol.* 2009;44:586-93.

Milleville-Pennel I, Charron C. Do mental workload and presence experienced when driving a real car predispose drivers to simulator sickness? An exploratory study. *Accid Anal Prev.* 2015;74:192-202.

Mitchell MC. Alcohol-induced impairment of central nervous system function: behavioral skills involved in driving. *J Stud Alcohol Suppl.* 1985;10:109-16.

Moe D. Evaluering av bilsimulator Autosim A/S – del1: Alphatest. SINTEF Samferdselsteknikk; 1995.

Moss AC, Albery IP. A dual-process model of the alcohol-behavior link for social drinking. *Psychol Bull.* 2009;135:516-30.

Mullen N, Charlton J, Devlin A, Bédard M. Simulator validity: behaviors observed on the simulator and on the road. In: Fisher DL, ed. *Handbook of driving simulation for engineering, medicine, and psychology.* Vol 1. Boca Raton: CRC Press; 2011.

Muller D, Judd CM, Yzerbyt VY. When moderation is mediated and mediation is moderated. *J Pers Soc Psychol.* 2005;89:852-63.

Musshoff F, Hokamp EG, Bott U, Madea B. Performance evaluation of on-site oral fluid drug screening devices in normal police procedure in Germany. *Forensic Sci Int.* 2014;238:120-4.

Muttray A, Breitinger A, Goetze E, Schnupp T, Geissler B, Kaufmann T, Golz M, Letzel S. Further development of a commercial driving simulation for research in occupational medicine. *Int J Occup Med Environ Health.* 2013;26:949-65.

Möller JC, Stiasny K, Hargutt V, Cassel W, Tietze H, Peter JH, Krüger HP, Oertel WH. Evaluation of sleep and driving performance in six patients with Parkinson's disease reporting sudden onset of sleep under dopaminergic medication: a pilot study. *Mov Disord.* 2002;17:474-81.

Mørland J, Setekleiv J, Haffner JF, Strømsæther CE, Danielsen A, Wethe GH. Combined effects of diazepam and ethanol on mental and psychomotor functions. *Acta Pharmacol Toxicol (Copenh).* 1974;34:5-15.

Mørland J. Driving under the Influence of Non-Alcohol Drugs. *Forensic Sci Rev.* 2000;12:79-105.

Narahashi T, Kuriyama K, Illes P, Wirkner K, Fischer W, Muhlberg K, Scheibler P, Allgaier C, Minami K, Lovinger D, Lallemand F, Ward RJ, DeWitte P, Itatsu T, Takei Y, Oide H, Hirose M, Wang XE, Watanabe

S, Tateyama M, Ochi R, Sato N. Neuroreceptors and ion channels as targets of alcohol. *Alcohol Clin Exp Res.* 2001;25:182S-8S.

Neutel CI. Risk of traffic accident injury after a prescription for a benzodiazepine. *Ann Epidemiol.* 1995;5:239-44.

Nilsen HK, Landrø NI, Kaasa S, Jenssen GD, Fayers P, Borchgrevink PC. Driving functions in a video simulator in chronic non-malignant pain patients using and not using codeine. *Eur J Pain.* 2011;15:409-15.

O'Hanlon JF, Haak TW, Blaauw GJ, Riemersma JB. Diazepam impairs lateral position control in highway driving. *Science.* 1982;217:79-81.

O'Hanlon JF. Driving performance under the influence of drugs: rationale for, and application of, a new test. *Br J Clin Pharmacol.* 1984;18 Suppl 1:121S-9S.

O'Hanlon JF, Ramaekers JG. Antihistamine effects on actual driving performance in a standard test: a summary of Dutch experience, 1989-94. *Allergy.* 1995;50:234-42.

Ogden EJ, Moskowitz H. Effects of alcohol and other drugs on driver performance. *Traffic Inj Prev.* 2004;5:185-98.

Owens K, Ramaekers JG. Drugs, driving, and models to measure driving impairment. In: Verster JC, Pandi-Perumal SR, Ramaekers JG, de Gier JJ, eds. Drugs, driving and traffic safety. Basel, Switzerland: Birkhäuser Verlag AG; 2009:43-58.

Peck RC, Gebers MA, Voas RB, Romano E. The relationship between blood alcohol concentration (BAC), age, and crash risk. *J Safety Res.* 2008;39:311-9.

Ramaekers JG. Antidepressants and driver impairment: empirical evidence from a standard on-the-road test. *J Clin Psychiatry.* 2003;64:20-9.

Ramaekers JG, Berghaus G, van Laar M, Drummer OH. Dose related risk of motor vehicle crashes after cannabis use. *Drug Alcohol Depend.* 2004;73:109-19.

Ramaekers JG, Kuypers KP, Bosker WM, Brookhuis KA, Veldstra JA, Simons R, Martens M, Hjalmdahl M, Forsman A, Knoche A. Effects of stimulant drugs on actual and simulated driving: perspectives from four experimental studies conducted as part of the DRUID research consortium. *Psychopharmacology (Berl).* 2012;222:413-8.

Reed MP, Green PA. Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialling task. *Ergonomics.* 1999;42:1015-37.

Rimm DC, Sininger RA, Faherty JD, Whitley MD, Perl MB. A balanced placebo investigation of the effects of alcohol vs. alcohol expectancy on simulated driving behavior. *Addict Behav.* 1982;7:27-32.

Rohsenow DJ, Marlatt GA. The balanced placebo design: methodological considerations. *Addict Behav.* 1981;6:107-22.

Rosner B. Fundamentals of biostatistics, 6th ed. Belmont, CA: Thomson-Brooks/Cole; 2006.

Sakshaug K. VALIDAD pilot experiments in simulator and on test track. Results of data analysis. SINTEF Technology and Society internal report project 503460; 2008.

Sayette MA, Breslin FC, Wilson GT, Rosenblum GD. An evaluation of the balanced placebo design in alcohol administration research. *Addict Behav.* 1994;19:333-42.

Seltman HJ. Experimental design and analysis. http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf. Accessed 22.05.2016.

Seppala K, Korttila K, Hakkinen S, Linnoila M. Residual effects and skills related to driving after a single oral administration of diazepam, medazepam or lorazepam. *Br J Clin Pharmacol.* 1976;3:831-41.

Seppala T, Linnoila M, Mattila MJ. Drugs, alcohol and driving. *Drugs.* 1979;17:389-408.

Shechtman O, Classen S, Awadzi K, Mann W. Comparison of driving errors between on-the-road and simulated driving assessment: a validation study. *Traffic Inj Prev.* 2009;10:379-85.

Silber BY, Croft RJ, Downey LA, Papafotiou K, Camfield DA, Stough C. The effect of *d*-methamphetamine on simulated driving performance. *Hum Psychopharmacol.* 2012;27:139-44.

Simen AA, Gargano C, Cha JH, Drexel M, Bautmans A, Heirman I, Laethem T, Hochadel T, Gheyle L, Bleys K, Beals C, Stoch A, Kay GG, Struyk A. A randomized, crossover, placebo-controlled clinical trial to assess the sensitivity of the CRCDS Mini-Sim to the next-day residual effects of zopiclone. *Ther Adv Drug Saf.* 2015;6:86-97.

Simons R, Martens M, Ramaekers J, Krul A, Klopping-Ketelaars I, Skopp G. Effects of dexamphetamine with and without alcohol on simulated driving. *Psychopharmacology (Berl).* 2012;222:391-9.

Statens institutt for rusmiddelforskning (SIRUS). Rusmidler i Norge 2015.

Skegg DC, Richards SM, Doll R. Minor tranquillisers and road accidents. *Br Med J.* 1979;1:917-9.

Smiley A, Moskowitz H. Effects of long-term administration of buspirone and diazepam on driver steering control. *Am J Med.* 1986;80:22-9.

Spanagel R. Alcoholism: a systems approach from molecular physiology to addictive behavior. *Physiol Rev.* 2009;89:649-705.

Staner L, Ertle S, Boeijinga P, Rinaudo G, Arnal MA, Muzet A, Luthringer R. Next-day residual effects of hypnotics in DSM-IV primary insomnia: a driving simulator study with simultaneous electroencephalogram monitoring. *Psychopharmacology (Berl).* 2005;181:790-8.

Starkey NJ, Charlton SG. The effects of moderate alcohol concentrations on driving and cognitive performance during ascending and descending blood alcohol concentrations. *Hum Psychopharmacol.* 2014;29:370-83.

Steentoft A, Simonsen KW, Linnet K. The frequency of drugs among Danish drivers before and after the introduction of fixed concentration limits. *Traffic Inj Prev.* 2010;11:329-33.

Strand MC, Gjerde H, Mørland J. Driving under the influence of non-alcohol drugs - an update. Part II: Experimental studies *Forensic Sci Rev.* 2016;28:79-102.

Testa M, Fillmore MT, Norris J, Abbey A, Curtin JJ, Leonard KE, Mariano KA, Thomas MC, Nomensen KJ, George WH, Vanzile-Tamsen C, Livingston JA, Saenz C, Buck PO, Zawacki T, Parkhill MR, Jacques AJ, Hayman LW, Jr. Understanding alcohol expectancy effects: revisiting the placebo condition. *Alcohol Clin Exp Res.* 2006;30:339-48.

Thompson I, Williams G, Caldwell B, Aldington S, Dickson S, Lucas N, McDowall J, Weatherall M, Robinson G, Beasley R. Randomised double-blind, placebo-controlled trial of the effects of the 'party pills' BZP/TFMPP alone and in combination with alcohol. *J Psychopharmacol.* 2010;24:1299-308.

Törnros J. Driving behaviour in a real and a simulated road tunnel - A validation study. *Accident Analysis and Prevention.* 1998;30:497-503.

Uc EY, Rizzo M, Anderson SW, Dastrup E, Sparks JD, Dawson JD. Driving under low-contrast visibility conditions in Parkinson disease. *Neurology.* 2009;73:1103-10.

Vakulin A, Baulk SD, Catcheside PG, Antic NA, van den Heuvel CJ, Dorrian J, McEvoy RD. Effects of alcohol and sleep restriction on simulated driving performance in untreated patients with obstructive sleep apnea. *Ann Intern Med.* 2009;151:447-55.

Veldstra JL, Brookhuis KA, de Waard D, Molmans BH, Verstraete AG, Skopp G, Jantos R. Effects of alcohol (BAC 0.5 per thousand) and ecstasy (MDMA 100 mg) on simulated driving performance and traffic safety. *Psychopharmacology (Berl).* 2012;222:377-90.

Veldstra JL. When the party is over. Investigating the effects of Alcohol, THC and MDMA on simulator driving performance. Thesis. University of Groningen, the Netherlands; 2014.

Veldstra JL, Bosker WM, de Waard D, Ramaekers JG, Brookhuis KA. Comparing treatment effects of oral THC on simulated and on-the-road driving performance: testing the validity of driving simulator drug research. *Psychopharmacology (Berl).* 2015;232:2911-9.

Vengeliene V, Bilbao A, Molander A, Spanagel R. Neuropharmacology of alcohol addiction. *Br J Pharmacol.* 2008;154:299-315.

Verster JC, Veldhuijzen DS, Volkerts ER. Residual effects of sleep medication on driving ability. *Sleep Med Rev.* 2004;8:309-25.

Verster JC, Veldhuijzen DS, Patat A, Olivier B, Volkerts ER. Hypnotics and driving safety: meta-analyses of randomized controlled trials applying the on-the-road driving test. *Curr Drug Saf.* 2006;1:63-71.

Verster JC, Roth T. Standard operation procedures for conducting the on-the-road driving test, and measurement of the standard deviation of lateral position (SDLP). *Int J Gen Med.* 2011;4:359-71.

Verster JC, Bervoets AC, de Klerk S, Vreman RA, Olivier B, Roth T, Brookhuis KA. Effects of alcohol hangover on simulated highway driving performance. *Psychopharmacology (Berl).* 2014;231:2999-3008.

Verster JC, Roth T. Effects of central nervous system drugs on driving: speed variability versus standard deviation of lateral position as outcome measure of the on-the-road driving test. *Hum Psychopharmacol.* 2014;29:19-24.

Verster JC, Roth T. Excursions out-of-lane versus standard deviation of lateral position as outcome measure of the on-the-road driving test. *Hum Psychopharmacol.* 2014;29:322-9.

Verster JC, van de Loo AJ, Roth T. Mirtazapine as positive control drug in studies examining the effects of antidepressants on driving ability. *Eur J Pharmacol.* 2015;753:252-6.

Vindenes V, Jordbru D, Knapskog AB, Kvan E, Mathisrud G, Slørdal L, Mørland J. Impairment based legislative limits for driving under the influence of non-alcohol drugs in Norway. *Forensic Sci Int.* 2012;219:1-11.

Voas RB. Robert F. Borkenstein: an appreciation. *Addiction.* 2003;98:371.

Voas RB, Torres P, Romano E, Lacey JH. Alcohol-related risk of driver fatalities: an update using 2007 data. *J Stud Alcohol Drugs.* 2012;73:341-50.

Volkow ND, Koob GF, McLellan AT. Neurobiologic Advances from the Brain Disease Model of Addiction. *N Engl J Med.* 2016;374:363-71.

Walsh JM, de Gier JJ, Christopherson AS, Verstraete AG. Drugs and driving. *Traffic Inj Prev.* 2004;5:241-53.

Walsh JM, Verstraete AG, Huestis MA, Mørland J. Guidelines for research on drugged driving. *Addiction.* 2008;103:1258-68.

Weafer J, Camarillo D, Fillmore MT, Milich R, Marczinski CA. Simulated driving performance of adults with ADHD: comparisons with alcohol intoxication. *Exp Clin Psychopharmacol.* 2008;16:251-63.

Weafer J, Fillmore MT. Acute tolerance to alcohol impairment of behavioral and cognitive mechanisms related to driving: drinking and driving on the descending limb. *Psychopharmacology (Berl).* 2012;220:697-706.

Weiler JM, Bloomfield JR, Woodworth GG, Grant AR, Layton TA, Brown TL, McKenzie DR, Baker TW, Watson GS. Effects of fexofenadine, diphenhydramine, and alcohol on driving performance. A randomized, placebo-controlled trial in the Iowa driving simulator. *Ann Intern Med.* 2000;132:354-63.

World Health Organization. Global Status Report on Road Safety 2013: Supporting a Decade of Action.

Zhang X, Zhao X, Du H, Ma J, Rong J. Effect of different breath alcohol concentrations on driving performance in horizontal curves. *Accid Anal Prev.* 2014;72:401-10.

Errata

In the methods chapters of papers I and II, it is stated that the test subjects "as a group drove slightly less" than the general population. The correct statement is that the test subjects drove slightly *more* than the general driving population. This has been corrected in papers III and IV, as well as in this thesis (see table 3–2, chapter 3.2.2). Since the difference from the general population of drivers at any rate was quite small, this should not have important consequences for the interpretation of the data.

In the discussion chapter of paper I, it is stated, "There are few simulator studies using SDLP as outcome measure". As evident from the review article of Jongen *et al*. (2016), this is not accurate, and neither was it accurate at the time of writing the first paper. In fact, Jongen *et al*. found 37 simulator studies that had reported SDLP as an outcome, of which most were published before 2012. It would have been more correct to state that few simulator studies have been published that have validated the use of SDLP as an outcome measure for DUI research.

In paper II, the influence of BAC on the occurence of driving through red light was reportedly analyzed with a paired samples *t*-test. However, the number of drives through a red light per driving session could be either zero, one or two, and is obviously not a normally distributed variable; hence, a *t*-test would not be correct. We therefore attempted to repeat the analysis: There were 32 passings of red light at each BAC level in the simulator. The number of red light violations were one, two, and seven, respectively, at BAC 0, 0.5 and 0.9. The probability of red light violation increased significantly with increasing BAC level (exact *p*-value = 0.022, two-sided linear-by-linear test assuming independent observations). We have made two simplifications by assuming independence. First, for each BAC level and person, there are usually two observations. Assuming independence in this context tends to bias the *p*-value downwards. Second, for most persons there are observations within person at three BAC levels. Assuming independence in this context tends to bias the *p*-value upwards. Hence, we have reason to believe the *p*-value obtained assuming independence to be of the right size of order.

In paper III, simulator sickness is described as an operational confounder, i.e. a variable that influences the outcome measures (steering wheel reversals and average speed) along with the independent variable of interest (BAC). However, it seems more correct to interpret simulator sickness as a *mediator* of an apparent BAC effect on these outcomes (see figure 5-2 as well as the discussion in chapter 5.1.3).

In the footnote *e* of table 1 in Paper IV, “Perceived sedative in drink” should read “Perceived sedative in pill”.

Paper I

Accident Analysis and Prevention 53 (2013) 9–16

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Comparison of driving simulator performance with real driving after alcohol intake: A randomised, single blind, placebo-controlled, cross-over trial

Arne Helland[a,b,*], Gunnar D. Jenssen[c], Lone-Eirin Lervåg[c], Andreas Austgulen Westin[a], Terje Moen[c], Kristian Sakshaug[c], Stian Lydersen[d], Jørg Mørland[e], Lars Slørdal[a,b]

[a] Department of Clinical Pharmacology, St. Olav University Hospital, Trondheim, Norway
[b] Department of Laboratory Medicine, Children's and Women's Health, Norwegian University of Science and Technology, Trondheim, Norway
[c] SINTEF Technology and Society, Department of Transport Research, Trondheim, Norway
[d] The Regional Centre for Child and Adolescent Mental Health, Norwegian University of Science and Technology, Trondheim, Norway
[e] National Institute of Public Health, Oslo, Norway and Institute of Clinical Medicine, University of Oslo, Norway

ARTICLE INFO

Article history:
Received 19 September 2012
Received in revised form
20 December 2012
Accepted 28 December 2012

Keywords:
Automobile driving
Traffic accidents
Ethanol

ABSTRACT

The purpose of this study was to establish and validate a driving simulator method for assessing drug effects on driving. To achieve this, we used ethanol as a positive control, and examined whether ethanol affects driving performance in the simulator, and whether these effects are consistent with performance during real driving on a test track, also under the influence of ethanol. Twenty healthy male volunteers underwent a total of six driving trials of 1 h duration; three in an instrumented vehicle on a closed-circuit test track that closely resembled rural Norwegian road conditions, and three in the simulator with a driving scenario modelled after the test track. Test subjects were either sober or titrated to blood alcohol concentration (BAC) levels of 0.5 g/L and 0.9 g/L. The study was conducted in a randomised, cross-over, single-blind fashion, using placebo drinks and placebo pills as confounders. The primary outcome measure was standard deviation of lateral position (SDLP; "weaving"). Eighteen test subjects completed all six driving trials, and complete data were acquired from 18 subjects in the simulator and 10 subjects on the test track, respectively. There was a positive dose–response relationship between higher ethanol concentrations and increases in SDLP in both the simulator and on the test track ($p < 0.001$ for both). In the simulator, this dose–response was evident already after 15 min of driving. SDLP values were higher and showed a larger inter-individual variability in the simulator than on the test track. Most subjects displayed a similar relationship between BAC and SDLP in the simulator and on the test track; however, a few subjects showed striking dissimilarities, with very high SDLP values in the simulator. This may reflect the lack of perceived danger in the simulator, causing reckless driving in a few test subjects. Overall, the results suggest that SDLP in the driving simulator is a sensitive measure of ethanol impaired driving. The comparison with real driving implies relative external validity of the simulator.

1. Introduction

Impaired driving caused by ethanol and/or drugs is a major cause of traffic accidents, and thus a major public health problem (Blomberg et al., 2009). The relationship between blood ethanol concentrations (BAC) and accident risk is well established in large epidemiological studies (Borkenstein et al., 1974; Blomberg et al., 2009). With the exception of cannabis (Ramaekers et al., 2004), similar relationships have not been demonstrated for other psychoactive drugs and drugs of abuse. Case–control studies on non-alcohol drugs require screening and quantification of a large number of potentially impairing drugs, as well as a large number of cases, as each drug has a relatively low prevalence of detection in car crash drivers. Such studies have seldom been performed, leaving the relation between blood drug concentrations and crash risk largely unknown. Also, blood sampling for drug testing of controls – as compared to simple breath tests in ethanol studies – is necessary, and makes the recruitment of controls more difficult (Verster et al., 2009a). Furthermore, post-mortem drug concentration changes occur to a larger degree in non-alcohol drugs, making interpretation of toxicological data from studies of killed drivers difficult.

Epidemiological approaches cannot establish causal relationships, and are fraught with methodological difficulties, including

* Corresponding author at: Department of Clinical Pharmacology, St. Olav University Hospital, Pb 3250 Sluppen, NO-7006 Trondheim, Norway. Tel.: +47 72 82 91 00; fax: +47 72 82 91 10.
E-mail addresses: arne.helland@legemidler.no, arnehelland@yahoo.no (A. Helland).

0001-4575
http://dx.doi.org/10.1016/j.aap.2012.12.042

the possibility of confounding factors. Thus, experimental studies are crucial to investigate the impairing effects of drugs and the relationship between drug concentrations, impaired performance and possible accident risk. All experimental settings are *a priori* artificial, and may thus have limited external validity when applied to real driving conditions. For instance, laboratory testing of cognitive and psychomotor functioning may measure some skills that are considered essential to safe driving, but can never fully reproduce the complexity of actual driving. Real on-road driving with measurements of standard deviation of lateral position (SDLP) has come to be considered the method of reference for assessing driving impairment from CNS depressant drugs (Verster et al., 2004), although this measure reflects mainly one (i.e., automatic behaviour) of the three "core levels" of driving (Walsh et al., 2008). Much of the on-road experiments have so far been conducted in The Netherlands on flat, straight multi-lane motorways; a driving scenario that may not reflect conditions elsewhere. Also, legal issues and safety considerations may hinder on-road experiments, and the costs of such experiments may be prohibitive.

Experimental studies utilising driving simulators may avoid some of the problems listed above. However, even very sophisticated simulators cannot fully replicate real driving conditions (Verster et al., 2004; Shechtman et al., 2009). Driving simulator studies of effects of depressant drugs on driving ability frequently yield inconclusive results due to the lack of validation against a known positive control; in practice, ethanol. The positive control is necessary to ensure that correlations between drug intake and driving related outcome measures actually reflect a drug related impairment of driving ability, and not simply randomly observed correlations with no relevance to impairment (Walsh et al., 2008). Ethanol as a positive control also ensures that the experimental design is sufficiently sensitive to the impairing effects of depressant drugs. Another common limitation of driving simulators is the lack of validation against a real driving scenario; i.e., the external validity. This leaves doubt as to whether test subject performance in the simulated scenario may predict performance in real driving situations.

We wanted to develop a valid and functional tool for assessing drug effects on driving performance, taking into account the recommendations made in the guidelines for research on drugged driving. To achieve this, we conducted a validation study of the SINTEF driving simulator. The purpose of the study was to establish a driving simulator test battery that is sensitive to ethanol effects, and to validate the test battery by comparing performance in the simulator with actual driving performance on a closed-circuit test track resembling rural driving conditions. Even though both simulator and closed circuit driving constitute experimental conditions, which do not fully reproduce the real life driving experience, both are widely used for assessing driving performance, and real driving is generally considered to be the reference methodology as far as validity is concerned. In this paper we present results from the primary outcome measure SDLP, measured in the simulator and on the test track.

2. Materials and methods

2.1. Test subjects

Twenty healthy, Caucasian, male volunteers aged 25–35 years (mean 28.7 years) who had been in possession of a driver's license for at least 5 years (mean 10.6 years), were included in the study. They were all recreational users of alcohol, and as a group drove slightly less and had a somewhat higher educational level than the general population. Women and non-Caucasians were excluded because of the teratogenic risk associated with ethanol use in the former group, and the possibility of deviant ethanol metabolism in the latter. The other exclusion criteria were previous or present drug or alcohol abuse or atypical reactions to alcohol, previous history of driving under the influence, significant adverse reactions to previous blood sampling, regular (daily) intake of any prescribed drug, or high likelihood of motion sickness as assessed with a modified version of the Apfel risk score for postoperative vomiting (Apfel et al., 1998). Each participant underwent a screening for eligibility, received written and oral information about the study and provided a written consent to participate. The study was approved by the Regional Ethics Committee, and was registered as a clinical trial in the ClinicalTrials.gov database. All participants received a gift certificate worth NOK 1000 (approx. USD 150) upon completion of the study.

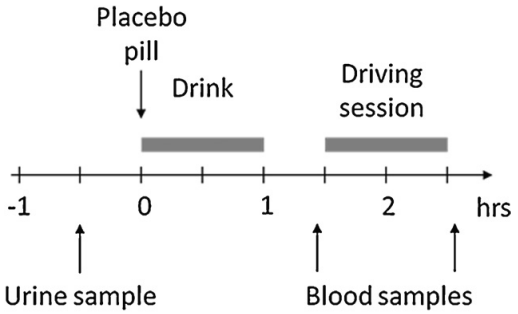

Fig. 1. Outline of trial test design.

2.2. Trial design

The experiment was designed as a randomised, placebo-controlled, single blind, crossover study. Only the necessary personnel were informed about which interventions were given. An outline of the design is presented in Fig. 1. Each participant underwent three driving tests of 1 h duration, both on a closed-circuit test track and in an advanced driving simulator, on six different test days with washout periods of minimum two days between test days to allow the dissipation of any learning or fatigue effects. The driving scenario in the simulator was modelled to mimic the test track, as illustrated in Fig. 2, to ensure that the driving experience would be as similar as possible in the two test conditions. Before testing commenced, the study subjects undertook a training session, both on the test track and in the simulator, in order to familiarise themselves with the testing scenario and minimise the impact of possible learning effects. On test days, the participants were obliged to deliver a urine sample on arrival at the test site to exclude the presence of drugs. The subjects' weight was registered each test day, after which they were administered a weight-adjusted dose of ethanol (0, 0.7 and 1.05 g per kg body weight), calculated to obtain an intended blood alcohol concentration (BAC) during testing of 0, 0.5 and 0.9 g/L on the three different test days both in the simulator and on the test track, respectively. The Widmark equation (Andréasson and Jones, 1995), was used to estimate the ethanol doses, assuming a total body water to total body mass ratio of 0.68, a bioavailability of 75%, and a metabolic rate for ethanol of 0.15 g/L per hour. We used vodka mixed with fruit extracts, orange and lime juice to make the drinks palatable. The placebo drinks were spiked with non-alcoholic vodka flavour in water to mimic the vodka taste. The drinks were served in closed plastic containers, from which the participants were instructed to sip the drink through a straw. To avoid an obvious ethanol taste, no drinks were stronger than 10% (v/v) ethanol, and they were kept cold by the addition of ice.

Fig. 2. Example of the driver's visual impression on the closed-circuit test track (left) and in the driving simulator (right).

The participants were allowed 1 h to finish their drinks, after which they waited another 30 min before the driving test started, to allow for absorption of the administered ethanol. The order in which the participants were tested at different BAC levels was randomised by use of a counterbalanced, multi-condition design. The same order of BAC levels was used for each participant both on the test track and in the simulator. As an additional confounder to enhance blinding, the study subjects were administered a placebo pill, which they were told may or may not contain a sedative drug, with the drink. Venous blood samples were drawn immediately before and after each driving session, and the mean value was used as the best estimate of the mean BAC during testing.

2.3. Real driving on test track

The test track driving was undertaken during a frost-free period of six weeks in the autumn. All study sessions were done after nightfall, between 20:00 and 01:00 h. The test track circuit was 1.37 km long, closed to ordinary traffic, and laid out in hilly terrain, with both gentle and sharper curves. The track was hard-surfaced, with two lanes each approx. 2.75 m wide, and had midline and side markings similar to standard Norwegian road markings. Thus, the test track closely resembled roads typical of rural Norway. Surprise obstacles (1 m^3 foam rubber cubes) were placed in two locations on two occasions, one at the beginning and one towards the end of each driving trip, and were to be avoided by the test subjects. Stoplights present in two locations turned red on one occasion during each trip. The participants drove an instrumented car (Volvo V70 2.4s) with automatic transmission, fitted with a double set of pedals. They were instructed to drive as they would normally do on a regular road. A professional driving instructor was present in the front passenger seat during all sessions of test track driving, in order to intervene if necessary. A physician was present on the site at all times during test drives. Permission to carry out the test track driving was granted from the local police. To enable continuous recording of lateral position in the road lane, the test car was equipped with an infrared wide-angle camera fixed to the roof of the car, and pointing at a downward angle to the rear of the car. The data were stored in a database and analysed in a program for photo analysis (Open Source Computer Vision Library). A filtering algorithm (Hough transformation) was used to identify roadside markings. The car also featured other equipment for recording the location of the car on the test circuit (global positioning system; GPS), speed, pedal use and steering wheel movements.

2.4. Driving simulator

Testing in the driving simulator took place in late autumn after the test track driving tests were completed. Test sessions were done at the same times during the evening and night as on the test track, using a virtual model of the test track and a nighttime scenario (Fig. 2), to ensure comparable results and eliminate differences in circadian influences. In addition to obstacles and stoplights, the simulator scenario also included two incidents (a car abruptly entering the road and a pedestrian crossing the road in front of the driver) that each occurred once at the end of the driving session. The simulator had the appearance of a regular car (Renault Scenic) with automatic transmission and original controls (Fig. 3). Information from the use of steering wheel, pedals, transmission etc. was fed into a dedicated driving scenario graphics computer. The driving scenario was depicted on screens covering 180° of the driver's forward field of vision and 90° of the rear field of vision, and synchronously in internal and external mirrors. The vertical field of view was 47° both to the front and to the rear. The simulator reproduced realistic motion, vibration and sound through a three-axis moving platform, a vibration system in the chassis and a four-channel sound system. Data on lateral position, speed, pedal use and steering wheel movements over the entire duration of the test sessions were extracted directly from the simulator computer and logged 20 times per second. A detailed description of the SINTEF simulator can be found in Engen (2008).

2.5. Measurements

The predefined primary outcome measure was the standard deviation of lateral position (SDLP), which is a measure of the degree of weaving of the car on the road. SDLP has been shown to correlate with BAC levels in a dose dependent manner, and is a thoroughly validated measure of the degree of driving impairment (Verster et al., 2004). Secondary outcome measures were number of brake pedal pressures per lap, number of accelerator

Fig. 3. Setup of the driving simulator. Vehicle and surrounding frontal screens.

pedal pressures per lap, steering wheel movement speed, steering wheel movement per lap, steering wheel reversals per lap, steering wheel reversal frequency, average speed, standard deviation of speed (measured continuously throughout the driving sessions), driving behaviour at unexpected incidents, and driving against red light. We aim to present the secondary outcome measures in a subsequent article.

Before and after each driving session, the participants completed a questionnaire, with items covering their feelings of intoxication, mastery, safety, sleepiness, alertness, whether they thought the drink had contained ethanol, and whether they thought the pill had contained a sedative drug. At the test track, driving instructors were also asked to rate the test subjects' degree of intoxication and driving performance.

Blood ethanol concentrations were quantified using a headspace gas chromatography–mass spectrometry (GC–MS) method. In brief, 200 μL blood was mixed with 50 μL internal standard (d6-ethanol). Samples were left for 30 min to achieve equilibrium before the gas fraction was aspirated into an Agilent HP 6890-5973 GC–MS system (Agilent, Palo Alto, CA). Separation was performed on a J&W Scientific 123-9134 DB-ALC1 (30 m × 1.2 mm) column with a helium mobile phase and a run time of 0.90 min. Ethanol was monitored at m/z 31 and the internal standard at m/z 33. The level of quantification (LOQ) was 2 mmol/L (approx. 0.09 g/L). Between-day coefficient of variation (CV) calculated from quality control samples was 4.5% at 5 mmol/L (0.22 g/L) and 1.8% at 50 mmol/L (2.2 g/L).

2.6. Statistical analyses

An *a priori* sample size estimation performed with one-tailed, paired t-tests indicated that a total sample size of $n = 11$ would be sufficient to detect significant differences in BAC level influence on SDLP with significance level (α) of 0.05 and power ($1-\beta$) of 0.95. Although theoretically 11 subjects would suffice, we chose to include 20 subjects in the study, to allow for the uncertainty in the underlying assumptions of the sample size estimation, as well as the possibility of dropouts, for instance due to simulator sickness.

In the results analyses, we used a linear mixed model with SDLP as dependent variable, measured BAC as covariate, and participant as random effect. Separate analyses were performed for test track and simulator. Reported results are from restricted maximum likelihood estimation. The maximum likelihood estimation did not always converge. The independent variables tested for significance were BAC level, curved/straight section and part of trip driven (each trip was divided in four equal parts of 15 min). To identify possible learning effects that could interfere with the results, the impact of the number of trips driven before the actual one was also analysed. Two-sided p-values <0.05 were considered significant. The analyses were performed in SPSS 18 and Stata 12.

3. Results

Of the 20 participants enrolled in the study, all completed three driving sessions on the test track, while 18 out of 20 completed all three sessions in the driving simulator. Two subjects did not complete the simulator testing; one because of intolerable nausea, and the other because of a surgical procedure unrelated to the study. On the test track, 10 out of the 60 driving sessions did not yield sufficient SDLP data to be included in the analyses. The car-mounted camera was out of position in eight sessions, the camera was not switched on in one instance, and one participant in his first session misinterpreted the instructions to drive in lane. Thus, a complete set of outcome data was obtained from 10 participants on the test track and 18 participants in the simulator. Data from the valid driving sessions of all subjects were included in the analyses.

Table 1
Measured blood ethanol concentrations (BAC) in simulator driving and on test track at the three designated BAC levels of 0, 0.5 g/L and 0.9 g/L among all test subjects with samples.

Test scenario	Intended BAC	Mean BAC (±SD)
Simulator ($n = 19$)	0	0
	0.5 g/L	0.38 (±0.10) g/L
	0.9 g/L	0.82 (±0.19) g/L
Test track ($n = 20$)	0	0
	0.5 g/L	0.42 (±0.09) g/L
	0.9 g/L	0.88 (±0.12) g/L

3.1. Safety and adverse events

No safety violations or serious or unexpected adverse events occurred during the study. The most common adverse event in the simulator was nausea, which is a known disadvantage of driving simulators. Six subjects (four at BAC 0 and two at BAC 0.5) had to terminate their first simulator session early because of this, but five of them were eventually able to complete all three sessions. Thus, only one subject had to withdraw from the study due to nausea. Prior experience suggests that ethanol may protect against simulator sickness, and repeated exposures to the simulator tend to attenuate the nausea. Therefore, in order to prevent dropouts, all participants who terminated their sessions early due to nausea were tested at the highest BAC level in the subsequent session. The random order was also modified in an additional three subjects due to other practical causes. These modifications to the randomisation did not affect concealment of the interventions, and did not appear to introduce systematic bias, since there was no statistically significant correlation between BAC level and the number of previous test sessions (Pearson correlation 0.241 ($p = 0.080$) in simulator and 0.094 ($p = 0.477$) on test track).

3.2. Blood alcohol concentrations

The ethanol concentrations are presented in Table 1. Ethanol concentrations were slightly lower than intended both in the simulator and on the test track, with concentrations closer to 0.4 g/L at the intended level of 0.5 g/L. The BAC also tended to be slightly lower in the simulator than on the test track. Paired sample t-test showed a statistically significant difference between the BAC levels in simulator and on test track for the designated BAC level of 0.5 g/L ($p = 0.041$); however, the mean difference was only 0.039 g/L. For the designated BAC level of 0.9 g/L, there was no statistically significant difference between BAC levels in simulator and on test track ($p = 0.21$). In the following, ethanol levels are referred to as the intended levels (BAC 0, BAC 0.5 and BAC 0.9, respectively).

3.3. Questionnaires

After each driving session, the participants were asked whether they thought the drink and the pill had contained alcohol and a sedative drug, respectively. Most subjects correctly identified the drink as containing/not containing ethanol (in 32 of 38 placebo trials, 35 of 38 BAC 0.5 trials and 37 of 38 BAC 0.9 trials, respectively). However, a few misidentified their drinks, and quite a few wrongly identified the pill as containing a sedative drug (in 15 of 38 placebo trials, 3 of 38 BAC 0.5 trials and 7 of 38 BAC 0.9 trials, respectively).

There were significant correlations between higher BAC levels and subjective (self reported) ratings of poorer driving performance both in the simulator ($R = 0.35$, $p = 0.013$) and on the test track ($R = 0.63$, $p < 0.001$). Likewise, there was a strong correlation between higher BAC levels and objective (driving instructor

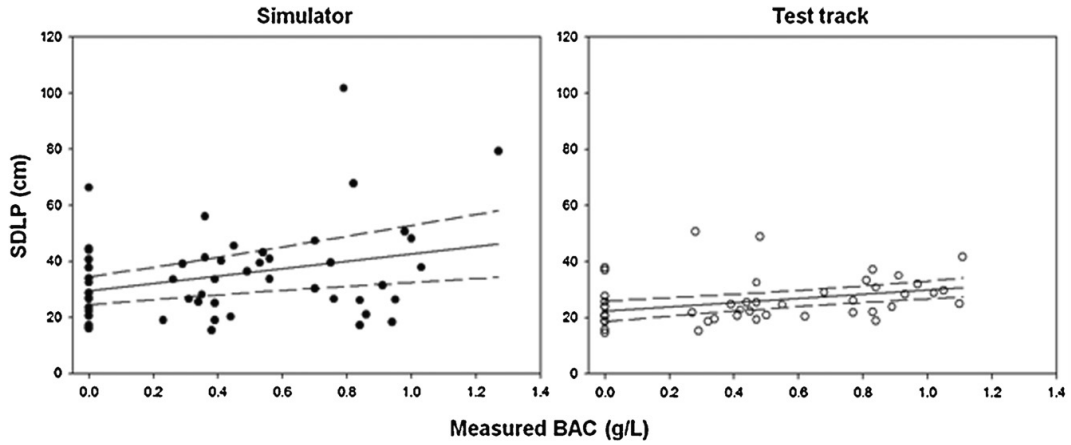

Fig. 4. Regression analysis of the relationship between blood alcohol concentration (BAC) and standard deviation of lateral position (SDLP) in simulator (left; filled circles) and on test track (right; open circles). The circles represent individual BAC and the corresponding SDLP value. The regression lines and their 95% confidence intervals are shown as continuous lines and broken lines, respectively.

reported) ratings of poorer driving performance on the test track ($R = 0.52$, $p < 0.001$).

3.4. SDLP

Fig. 4 shows the individual SDLP values at the corresponding BAC, with the estimated regression line and its 95% confidence interval. Both in the simulator and on the test track, there were significant positive correlations between BAC and SDLP (positive regression slope with $p < 0.001$). The estimated regression lines for the simulator (Eq. (1)) and the test track (Eq. (2)) are as follows, with standard errors for the estimates in parentheses:

$$(\text{simulator}) : \text{SDLP}(\text{cm}) = 29.43\,(\pm 2.57) + 13.20\,(\pm 3.61) \times \text{BAC} \quad (1)$$

$$(\text{test track}) : \text{SDLP}(\text{cm}) = 22.30\,(\pm 1.89) + 7.61\,(\pm 1.91) \times \text{BAC} \quad (2)$$

SDLP values were higher in the simulator than on the test track at baseline (placebo) conditions (29.4 cm vs. 22.3 cm, respectively), and showed a steeper increase with increasing BAC, as seen from Eqs. (1) and (2), as well as Fig. 4. As evident from Fig. 4, SDLP variance was also larger in simulator driving than in test track driving.

The relationship between BAC levels and SDLP results show a dose–response effect, as quantified by the slopes 13.20 and 7.61 in Eq. (1) and (2). Furthermore, a visual comparison of SDLP results in the simulator and on the test track in each of the 20 individual subjects shows similar, positive slopes in most subjects (Fig. 5).

To identify possible differential effects of test duration and curved/straight sections on SDLP, the SDLP results were analysed with respect to time intervals (four equal intervals of 15 min each), and performance on curved and straight sections of the driving scenario. In the simulator, mean SDLP values were significantly higher in curved sections than in straight sections ($p = 0.047$), whereas there were no such differences on the test track ($p = 0.17$). In the simulator, statistically significant differences in SDLP between BAC levels were seen in all four time intervals. On the test track, the differences in SDLP were similar but less pronounced, and mostly did not reach significance during the first half hour of the test. In the simulator, there was a trend towards higher SDLP values with longer test duration, especially at the highest BAC level. No such tendency was evident on the test track.

To identify possible learning effects that would be expected to reduce SDLP with the number of prior test sessions, the number of trips driven before the actual one was also analysed as an independent variable. However, this had no statistically significant correlation with SDLP results either in the simulator ($p = 0.70$) or on the test track ($p = 0.66$).

4. Discussion

4.1. SDLP

Our results show a positive dose–response correlation between BAC and SDLP in the simulator and on the test track, both for individual and mean data. A high degree of intra-individual similarity in the BAC-correlated increase in SDLP in the simulator and on the test track, suggests that SDLP is a valid and sensitive measure of ethanol-induced driving impairment in the simulator.

Absolute values of SDLP were higher in the simulator than on the test track, with mean SDLP at BAC 0 (sober state) of 29 cm and 22 cm, respectively. SDLP values during placebo conditions in the simulator were also considerably higher than those seen in Dutch on-road driving tests, where mean baseline SDLP is approx. 19 cm (range 9–30 cm) (Verster and Roth, 2011). The relatively demanding driving scenario that was used in our experiment may account for the slightly higher SDLP values on the test track than those seen during previous on-road tests. Higher absolute SDLP values in the simulator compared to real driving may be explained by unfamiliarity with the driving experience in the simulator, a lack of perceived danger, and lack of gravitational cues and feedback that will normally adjust steering. This notion is also supported by the observation that SDLP values were higher in curved sections than in straight sections in the simulator, whereas such a difference was not observed on the test track. Together with the more demanding driving scenario in our experiment, this may account for the considerably higher SDLP values than those seen for instance in the Dutch STISIM simulator employing a monotonous highway scenario (Mets et al., 2011b).

Most test subjects showed similar SDLP increases in the simulator and on the test track. However, from the individual SDLP data shown in Fig. 5, a few subjects behave differently, evidenced by excessive SDLP values in the simulator. For instance, test subject no. 15 had a mean SDLP exceeding 1 m at the highest BAC level. This would correspond to the car being located mostly out of lane during the trip, which is in accordance with the actual observations made

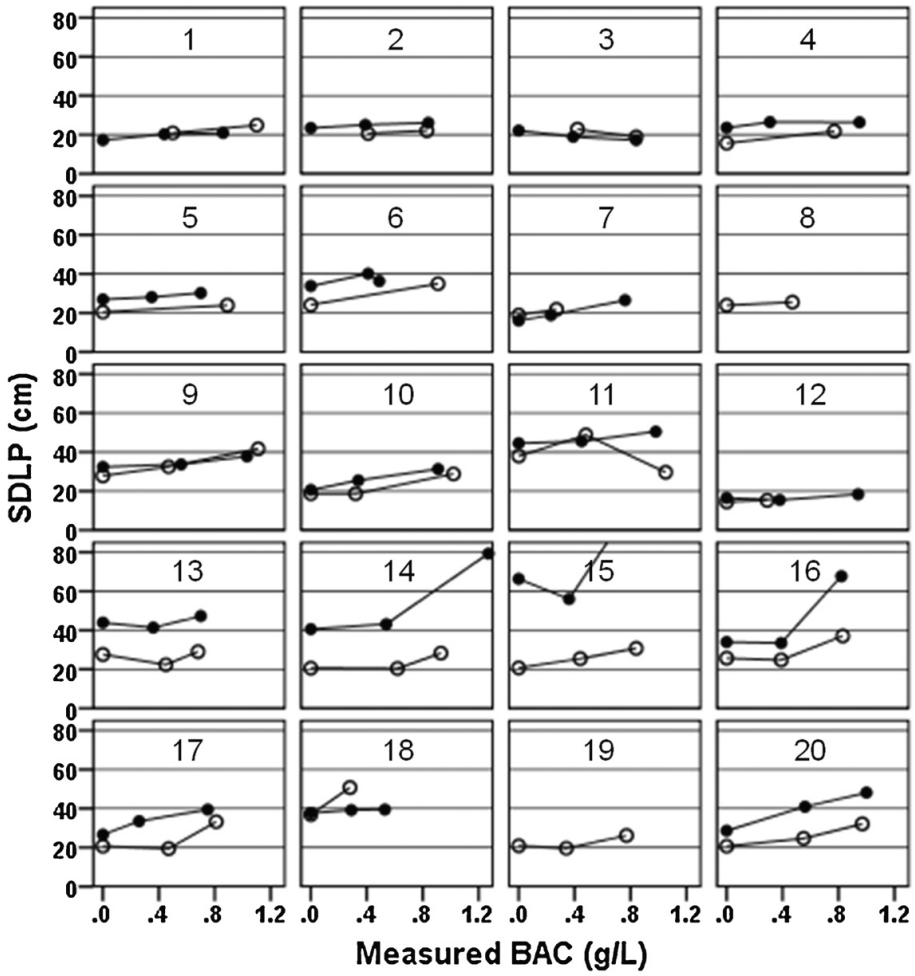

Fig. 5. Individual SDLP data at actual BAC levels in simulator (filled circles) and on test track (open circles). For test subject 15, the BAC and SDLP values at the highest BAC level in the simulator were 0.79 g/L and 102 cm, respectively.

during this individual's simulator driving. It is our experience from the present and earlier simulator experiments that some participants regard the simulator as a kind of game and behave more like virtual rally drivers instead of following the instructions to drive appropriately according to conditions. This can explain the large discrepancies in SDLP between test track and simulator seen in a few of the subjects. Subject no. 14 attained an unexpectedly high BAC at his highest BAC level in the simulator (1.25 g/L), which may explain the high SDLP observed in that driving session. Also, we cannot exclude the possibility that some participants' SDLP scores were influenced by simulator sickness.

4.2. BAC

Mean subjective and objective ratings of intoxication and driving performance correlated with BAC level in the expected manner. The somewhat lower BAC levels in simulator than on test track may be due to a possible conditioned nausea response in the simulator that could have caused retention of stomach content with delayed ethanol absorption. One participant (subject no. 6) was unable to finish his drink at the intended BAC 0.9 level in the simulator, and consequently acquired a low BAC.

Most participants correctly identified their drink as containing/not containing ethanol and the pill as containing/not containing a sedative drug, although quite a few of the participants misidentified the placebo pill, especially in the BAC 0 trials. This probably reflects an expectation bias in some subjects, and indicates that the use of placebo pills to enhance blinding of the intervention in experimental trials with ethanol may be worthwhile. Previous experience suggests that concealment of ethanol is difficult in blinded studies due to the distinctive taste and smell and the characteristic and familiar effects of ethanol.

4.3. Comparison with other driving simulator studies and on-road tests

To date, there are few other studies validating the use of driving simulators for drug and/or ethanol impairment research. A simulator validation study published in 2009 used data from two separate previous studies (on-road and in simulator). The

description of the simulator they used suggests that it was similar to the SINTEF simulator, but the driving scenario and outcome measure were different (urban traffic and number of driving errors at intersections as assessed by a driving instructor, respectively). No ethanol or other drugs were used. Results indicated relative validity for the simulator, and suggested absolute validity for the type of errors pertaining lane maintenance, adjustment to stimuli and visual scanning (Shechtman et al., 2009).

There are few previous simulator studies using SDLP as outcome measure. Only one study has validated SDLP as an indicator of unsafe driving in the simulator that was used. Mets et al. published a validation study in 2011 showing the ability of the STISIM driving simulator to differentiate between different BAC levels based on SDLP results. In this study, 27 healthy volunteers underwent a simulator adaptation of the standardised Dutch on-road test scenario (multi-lane highway driving for 1 h). BAC levels of 0 g/L, 0.5 g/L, 0.8 g/L and 1.1 g/L yielded mean SDLP values of 28.0 cm, 29.7 cm, 33.8 cm and 36.3 cm, respectively. This study did not validate the simulator results against a real driving test (Mets et al., 2011b). Apart from this, only two simulator studies concerning driving performance after drug intake have been published using SDLP as an outcome measure. Mets et al. have investigated the effects of caffeine (given in the form of the energy drink Red Bull® and coffee, respectively) on driving performance in healthy volunteers in two studies in the Dutch STISIM simulator, and found small but significant reductions in SDLP after caffeine administration in both studies (Mets et al., 2011a, 2012).

In 2009, a validation study with ethanol in a divided-attention steering simulator (DASS) was published. As the name suggests, the simulator is designed to measure ability of divided attention. Accordingly, it employs a rather artificial test scenario, where subjects must keep the car in lane and simultaneously respond to peripheral visual stimuli. Also, the simulator used did not resemble a normal car. Dose-dependent impairment was found with higher ethanol levels (Verster et al., 2009b).

The standardised on-road driving test with SDLP as the outcome measure developed in The Netherlands remains the method of reference to examine driving impairment from drugs. In such testing, BAC levels of 0.5 g/L and 0.8 g/L on average increases SDLP from placebo conditions with 2.4 cm and 4.3 cm, respectively (Verster and Roth, 2011). Our results from the test track show slightly larger increases in SDLP, whereas the BAC-related increases in the simulator were considerably larger. Again, the discrepancy between our results and the Dutch on-road results may be explained by the more demanding driving scenario employed in our validation study.

4.4. Implications for the validity and further use of the simulator

External validity of a driving simulator refers to the test scenario's ability to invoke similar reactions in the drivers as a real driving scenario. Validity is specific for the particular type of scenario and simulator, test, and population used in the validation experiments, and will not necessarily be transferable to other driving scenarios, simulators, tests, or populations. External validity is absolute if the same effect is invoked to the same extent both in the simulator and in the real driving environment. Relative external validity implies that there exists a trend of change in the same direction both in the simulator and in the real driving environment, but the magnitude of change is different (Shechtman et al., 2009).

There was a large degree of similarity in the relationship between SDLP and BAC levels in the simulator and on the test track. However, the absolute values of SDLP in the simulator were consistently higher than on the test track. Thus, the relative (but not the absolute) external validity of the SINTEF simulator has been established when validated against test track driving in a driving scenario that is representative of the demanding rural driving conditions in Norway, using ethanol as a positive control. We believe that this validation may be extended to real driving under similar conditions; however, this assumption has not been proven.

In the simulator, we found consistent and significant BAC-related increases in SDLP in all time intervals when the hour-long test was divided into four 15-min time intervals. This suggests that the duration of the simulator test in order to reach significant results may be shortened in future studies.

4.5. Limitations of the study

In our study, all test subjects were healthy young male volunteers, who are not representative for the general driving population. Our results may therefore give a somewhat inaccurate estimation of the impact of BAC on SDLP in the general population.

There are three levels of behaviour relevant to traffic safety: automatic, control and executive planning behaviour (Michon, 1985; Walsh et al., 2008). SDLP as the primary outcome measure in this study is mainly representative for the effect of ethanol on automated actions at a behavioural control level. Outcome measures of driving behaviour at manoeuvring and strategic levels will be reported in a separate publication. Driving simulators may be especially suitable to test higher behavioural levels like hazard avoidance, dual attention, risk taking and impulsivity, both for ethical (risk of injury) and practical (ease and reliability of measurements) reasons.

We employed a single blind design, keeping the intervention concealed from the test subjects but not from the study personnel or those responsible for analysing the outcome data.

Unlike some of the most advanced simulators in use, the SINTEF simulator allows only limited tilting (three degrees of freedom). Motion-based simulators with full tilting technology might increase the realism of the driving experience, and thus heighten the external validity of the simulator.

Several of the test subjects experienced nausea in the simulator, which caused one subject to withdraw from the study, and may have affected driving behaviour in others. This is a general drawback of driving simulators, which may to some extent be unavoidable, even when using screening procedures including test drives before enrolment. We also employed a rather challenging driving scenario, with many curves and long duration, which may have exacerbated the problems related to nausea.

The validation against real driving was done on a closed test track. The length (approx. 1.4 km) and layout (curvy, hard-top road approx. 5.5 m wide with midline and side markings) of the test track ensured that the driving experience resembled real driving on rural Norwegian roads. However, it may be impossible to fully eliminate the feeling of an artificial situation when driving on a closed test track. For safety reasons, a driving instructor was present in the passenger seat at all times on the test track, as well as a police officer on the test track site. This may have constituted a restraining effect as well as heightened the attention of test subjects, causing them to drive more carefully and attentively than they would otherwise have done.

Finally, our study had a limited sample size, which generally increases the risk of type II errors (i.e., failing to detect real differences). Also, missing data from 10 of 60 driving sessions on the test track may have limited the statistical significance of our findings. The missing data occurred due to random incidents, and we have no reason to believe this introduced systematic bias.

5. Conclusions

In healthy volunteers, SDLP as a measure of drug-impaired driving shows qualitatively similar outcomes during test track driving and in a driving simulator designed to mimic the test track, both sober and under the influence of ethanol. However, SDLP is amplified in the simulator as compared to real driving. Although closed circuit driving is an experimental situation and thus of limited external validity, the quantitative and qualitative similarities between simulator and test track driving nevertheless imply external validity of the simulator. In conclusion, the SINTEF driving simulator is a sensitive and valid tool to assess driving impairment from ethanol, and this may be extended to include other CNS depressant drugs.

Acknowledgements

The research project was funded by a grant from the Research Council of Norway.

The authors would like to thank the following employees at SINTEF Technology and Society and the Department of Clinical Pharmacology at St. Olav University Hospital who made valuable contributions to the research project: Thomas Engen, Dagfinn Moe, Kenneth Sørensen, Terje Giæver, Cato Mausethagen, Nina Husby, Marianne Nordtømme, Odd Hjelkrem, Guro Haltbrekken, Anita Skogholt, Trude Kristine Vaag Gjerde.

We would also like to thank Prof. Dr. Jan Ramaekers at Maastricht University, The Netherlands, for valuable advice and helpful discussions during the preparations of the study.

Finally, we would like to express our gratitude to the study subjects who uncomplainingly underwent the rather demanding trials.

References

Andréasson, R., Jones, A.W., 1995. Erik M. P. Widmark (1889–1945): Swedish pioneer in forensic alcohol toxicology. Forensic Science International 72 (1), 1–14.

Apfel, C.C., Greim, C.A., Haubitz, I., Goepfert, C., Usadel, J., Sefrin, P., Roewer, N., 1998. A risk score to predict the probability of postoperative vomiting in adults. Acta Anaesthesiologica Scandinavica 42 (5), 495–501.

Blomberg, R.D., Peck, R.C., Moskowitz, H., Burns, M., Fiorentino, D., 2009. The Long Beach/Fort Lauderdale relative risk study. Journal of Safety Research 40 (4), 285–292.

Borkenstein, R.F., Crowther, R.F., Shumante, R.P., Ziel, W.B., Zylman, R., 1974. Die Rolle des alkoholisierten Fahrers bei Verkehrsunfällen (the Grand Rapids study). Blutalkohol 11, 1–7.

Engen, T., 2008. Use and validation of driving simulators. Thesis. Norwegian University of Science and Technology.

Mets, M.A., Baas, D., Van Boven, I., Olivier, B., Verster, J.C., 2012. Effects of coffee on driving performance during prolonged simulated highway driving. Psychopharmacology (Berlin) 222 (2), 337–342.

Mets, M.A., Ketzer, S., Blom, C., Van Gerven, M.H., Van Willigenburg, G.M., Olivier, B., Verster, J.C., 2011a. Positive effects of Red Bull® energy drink on driving performance during prolonged driving. Psychopharmacology (Berlin) 214 (3), 737–745.

Mets, M.A., Kuipers, E., Senerpont Domis, L.M., Leenders, M., Olivier, B., Verster, J.C., 2011b. Effects of alcohol on highway driving in the STISIM driving simulator. Human psychopharmacology. [Epub ahead of print].

Michon, J.A., 1985. A critical view of driver behavior models: what do we know, what should we do? In: Evans, L., Schwing, R.C. (Eds.), Human Behavior and Traffic Safety. Plenum Press, New York, pp. 485–520.

Ramaekers, J.G., Berghaus, G., Van Laar, M., Drummer, O.H., 2004. Dose related risk of motor vehicle crashes after cannabis use. Drug and Alcohol Dependence 73 (2), 109–119.

Shechtman, O., Classen, S., Awadzi, K., Mann, W., 2009. Comparison of driving errors between on-the-road and simulated driving assessment: a validation study. Traffic Injury Prevention 10 (4), 379–385.

Verster, J.C., Pandi-Perumal, S.R., Ramaekers, J.G., De Gier, J.J., 2009a. Drugs, Driving and Traffic Safety. Birkhäuser Verlag AG, Basel, Switzerland.

Verster, J.C., Roth, T., 2011. Standard operation procedures for conducting the on-the-road driving test, and measurement of the standard deviation of lateral position (SDLP). International Journal of General Medicine 4, 359–371.

Verster, J.C., Veldhuijzen, D.S., Volkerts, E.R., 2004. Residual effects of sleep medication on driving ability. Sleep Medicine Reviews 8 (4), 309–325.

Verster, J.C., Wester, A.E., Goorden, M., Van Wieringen, J.P., Olivier, B., Volkerts, E.R., 2009b. Novice drivers' performance after different alcohol dosages and placebo in the divided-attention steering simulator (DASS). Psychopharmacology (Berlin) 204 (1), 127–133.

Walsh, J.M., Verstraete, A.G., Huestis, M.A., Mørland, J., 2008. Guidelines for research on drugged driving. Addiction 103 (8), 1258–1268.

Paper II

Is not included due to copyright

Paper III

Accident Analysis and Prevention 94 (2016) 180–187

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Driving simulator sickness: Impact on driving performance, influence of blood alcohol concentration, and effect of repeated simulator exposures

Arne Helland [a,b,*], Stian Lydersen [c], Lone-Eirin Lervåg [d], Gunnar D. Jenssen [d], Jørg Mørland [e], Lars Slørdal [a,b]

[a] Department of Clinical Pharmacology, St. Olav University Hospital, Trondheim, Norway
[b] Department of Laboratory Medicine, Children's and Women's Health, Norwegian University of Science and Technology, Trondheim, Norway
[c] Regional Centre for Child and Youth Mental Health and Child Welfare—Central Norway, Norwegian University of Science and Technology, Trondheim, Norway
[d] SINTEF Technology and Society, Department of Transport Research, Trondheim, Norway
[e] National Institute of Public Health, Oslo, Norway and Institute of Clinical Medicine, University of Oslo, Norway

ARTICLE INFO

Article history:
Received 9 October 2015
Received in revised form 25 April 2016
Accepted 8 May 2016

Keywords:
Automobile driving
Traffic accidents
Ethanol
Simulated driving
Simulator sickness

ABSTRACT

Simulator sickness is a major obstacle to the use of driving simulators for research, training and driver assessment purposes. The purpose of the present study was to investigate the possible influence of simulator sickness on driving performance measures such as standard deviation of lateral position (SDLP), and the effect of alcohol or repeated simulator exposure on the degree of simulator sickness. Twenty healthy male volunteers underwent three simulated driving trials of 1 h's duration with a curvy rural road scenario, and rated their degree of simulator sickness after each trial. Subjects drove sober and with blood alcohol concentrations (BAC) of approx. 0.5 g/L and 0.9 g/L in a randomized order. Simulator sickness score (SSS) did not influence the primary outcome measure SDLP. Higher SSS significantly predicted lower average speed and frequency of steering wheel reversals. These effects seemed to be mitigated by alcohol. Higher BAC significantly predicted lower SSS, suggesting that alcohol inebriation alleviates simulator sickness. The negative relation between the number of previous exposures to the simulator and SSS was not statistically significant, but is consistent with habituation to the sickness-inducing effects, as shown in other studies. Overall, the results suggest no influence of simulator sickness on SDLP or several other driving performance measures. However, simulator sickness seems to cause test subjects to drive more carefully, with lower average speed and fewer steering wheel reversals, hampering the interpretation of these outcomes as measures of driving impairment and safety. BAC and repeated simulator exposures may act as confounding variables by influencing the degree of simulator sickness in experimental studies.

1. Introduction

Driving simulation has numerous uses, such as training purposes, assessment of possibly unfit drivers and research in the fields of traffic safety and driving under the influence of alcohol and drugs (DUI) (Classen and Brooks, 2014). Driving simulators enable researchers to assess performance in various driving environments (i.e., city driving, highway driving, or situations or settings with high accident risk) under controlled laboratory conditions. Furthermore, simulators allow convenient measurement of several aspects of driving behavior.

One major obstacle to the use of driving simulators is the phenomenon of simulator sickness, a syndrome resembling motion sickness with symptoms including dizziness, cold sweats, drowsiness, nausea and vomiting. Simulator sickness is most likely caused by an incongruity of sensory input, with conflicting signals from simulated and actual motion, although other theories of causation also exist (Brooks et al., 2010). A variable but considerable proportion of test subjects in simulator trials experience simulator sickness, some to the extent that they are unable to complete simulator testing. For example, a study combining the results from

* Corresponding author at: Department of Clinical Pharmacology, St. Olav University Hospital Pb 3250 Sluppen, NO–7006 Trondheim, Norway.
E-mail address: arne.helland@stolav.no (A. Helland).

http://dx.doi.org/10.1016/j.aap.2016.05.008
0001-4575/

several simulator studies reported a dropout rate due to simulator sickness of 17% (Brooks et al., 2010). Increased dropout rates reduce power and, perhaps more problematic, could introduce bias in the study population and confound results (Brooks et al., 2010; Classen et al., 2011).

The scientific literature on what influences driving simulator sickness and its impacts on performance is limited. Some factors that increase the likelihood of simulator sickness have been identified. These are related to the test subjects (i.e., older age, female sex, certain psychological states and traits), the test scenario (longer duration, more curves and turns, higher speeds, increased visual detail) and the technical setup of the simulator (broader field of vision, disagreement or delay between instrument operation and response of the virtual car) (Classen et al., 2011; Milleville-Pennel and Charron, 2015). Some techniques to alleviate simulator sickness have also been identified, including adaptation over time and neural or sensory stimulation (Domeyer et al., 2013; Galvez-Garcia et al., 2015). Hence, researchers of simulated driving may employ measures to limit the problem of simulator sickness to a certain extent. Various screening questions (i.e., history of motion sickness) and pre-trial testing are commonly used to exclude subjects that are prone to severe simulator sickness from experimental studies. Nevertheless, it is presently impossible to avoid completely the occurrence of simulator sickness in such studies (Brooks et al., 2010).

External validity is a precondition to the use of simulators – we must be able to trust that the data are relevant to real life. Thus, aspects of the simulator experience that differ significantly from the real-life driving experience must be investigated to determine if they influence measurements of driving safety directly, or if they in some way introduce bias in the interpretation of data. When present, simulator sickness may cause significant behavioral changes that could conceivably influence outcomes. Therefore, research on simulator sickness is important to assess the validity of simulator data, and to be better able to minimize the impact of simulator sickness on the results.

Although negative effects of virtual reality-induced symptoms (a syndrome resembling simulator sickness) on psychomotor control have been described (Cobb et al., 1999), little is known about the influence of simulator sickness on validated and commonly used measures of impaired driving in experimental studies, such as standard deviation of lateral position (SDLP). Thus, there is a risk that simulator sickness may confound the results. In addition, if simulator sickness leads to significant changes in the way test subjects drive, this could weaken the generalizability and validity of driving simulation results. In DUI research, alcohol is often used as a positive control (Walsh et al., 2008), yet alcohol inebriation may be associated with nausea as well as other complex central nervous effects that could influence symptoms of simulator sickness. Therefore, simulator sickness could be a source of operational confounding in such studies. Moreover, many studies use a design with repeated driving trials, where for instance a drug is given in different doses and/or compared to a placebo. Repeated exposures to the simulator might influence the degree of simulator sickness through either habituation or sensitization, which could pose a risk of procedural confounding. Two previous studies lend support to a habituation effect of repeated exposures (Kennedy et al., 2000; Domeyer et al., 2013). In an unpublished pilot study we conducted, we observed that the test subjects tended to complain less about simulator sickness when driving under the influence of alcohol, and after repeated exposures to the simulator. Given these observations, it seems prudent to further investigate the influence of such factors on the degree of simulator sickness.

In this paper, we explore the possible influence of simulator sickness on several measures of impaired driving, including SDLP, without making any pre-specified predictions regarding the direction of the outcomes. Based on findings in our pilot study, we also investigate the effect of blood alcohol concentration (BAC) and repeated exposures to the simulator on the reported degree of simulator sickness, hypothesizing that alcohol and repeated exposures attenuate simulator sickness.

2. Material and methods

The data presented in this article were generated in a validation study designed to compare driving performance in real and simulated driving at three levels of alcohol inebriation.

2.1. Test subjects

Twenty healthy, Caucasian males aged 25–35 years (mean: 28.7 years) were included in the study. The test subjects were recruited through medical students' organizations, student- and employee networks at the Norwegian University of Science and Technology, and the employee website of the SINTEF research institute. They were all recreational drinkers, and had all been in possession of a driver's license for at least 5 years (mean: 10.6 years) As a group they drove slightly more and were somewhat higher educated than the average population. For instance, 25% of our test subjects drove <10,000 km/year, compared to 35% in the general population, and 25% drove >20,000 km/year, compared to 18% in the general population. We recruited a rather narrow age group to minimize variability in driving experience and ethanol tolerance. Exclusion criteria were female sex, non-Caucasian ethnicity, prior or present drug/alcohol abuse, previous history of deviant (violent or aggressive) alcohol reactions or driving under the influence, intolerance to blood sampling, daily intake of any drug, or high likelihood of simulator sickness. We chose to exclude females because of the teratogenic effects of ethanol, which would necessitate interviews and administration of pregnancy tests before each test run. Non-Caucasians were excluded to avoid uncontrolled variation in ethanol tolerance and metabolism. The subjects received written information about the possibility of nausea/simulator sickness prior to inclusion, and that they were free to terminate the simulator driving anytime during the session. To avoid a high likelihood of simulator sickness, all volunteers were assessed with a modified version of the Apfel risk scale for postoperative vomiting (Apfel et al., 1998). The scale contained three items: Smoking status (yes = 0, no = 1), previous nausea/vomiting after surgery or other invasive procedures (yes = 1, no = 0), and car sickness after the age of 10 (yes = 1, no = 0). Persons with a score of 2 or higher were excluded. This method has not been validated to identify persons with high risk for simulator sickness. Before final inclusion, prospective participants underwent a screening trial of 20 min' duration in the simulator to exclude persons with excessive simulator sickness and familiarize them with the simulator to minimize learning effects. Three potential participants were excluded due to simulator sickness during the pretest trial. Information about the possibility of simulator sickness was repeated orally both at the pretest trial and at each study session. Each participant gave his informed consent and the study was approved by the Regional Ethics Committee.

2.2. Trial design

Each participant underwent three 1-h nighttime driving tests in the simulator, with at least 2 days between each test. The experiment was conducted as a randomized, placebo-controlled, single blind study, using a counterbalanced, multi-condition design to randomize the order in which the subjects were tested at different BAC. The intervention was concealed from study subjects, who also received sham treatment in the form of a placebo pill before

Fig. 1. Appearance of the driving simulator.

each driving session, which they were told may or may not contain a sedative drug, to further enhance concealment. Vodka or ethanol-free vodka extract was mixed with fruit juices and administered in weight-adjusted doses (0, 0.7 and 1.05 g per kg body weight) 1.5 h before the start of the driving task, aiming to achieve mean BAC of zero, approx. 0.5 and 0.9 g/L on the three different test days. The Widmark equation was used to estimate the ethanol doses, assuming a total body water to total body mass ratio of 0.68, a bioavailability of 75%, and a metabolic rate for ethanol of 0.15 g/L per hour. The drinks were served in closed plastic containers, from which the participants were instructed to sip the drink through a straw. To avoid an obvious ethanol taste, no drinks were stronger than 10% (v/v) ethanol, and they were kept cold by the addition of ice. The participants were allowed 1 h to finish their drinks, after which they waited another 30 min before the driving test started, to allow for absorption of the administered ethanol. Placebo effectiveness was assessed by questionnaire. Fifty percent of the participants believed they had received an impairing drug (either alcohol or a sedating drug) under the placebo condition, whereas only 8% believed they were sober under ethanol conditions. The possible impact of placebo effectiveness, including the sham pill placebo, does not constitute a part of the current study.

In the following, ethanol levels are referred to as BAC 0, BAC 0.5 and BAC 0.9, respectively. Blood was sampled immediately before and after each driving session, and the mean value was used as the best estimate of the mean BAC during driving. Immediately after each drive, the subjects rated their degree of simulator sickness from 0 (very little) to 10 (very much) on a numerical scale, according to the following question: "To what extent did you experience simulator sickness during the driving test?"

2.3. Simulated driving

The simulator had the appearance of a normal car (Renault Scenic) with automatic transmission and original controls (Fig. 1). The driving scenario was depicted on screens covering 180° of the driver's forward field of vision and 90° of the rear field of vision, with synchronized displays in internal and external mirrors. The vertical field of view was 47° both to the front and to the rear. The simulator reproduced realistic motion, vibration and sound through a three-axis motion platform, a vibration system in the chassis and a four-channel sound system. Data on lateral position, speed, pedal use and steering wheel movements over the entire duration of the test sessions were extracted directly from the simulator computer and logged 20 times per second. The participants drove on average 34 laps during each test, corresponding to 46.8 km.

The nighttime driving scenario consisted of a narrow, hilly and curvy road circuit that was 1.37 km long and closely resembled a typical rural two-lane Norwegian road, with midline and side markings. Traffic lights present in two locations turned red on one occasion during each trip. Two sudden incidents (a car abruptly entering the road and a pedestrian crossing the road in front of the driver) each occurred once towards the end of the driving session. Apart from this, there was no other traffic. The participants were instructed to keep in the middle of the lane, adjust speed according to the driving conditions and otherwise drive as they would normally have done.

2.4. Measurements

The following measures of driving behavior were obtained: standard deviation of lateral position (SDLP), number of brake pedal pressures per lap, number of accelerator pedal pressures per lap, steering wheel movement speed, steering wheel movement per distance driven, steering wheel reversals per distance driven, steering wheel reversal frequency, average speed, and standard deviation of speed. Collisions at the potential crash events were also recorded. The measurements were chosen to cover important behavioral levels of driving (Michon 1985; Walsh et al., 2008), although they do not represent a full range of skills necessary for safe driving.

Blood ethanol concentrations were quantified using a headspace gas chromatography-mass spectrometry (GC–MS) method as previously described (Helland et al., 2013).

2.5. Statistical analyses

Sample size estimates were based on a pilot study measuring SDLP in the simulator, and performed to determine the appropriate sample size in a validation study of the simulator designed to compare driving performance in real and simulated driving at dif-

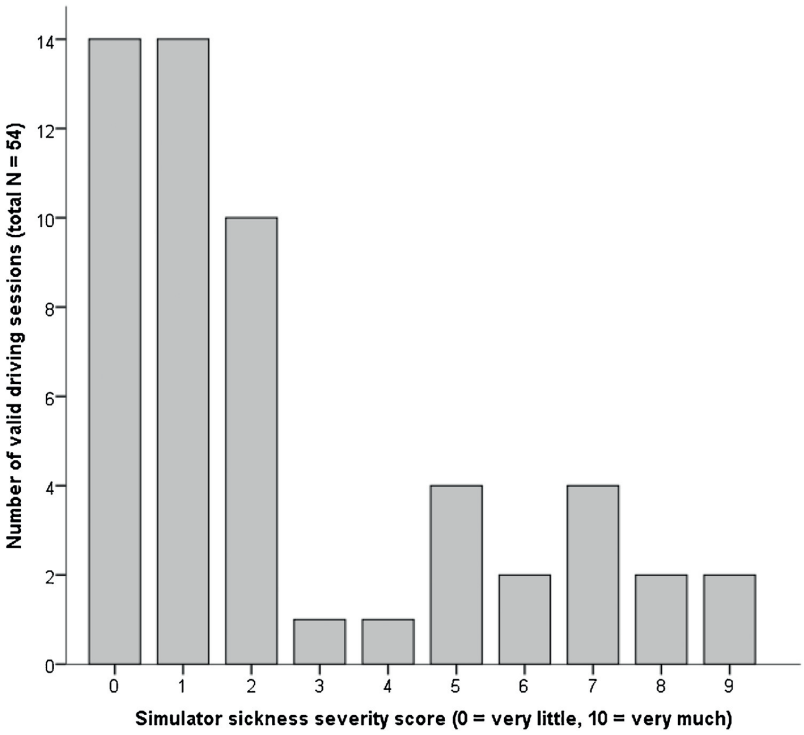

Fig. 2. Frequency distribution of simulator sickness scores in the 54 completed driving sessions in the study.

ferent BAC. We did not perform separate sample size estimations for the assessment of simulator sickness. To investigate the relation between simulator sickness and driving outcomes, we used linear mixed model analyses with the driving outcomes as dependent variables, simulator sickness severity (SSS) and measured BAC as covariates, and participant as random effect. Where a significant main effect of SSS was shown, we also added the interaction term SSS × BAC to the model, and separate plots were made to explore the effects of SSS at different intended BAC.

Linear mixed model analyses were also used to explore the relation between BAC and SSS, and repeated exposures to the simulator and SSS. In these analyses, the dependent variable SSS is not normally distributed. Normality of residuals was judged by visual inspection of QQ-plots. We therefore performed analyses with log-transformed simulator sickness scores. Since the SSS scale included zero, we added one to all scores before log-transforming the variable.

For the mixed models, we report R squared values computed as the proportional reduction in the estimated total residual variance comparing the null model without covariates with the model with covariates (Rabe-Hesketh and Skrondal, 2012).

Two-sided p-values < 0.05 were considered significant. The analyses were performed in SPSS 21 and Stata 12.

3. Results

Eighteen out of 20 participants completed three sessions in the driving simulator. One subject withdrew from simulator testing because of intolerable nausea during his first test drive, the other because of a surgical procedure unrelated to the study. These subjects were excluded from the analyses. Five subjects interrupted their first simulator session because of simulator sickness, but were re-tested and eventually completed all three sessions. These subjects were excluded from the analysis of the relationship between previous exposures to the simulator and simulator sickness severity, as their data would not be comparable to the rest. Otherwise, data from all valid sessions were included in the final analyses.

Overall, in the 54 completed driving sessions, the mean and median simulator sickness score was 2.5 and 1, respectively, with a standard deviation of 2.7 and a range of 0–9. The distribution was highly skewed, with a majority of driving sessions scored 0 or 1 (Fig. 2).

The mean blood alcohol concentrations achieved were generally slightly lower than the intended levels (0.38 g/L and 0.82 g/L at the intended BAC 0.5 and BAC 0.9 levels, respectively). In the statistical analyses, the actual BAC measured at each driving session was used.

3.1. Simulator sickness effects on measures of driving impairment

The results from linear mixed model analyses are presented in Table 1. The severity of simulator sickness significantly predicted lower values of the dependent variables steering wheel reversal frequency and average speed. The effect estimates predicted from the regression model correspond to an expected reduction in steering wheel reversal frequency and average speed of 23% and 18%, respectively, at a maximum simulator sickness score of 10. For the other outcomes, there were no statistically significant effects, nor even trends towards significance, of simulator sickness. We recorded no collisions at the potential crash events.

Additional mixed model analyses that allowed for possible interaction between BAC and simulator sickness showed that there was a statistically significant interaction in the case of steering

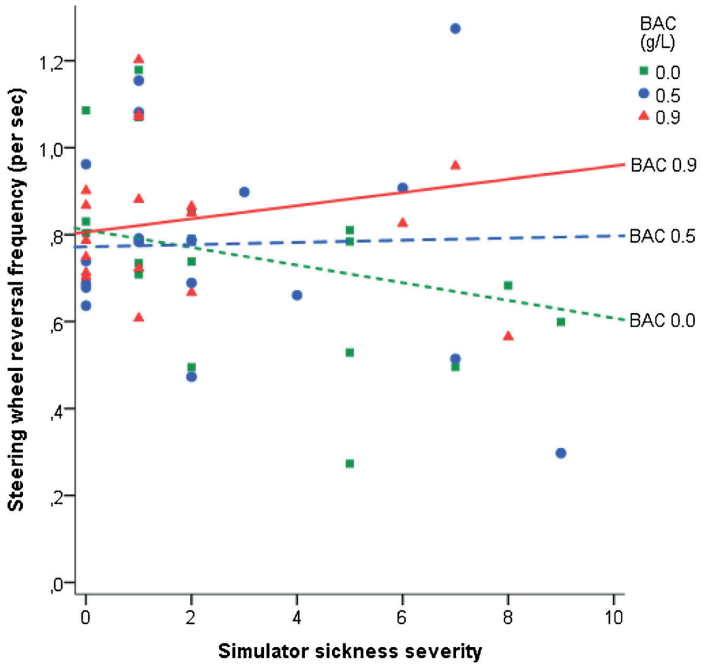

Fig. 3. Relationship between simulator sickness severity and steering wheel reversal frequency at different blood alcohol concentrations.

Table 1
Linear mixed model analyses with driving measures as dependent variable and blood alcohol concentration (BAC; g/L) and severity of simulator sickness (SSS; 0–10) as covariates. Significant associations between SSS and driving measures in bold.

Dependent variable: Driving measure (unit)	Intercept	R^2	Covariate	Regression coefficient	95% CI		p value
					Lower	Upper	
SDLP (cm)	30	0.10	BAC	13	6.6	19	<0.001
			SSS	−0.34	−1.4	0.76	0.54
Steering wheel movement speed (rad./sec)	0.50	0.15	BAC	0.22	0.15	0.30	<0.001
			SSS	−0.0028	−0.016	0.011	0.67
Steering wheel movement per meter (m^{-1})	0.042	0.19	BAC	0.012	0.0079	0.017	<0.001
			SSS	0.00016	−0.00057	0.00089	0.66
Steering wheel reversal frequency[a] (sec^{-1})	0.80	0.10	BAC	0.0070	−0.087	0.10	0.88
			SSS	**−0.018**	**−0.033**	**−0.0025**	**0.024**
			BAC × SSS[a]	**0.041**	**0.011**	**0.071**	**0.009**
Average speed[a] (km/h)	51	0.18	BAC	3	−1.2	7.2	0.16
			SSS	**−0.9**	**−1.6**	**−0.23**	**0.010**
			BAC × SSS[a]	1.1	−0.28	2.4	0.12
Standard deviation of speed (km/h)	4.9	0.10	BAC	1.1	0.48	1.7	0.014
			SSS	−0.028	−0.13	0.076	0.59
Brake pedal pressures (km^{-1})	4.7	0.044	BAC	1.4	0.58	2.22	0.011
			SSS	0.10	−0.042	0.24	0.16
Accelerator pedal pressures (km^{-1})	7.9	0.15	BAC	2.4	1.5	3.3	<0.001
			SSS	−0.063	−0.22	0.095	0.42

[a] For the outcomes showing a statistically significant association with SSS, mixed model analyses allowing for interaction between BAC and SSS are also reported.

wheel reversal frequency, and no such interaction in the case of average speed.

In order to explore the differential effects of BAC and simulator sickness further, we investigated the effect of simulator sickness on steering wheel reversal frequency in the intended BAC groups of zero, 0.5 and 0.9 (Fig. 3). There is a negative association between severity of simulator sickness and steering wheel reversal frequency for the BAC 0 group ($p = 0.027$), whereas this is not the case for the BAC 0.5 and BAC 0.9 groups. Likewise, the negative association between sickness severity and average speed is most pronounced in the BAC 0 group ($p = 0.036$), even though the interaction between BAC and SSS did not reach statistical significance for this parameter.

3.2. Effect of BAC on simulator sickness

There is a statistically significant, negative relationship between BAC and the degree of simulator sickness ($p = 0.049$, log-transformed SSS; $R^2 = 0.054$). The regression analysis shows an expected effect of approximately 1.6 points lower simulator sickness score at a BAC of 1 g/L compared with sober driving (Fig. 4).

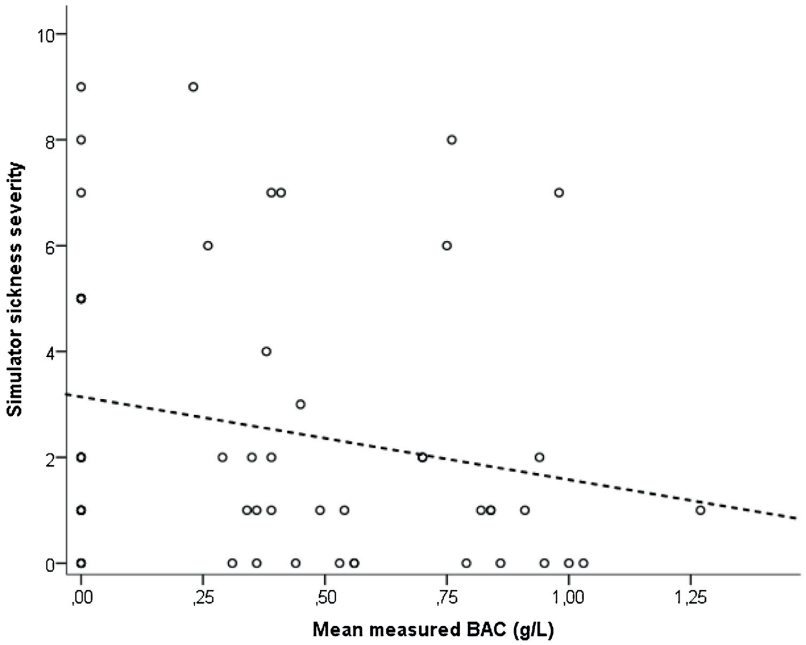

Fig. 4. Relationship between BAC and the degree of simulator sickness, with fitted linear regression line from mixed model analysis.

3.3. *Effect of repeated exposures to the simulator on simulator sickness*

The mean simulator sickness score in the 13 subjects who did not interrupt any driving sessions was 3.4, 1.8 and 1.5 in the first, second and third driving session, respectively. There was a larger spread, with more subjects scoring high on simulator sickness, in the first round (Fig. 5). The negative relation between the number of exposures and the degree of simulator sickness is not statistically significant ($p = 0.23$, log-transformed SSS). The five subjects that were excluded from this analysis due to interruption of their first driving session generally had higher sickness scores (overall mean 3.2) in their three completed sessions than the other participants (overall mean 2.2).

4. Discussion

This study demonstrates that there is no significant influence of simulator sickness on the important driving impairment measure SDLP, in a curvy and hilly rural road scenario of long duration. Nor is there any significant interaction between BAC and simulator sickness for SDLP. This strengthens the notion that SDLP is a robust parameter of drug related driving impairment in simulator studies (Mets et al., 2011; Helland et al., 2013; Helland et al., 2016). Our results are in accordance with the findings in another simulator study (Muttray et al., 2013), where simulator sickness was found not to influence lane keeping behavior; however, the participants of that study reported very low simulator sickness scores. Similarly, there were no significant relations between simulator sickness and several other measures of driving behavior in the simulator, such as standard deviation of speed, steering wheel movement measures, and brake or accelerator pedal pressures per distance driven. The driving simulator test has previously been shown to be sensitive to ethanol effects, showing strong BAC-related increments in SDLP as well as several other measurements of driving performance (Helland et al., 2013; Helland et al., 2016).

We found significant, negative associations between the severity of simulator sickness and the measures of average speed and steering wheel reversal frequency. The reduction in steering wheel reversal frequency to some degree may be a consequence of reduced speed and thus may not constitute an independent finding. For steering wheel reversal frequency, there appears to be an interaction between BAC and simulator sickness, so that the effects of simulator sickness are most pronounced when driving sober. For average speed, there is no statistically significant interaction between BAC and simulator sickness, yet the negative association between sickness severity and average speed is more pronounced in the BAC 0 group. Thus, simulator sickness may have a moderating effect on driving, leading to lower speeds and less steering wheel reversals, and ethanol seems to cancel this effect, at least in the case of steering wheel reversal frequency.

Our interpretation of these findings is that simulator sickness primarily causes the subjects to drive more slowly and avoid unnecessary steering wheel reversals in an attempt to ease symptoms. This may confound the interpretation of these measures as indicators of driving impairment. In a challenging scenario with many curves, and no specific speed instructions, average speed could arguably be regarded as an outcome with relevance to traffic safety, reflecting risk willingness and self-assessment at a strategic planning behavior level (Michon 1985; Fillmore et al., 2008). Average speed has been shown to be positively correlated to BAC in other driving simulator studies (Zhang et al., 2014), and was also reported by us in a previous paper (Helland et al., 2016). The observation that the effect of simulator sickness on average speed is most pronounced in sober subjects indicates that the observed increase in average speed with rising BAC may actually reflect a mitigating effect on simulator sickness, and not an effect of BAC on average speed *per se*. Hence, we believe that the apparent BAC-related increases in average speed and steering wheel reversal frequency

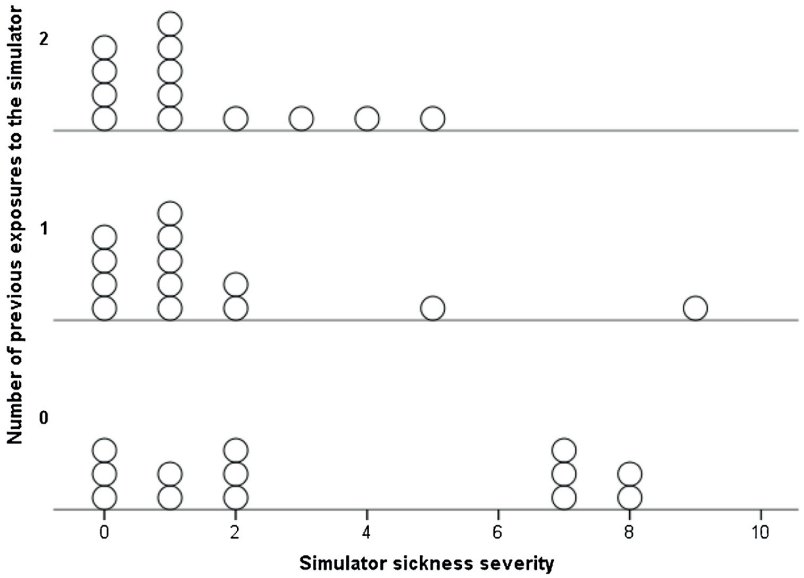

Fig. 5. Degree of simulator sickness according to the number of previous exposures to the simulator in the 13 subjects who did not interrupt any driving sessions.

described in our previous article (Helland et al., 2016) may in fact be a consequence of less simulator sickness at higher BACs. This suggests that simulator sickness may act as an operational confounder, leading to incorrect conclusions about BAC effects on driving performance. The finding that there is a significant, inverse relationship between simulator sickness and steering wheel reversal frequency in the placebo group but not in the ethanol groups supports this notion. The mechanism by which ethanol reduces the effect of simulator sickness on these measures of driving behavior is unknown. In the case of average speed, our findings should be interpreted with caution since the interaction between BAC and simulator sickness did not reach statistical significance.

There was a negative relationship between BAC and simulator sickness severity. Thus, ethanol seems to protect against simulator sickness to a certain extent. At a BAC of 1.0 g/L, the effect corresponds to a reduction of 1.6 points on the 0–10 numerical scale of simulator sickness severity that we used in our study. It is unlikely that the BAC-related reduction in simulator sickness will influence measures of driving ability significantly. However, the effect may be sufficient to cause lower dropout rates at higher BAC in studies using ethanol as a test substance, which could be a source of possible data bias. The mechanism by which ethanol reduces simulator sickness is not clear, and we are not aware of any previous studies that have reported such an effect. Ethanol may interfere with many different neurochemical systems and neuronal networks when present at the concentrations measured in our study (Spanagel 2009). One could speculate that several of these effects could change various sensory inputs, thereby decreasing the discrepancy between sensory and vestibular responses by which simulator sickness probably occurs (Brooks et al., 2010). It is unclear whether this phenomenon is unique to ethanol or may also be a feature of other centrally acting drugs.

Although not statistically significant, simulator sickness scores tended to decrease with repeated exposures to the simulator. This is in accordance with the findings in other studies that have shown attenuation of simulator sickness with repeated exposures (Kennedy et al., 2000; Domeyer et al., 2013). Since five subjects had to be excluded from the analyses due to interruption of their first driving test, the analysis only includes data from 13 participants. In addition, a few potential test subjects were also excluded before the study commenced due to excessive simulator sickness during screening. The exclusion of those most prone to simulator sickness may account for the low simulator sickness scores in the study, as well as the lack of significant decrease in simulator sickness with repeated exposures.

The driving scenario used in our study was designed to reflect conditions in which a disproportionately high number of ethanol- or drug-related accidents occur in Norway, i.e. nighttime driving on narrow, winding roads (Norwegian Public Roads Administration, 2013). It has been shown that driving scenarios with many curves and long duration are prone to provoke simulator sickness in test subjects (Classen et al., 2011). Another simulator study that employed a rural driving scenario found low ratings of simulator sickness, assessed with the Simulator Sickness Questionnaire (Muttray et al., 2013). Differences in driving scenario, duration, technical specification of the simulator and measuring methods may explain the discrepancies to our findings. This underlines the importance of thorough validation of the specific simulator scenario in use, as results cannot readily be extrapolated to other simulators and scenarios (Shechtman et al., 2009).

The most important weakness of our study is that we did not use the Simulator Sickness Questionnaire (SSQ), which is regarded as the gold standard to assess the severity of simulator sickness (Kennedy et al., 1993; Classen et al., 2011). Instead, the participants simply rated their perceived degree of simulator sickness on a numerical scale from zero to 10 immediately after each driving session. They were informed about the possible occurrence and symptoms of simulator sickness upon inclusion. We believe that this method, albeit simple, provided a valid assessment of simulator sickness experienced in the study, but cannot exclude that the overly simplistic measurement may have influenced the data. The SSQ is time consuming, which makes it challenging to fit into an experimental design. Furthermore, the validity of the SSQ in monotonous driving tests of long duration has been questioned, since items such as "fatigue" and "difficulty concentrating" included

in the SSQ are not specific to simulator sickness and may instead be due to sleepiness (Muttray et al., 2013).

Another major weakness is the lack of specific sample size estimations for assessing effects of simulator sickness. Consequently, we cannot exclude the possibility that the study had inadequate power to detect a real effect for some of the associations we explored.

Our study has some additional limitations. We employed a single blind design, thus we cannot exclude bias affecting data analysis. Also, the distinctive taste, smell and effects of ethanol make effective blinding difficult. All test subjects were healthy young male volunteers, who are not representative for the general driving population, and are probably less vulnerable to simulator sickness than older subjects (Kawano et al., 2012).The narrow inclusion criteria limits the generalizability of our findings. Further research in this field should include a broader sample, and measure simulator sickness at several time points during single exposures to the simulator, using validated (i.e. SSQ) and/or objective (i.e. eye fixation, blinking) measurements of simulator sickness.

5. Conclusions

In summary, simulator sickness is associated with a reduction of average speed and steering wheel reversal frequency. These changes seem to be less pronounced in subjects driving under the influence of alcohol, and may hamper the use of these measures as indicators of unsafe driving in simulator studies of drug impairment. On the other hand, simulator sickness is not associated with changes in SDLP. In the young, healthy, male recreational drinkers tested in the present study, simulator sickness scores decreased with higher BAC. To our knowledge, the present study is the first to quantify the impact of simulator sickness on SDLP and other measures with relevance to driving safety, and to explore the relationship between BAC and simulator sickness. Our findings lend further support to the robustness of SDLP as a measure of drug impaired driving in simulator studies. Driving simulator researchers should beware the risks of simulator sickness confounding the results and introducing bias, and take action to minimize its occurrence. The possibility of simulator sickness acting as an operational confounder should be borne in mind when investigating the effect of drugs on driving. In light of the exploratory nature of our study and its shortcomings with regard to sampling bias, power and measuring technique of simulator sickness severity, our findings should be regarded as tentative. Further studies are needed to confirm or disprove our findings, and extend the characterization of simulator sickness to other driving scenarios and measures of driving impairment.

Acknowledgments

The research project was funded by a grant from the Research Council of Norway.

The authors would like to thank the following employees at SINTEF Technology and Society and the Dept. of Clinical Pharmacology at St. Olav University Hospital who made valuable contributions to the research project: Dagfinn Moe, Kenneth Sørensen, Terje Giæver, Cato Mausethagen, Nina Husby, Marianne Nordtømme, Odd Hjelkrem, Guro Haltbrekken, Anita Skogholt, Trude Kristine Vaag Gjerde, and Andreas Austgulen Westin.

We would also like to thank the study subjects for their participation.

Finally, we would like to express our gratitude to Kristian Sakshaug, who participated in the study but regrettably fell ill and died before its completion.

References

Apfel, C.C., Greim, C.A., Haubitz, I., Goepfert, C., Usadel, J., Sefrin, P., Roewer, N., 1998. A risk score to predict the probability of postoperative vomiting in adults. Acta Anaesthesiol. Scand. 42 (5), 495–501.

Brooks, J.O., Goodenough, R.R., Crisler, M.C., Klein, N.D., Alley, R.L., Koon, B.L., Logan Jr., W.C., Ogle, J.H., Tyrrell, R.A., Wills, R.F., 2010. Simulator sickness during driving simulation studies. Accid. Anal. Prev. 42 (3), 788–796.

Classen, S., Brooks, J., 2014. Driving simulators for occupational therapy screening, assessment, and intervention. Occup. Ther. Health Care 28 (2), 154–162.

Classen, S., Bewernitz, M., Shechtman, O., 2011. Driving simulator sickness: an evidence-based review of the literature. Am. J. Occup. Ther. 65 (2), 179–188.

Cobb, S.V.G., Nichols, S., Ramsey, A., Wilson, J.R., 1999. Virtual reality-induced symptoms and effects (VRISE). Presence-Teleoperators Virtual Environ. 8 (2), 169–186.

Domeyer, J.E., Cassavaugh, N.D., Backs, R.W., 2013. The use of adaptation to reduce simulator sickness in driving assessment and research. Accid. Anal. Prev. 53, 127–132.

Fillmore, M.T., Blackburn, J.S., Harrison, E.L., 2008. Acute disinhibiting effects of alcohol as a factor in risky driving behavior. Drug Alcohol Depend. 95 (1–2), 97–106.

Galvez-Garcia, G., Hay, M., Gabaude, C., 2015. Alleviating simulator sickness with galvanic cutaneous stimulation. Hum. Factors 57 (4), 649–657.

Helland, A., Jenssen, G.D., Lervåg, L.E., Westin, A.A., Moen, T., Sakshaug, K., Lydersen, S., Mørland, J., Slørdal, L., 2013. Comparison of driving simulator performance with real driving after alcohol intake: a randomised, single blind, placebo-controlled, cross-over trial. Accid. Anal. Prev. 53, 9–16.

Helland, A., Jenssen, G.D., Lervåg, L.E., Moen, T., Sakshaug, K., Lydersen, S., Mørland, J., Slørdal, L., 2016. Evaluation of measures of impairment in real and simulated driving: results from a randomized, placebo-controlled study. Traffic Inj. Prev. 17 (3), 245–250.

Kawano, N., Iwamoto, K., Ebe, K., Aleksic, B., Noda, A., Umegaki, H., Kuzuya, M., Iidaka, T., Ozaki, N., 2012. Slower adaptation to driving simulator and simulator sickness in older adults. Aging Clin. Exp. Res. 24 (3), 285–289.

Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G., 1993. Simulator sickness questionnaire (SSQ): a new method for quantifying simulator sickness. Int. J. Aviat. Psychol. 3 (3), 203–220.

Kennedy, R.S., Stanney, K.M., Dunlap, W.P., 2000. Duration and exposure to virtual environments: sickness curves during and across sessions. Presence-Teleoperators Virtual Environ. 9 (5), 463–472.

Mets, M.A., Kuipers, E., De Senerpont Domis, L.M., Leenders, M., Olivier, B., Verster, J.C., 2011. Effects of alcohol on highway driving in the stisim driving simulator. Hum. Psychopharmacol. 26 (6), 434–439.

Michon, J.A., 1985. A critical view of driver behavior models: what do we know, what should we do? In: Evans, L., Schwing, R.C. (Eds.), Human Behavior and Traffic Safety. Plenum Press, New York, pp. 485–520.

Milleville-Pennel, I., Charron, C., 2015. Do mental workload and presence experienced when driving a real car predispose drivers to simulator sickness? An exploratory study. Accid. Anal. Prev. 74, 192–202.

Muttray, A., Breitinger, A., Goetze, E., Schnupp, T., Geissler, B., Kaufmann, T., Golz, M., Letzel, S., 2013. Further development of a commercial driving simulation for research in occupational medicine. Int. J. Occup. Med. Environ. Health 26 (6), 949–965.

Norwegian Public Roads Administration, 2013: STRAKS road accident database.

Rabe-Hesketh, S., Skrondal, A., 2012. Multilevel and Longitudinal Modeling Using STATA, 3rd ed. Stata Press Publication, College Station, Texas, pp. 135.

Shechtman, O., Classen, S., Awadzi, K., Mann, W., 2009. Comparison of driving errors between on-the-road and simulated driving assessment: a validation study. Traffic Inj. Prev. 10 (4), 379–385.

Spanagel, R., 2009. Alcoholism: a systems approach from molecular physiology to addictive behavior. Physiol. Rev. 89 (2), 649–705.

Walsh, J.M., Verstraete, A.G., Huestis, M.A., Mørland, J., 2008. Guidelines for research on drugged driving. Addiction 103 (8), 1258–1268.

Zhang, X., Zhao, X., Du, H., Ma, J., Rong, J., 2014. Effect of different breath alcohol concentrations on driving performance in horizontal curves. Accid. Anal. Prev. 72, 401–410.

Paper IV

Is not included due to copyright