From dusk till yawn

From dusk till yawn:

The assessment of driving performance in clinical populations at risk of drowsy driving at the group and individual level

Frederick R.J. Vinckenbosch

Cover design: Eva Lynen (illustration) & Lukas Liefsoens (formatting)

Printing: Nathing – Hasselt, Belgium

ISBN: 978-94-6469-113-9

Printing of this dissertation was supported by Maastricht University, the Netherlands

From dusk till yawn:

The assessment of driving performance in clinical populations at risk of drowsy driving at the group and individual level

DISSERTATION

to obtain the degree of doctor at Maastricht University on the authority of Rector Magnificus Prof. Dr. Pamela Habibović, in accordance with the decision of the Board of Deans, to be defended in public on Thursday 10 November 2022 at 16:00

by

Frederick Remy Judith Vinckenbosch

Promotor:

Prof. Dr. J.G. Ramaekers

Co-promotor:

Dr. A.Vermeeren

Assessment Committee:

Prof. Dr. W.J. Riedel (chair)

Prof. Dr. T. Brijs, Universiteit Hasselt, Instituut voor mobiliteit (IMOB), Hasselt, Belgium

Prof. Dr. L. Downey, Swinburne University of Technology, Melbourne, Australia

Dr. E.L. Theunissen

Dr. F. Smulders

The research and analyses presented in this dissertation were conducted at the department of neuropsychology and psychopharmacology, faculty of psychology and neuroscience, Maastricht University, Maastricht, the Netherlands. The studies described in chapters 3 and 4 were sponsored by Jazz Pharmaceuticals, plc. Data collection for the analysis presented in chapter 5 was performed at the University of Groningen and Utrecht University, the Netherlands, in addition to Maastricht University. Funding for this study was provided by the Dutch ministry of infrastructure and water management. The studies underlying the analysis in chapter 6 were performed at Maastricht University and Utrecht university. Funding was obtained from various funding bodies. For more details, the reader is referred to the respective articles cited in chapter 6.

Table of contents

CHAPTER 1: General Introduction

Relevance and responsibilities of traffic medicine

Operating a motor vehicle in traffic is a complex task that requires the simultaneous execution of multiple subtasks, ranging from conceptually simple operations, such as road tracking, to more complex tasks, such as anticipating maneuvers by other road users (Keskinen, 1994; Michon, 1985). In turn, performance of these tasks depends on an even wider range of perceptual, cognitive, and motor functions (Food and Drug Administration, 2017). These numerous functions are performed by an equally great number of physiological systems, each of which constitutes a potential weak link in the chain of processes that underlie safe driving behavior.

Considering the complex symphony of physiological systems that is required to carry out the challenging task that driving is, it is not surprising that many medical conditions are known to negatively affect driving performance and subsequently increase traffic accident risk. These conditions include, but are not limited to impairments of the visual systems (Owsley & McGwin Jr, 2010; Rubin et al., 2007), mood disorders (Bulmash et al., 2006; van der Sluiszen et al., 2017), attentional disorders (Curry et al., 2017; Woodward, Fergusson, & Horwood, 2000), sleep-wake disorders (Ellen et al., 2006; Pizza et al., 2015), etc. These disorders either affect the ability of the driver to perceive and interpret and/or react appropriately to the traffic environment.

In addition to medical conditions, there are also many pharmaceutical drugs known to impair driving performance as a side effect. The International Council On Alcohol Drugs and Traffic Safety (ICADTS) provides a list containing approximately four hundred medicinal drugs that are considered to have a moderate or severe negative effect on driving performance (Alvarez, de Gier, & Verstraete, 2006). In general, these drugs directly or indirectly suppress central nervous system (CNS) activation, either as the intended therapeutic effect or as a side effect. Indeed, many of the medications listed as impairing driving performance by ICADTS have been associated with increased traffic accident risk

and impaired driving performance (Orriols et al., 2009; Ramaekers, 1998; Ravera, van Rein, De Gier, & de Jong-van den Berg, 2011).

The potential of medical conditions and their pharmaceutical treatments to impair driving performance is an important topic in patient care and policymaking. In the scope of traffic safety, it is paramount to keep impaired drivers of the roads as much as possible. However, losing the privilege to drive often poses a great limitation on independence and quality of life (Gilliam et al., 1997; Marottoli et al., 1997; Musselwhite & Haddad, 2010; Owsley, 2002; Timmermans et al., 2018). Unlike the use of alcohol or illicit drugs before or during driving, suffering from a medical condition or using pharmaceuticals for the treatment thereof is not a free choice by the driver. Researchers, clinicians, and policymakers carry the ethical responsibility of generating insights, advice and regulations that fend impaired drivers off the roads as much as possible, while also preventing the unjustified denial of patients' right to drive.

Next to generating insights, advice, and fair policies, it is ethically imperative to invest efforts into finding ways to re-enable impaired drivers to safely operate a motor vehicle in traffic. Various pharmacological treatments have been developed, albeit not always specifically, for this purpose. Prominent examples are antiepileptics for the prevention of epileptic seizures (Orriols et al., 2013) and the prescription of stimulants for the treatment of excessive daytime sleepiness as a result of certain sleep-wake disorders (Lin et al., 2020). It is clear that a thorough evaluation of a treatments' effectiveness to attenuate driving impairment is warranted to safeguard traffic safety, but also to provide the necessary evidence to support the resumption of driving after treatment initiation, and under which conditions.

The need for individual driving assessment

Clinical advice and legal policies regarding a patient's ability to drive are often based on empirical findings from epidemiological and cross-sectional studies, and clinical trials. However, findings from such studies arguably form an insufficient basis for formulating advice and policies on

this matter. This is because these types of studies are aimed at providing general insights that pertain to the average member of the respective clinical population. However, clinical populations are often inherently heterogeneous regarding clinical manifestation and symptom severity, as well as with respect to medication use (e.g. frequency, dose, co-medications, etc.) and response. Hence, group-level findings will often not apply to a significant part of the respective clinical population.

Nevertheless, the insights yielded by clinical trials, epidemiological and cross-sectional studies are an indispensable part of providing a proof-of-principle that a certain medical condition or medication is linked to impaired driving performance, or that a treatment can successfully counteract driving impairments. These study types can also help narrow down which members of a clinical population are most likely to be affected negatively by medical conditions or to respond (un)favorably to initiated treatments. However, it is argued that the identification of vulnerable (sub)populations would ideally be complemented with individual assessment of driving performance in order to maximize sensitivity to detect impaired drivers while also minimizing the risk for false positive labelling of unaffected drivers as being impaired. Hence, it is argued that there is an unequivocal need for sensitive and specific methods to identify individual drivers who as a result of their medical condition or prescribed medications experience difficulties to safely operate a motor vehicle in traffic, as well as for the evaluation of treatments and interventions designed to counteract these impairments.

Advancing the standardized on-the-road driving test for individual driving assessment

For over 30 years, the standardized on-the-road driving test has been the backbone of driving performance assessment at Maastricht University (O'Hanlon, 1984; Ramaekers, 2017). During this test, research participants drive a specially equipped test vehicle at a steady velocity of 95km/h on a public motorway under the supervision of a licensed driving instructor. During the 100 km drive, the vehicle's distance to the lane demarcation on its left, i.e. the lateral position, is continuously

measured. The variability of the lateral position throughout the drive is summarized in the primary outcome measure known as *the standard deviation of the lateral position* (SDLP). Increased SDLP values are assumed to reflect poorer road-tracking performance and vice versa (Ramaekers, 2017; Verster & Roth, 2011). The SDLP has proven itself to be a sensitive measure for detecting the impairing effects of various CNS depressing drugs such as benzodiazepines, tricyclic antidepressants, antihistamines, alcohol, cannabis, as well as the negative impact of sleepiness and mental load on driving performance (Arkell et al., 2020; Jongen, Perrier, Vuurman, Ramaekers, & Vermeeren, 2015; Jongen, van der Sluiszen, Brown, & Vuurman, 2018; Jongen et al., 2017; Leufkens, Ramaekers, de Weerd, Riedel, & Vermeeren, 2009; Ramaekers, 2003; van der Sluiszen et al., 2016; Vermeeren, 2004).

The prolonged duration and monotonous nature of the standardized on-the-road driving test, combined with its ecological validity, sensitivity and reliability makes it arguably one of the superior methods for the detection of driving impairment due to decrements of sustained attention and/or alertness. However, this only relates to its application in crossover studies, where each participant serves as their own reference. Although the SDLP is a very reliable parameter within any given individual, the normative range of SDLP values is very wide (Verster & Roth, 2011). A statistical consequence of this is that when the SDLP is applied in studies where a crossover design cannot be implemented, the required sample size increases drastically.

However, a more fundamental consequence of the wide range of normative SDLP values is that the applicability for the identification of individual drivers who exhibit road-tracking impairment is very limited. A representative SDLP for one healthy and capable driver might be indicative of road-tracking impairment for the other (Ramaekers, 2017; Verster & Roth, 2011). Also, only a minority of impaired drivers will exhibit SDLP values that lie outside the normative range (Ramaekers, 2017; Verster & Roth, 2011). Hence, the SDLP has little informative value for judging an individual drivers' road tracking performance.

Although some information on individual performance can still be derived when the comparison to an individual reference or baseline SDLP value can be made, as will be proposed in chapter 7 of this dissertation, it is clear that the SDLP alone is not ideal for individual assessments of road-tracking performance. However, as argued above, there is a clear need for assessment techniques that can sensitively and specifically identify impaired drivers at the individual level. In order to exploit the ecological validity of the standardized on-the-road driving test for this purpose, it is argued that alternative outcomes and applications of the lateral position signal are needed.

Purpose and outline of this thesis

The purpose of this dissertation is to present studies that apply the standardized on-the-road driving test in clinical samples, as well as exploring alternative ways of utilizing the outcomes of the on-the-road driving test for the identification of individual drivers who experience driving impairment.

Chapter 2 provides a general overview of the research methods applied in the field of traffic medicine. This chapter was written to be part of *Handbook of Forensic Medicine*, *2nd edition.* Hence, the chapter is directed toward forensic experts. Nevertheless, in essence it is an introduction to the field of traffic medicine, to which this dissertation belongs. The chapter provides an overview of the common study types in the field of traffic medicine, as well as the range of available tools for the assessment of driving performance, including a more detailed explanation of the on-the-road driving test.

In **chapter 3**, an experimental clinical trial is presented. In the presented study, the effects of the stimulant drug *solriamfetol* on road-tracking performance in patients suffering from narcolepsy is investigated. Narcolepsy is a neurological condition of which one hallmark feature is difficulty with maintaining wakefulness. It was investigated whether solriamfetol could assist narcolepsy patients in remaining attentive and awake during the standardized on-the-road driving test.

In **chapter 4**, a similar study on the effects of solriamfetol is presented. In this study, solriamfetol's potential to assist sleep apnea patients in remaining wakeful and attentive during driving is investigated. Sleep apnea is a sleep-breathing disorder which causes fragmented sleep and therefore impedes a restorative night. Sleep apnea patients often feel tired and have difficulties with maintaining wakefulness throughout the day.

Chapter 5 presents an explorative analysis of the outcomes of a previous study on the effects of long-term benzodiazepine (CNS depressants) use on cognitive functioning and driving performance. This cross-sectional study illustrates the methodological difficulties that are introduced by the heterogeneity regarding symptomatology and medication use that is inherent to clinical populations. The analysis attempts to answer the question whether long-term benzodiazepine users exhibit cognitive or driving impairment relative to healthy controls, taking into account the heterogeneity in medication use and clinical complaints.

Chapter 6 investigates the validity of *lane* drifts as an additional outcome measure of the on-the-road driving test. Lane drifts are large and abrupt lateral displacements of the test vehicle. The authors proposing this new outcome measure conceived these events as indicative of momentary lapses of attention. Indeed, signatures of attentional fluctuations in the lateral position signal might provide a useful outcome measure for identifying driving impairment at the individual level. However, it was concluded that the proposed operationalization did not have any added value as an additional outcome measures of the standardized on-the-road driving test. Nevertheless, the idea is arguably deserving of further consideration in the form of alternative operationalization. At the time of designing the studies presented in chapter 3 and 4, the outcome measure had been recently proposed and was therefore included as an outcome measure in both studies, despite the later negating findings presented in chapter 6.

In **chapter 7**, two statistical methods for identifying drivers with significantly improved or impaired road-tracking performance during a standardized on-the-road driving test are presented. The methods exploit the high test-retest reliability of SDLP to determine which drivers demonstrate

statistically significant in- or decreased SDLP compared to their own baseline SDLP. A limitation of the proposed methods is the requirement for a baseline SDLP measurement. Therefore, the methods cannot be applied directly for the identification of impaired drivers based on a single SDLP measurement. Nevertheless, it is argued in the general discussion (Chapter 8) that the identification of impaired drivers in experimental crossover trials provides opportunities to derive sensitive and specific signatures of driving impairment in the lateral position signal or alternative outcome measures that in turn can be applied for the identification of impaired drivers based on a single test-drive.

References

Alvarez, J., de Gier, H., & Verstraete, A. (2006). Categorization system for medicinal drugs affecting driving performance ICADTS Working Group.

Arkell, T. R., Vinckenbosch, F., Kevin, R. C., Theunissen, E. L., McGregor, I. S., & Ramaekers, J. G. (2020). Effect of cannabidiol and Δ9-tetrahydrocannabinol on driving performance: a randomized clinical trial. *JAMA, 324*(21), 2177-2186.

Bulmash, E. L., Moller, H. J., Kayumov, L., Shen, J., Wang, X., & Shapiro, C. M. (2006). Psychomotor disturbance in depression: assessment using a driving simulator paradigm. *Journal of affective disorders, 93*(1-3), 213-218.

Curry, A. E., Metzger, K. B., Pfeiffer, M. R., Elliott, M. R., Winston, F. K., & Power, T. J. (2017). Motor vehicle crash risk among adolescents and young adults with attention-deficit/hyperactivity disorder. *JAMA pediatrics, 171*(8), 756-763.

Ellen, R., Marshall, S. C., Palayew, M., Molnar, F. J., Wilson, K. G., & Man-Son-Hing, M. (2006). Systematic review of motor vehicle crash risk in persons with sleep apnea. *J Clin Sleep Med, 2*(2), 193-200.

Food and Drug Administration. (2017). Evaluating Drug Effects on the Ability to Operate a Motor Vehicle, Guidance for Industry. 2017.

Gilliam, F., Kuzniecky, R., Faught, E., Black, L., Carpenter, G., & Schrodt, R. (1997). Patient-validated content of epilepsy-specific quality-of-life measurement. *Epilepsia, 38*(2), 233-236.

Jongen, S., Perrier, J., Vuurman, E., Ramaekers, J., & Vermeeren, A. (2015). Sensitivity and validity of psychometric tests for assessing driving impairment: effects of sleep deprivation. *PLoS One, 10*(2), e0117045.

Jongen, S., van der Sluiszen, N. N., Brown, D., & Vuurman, E. F. (2018). Single-and dual-task performance during on-the-road driving at a low and moderate dose of alcohol: A comparison between young novice and more experienced drivers. *Human Psychopharmacology: Clinical and Experimental, 33*(3), e2661.

Jongen, S., Vermeeren, A., van der Sluiszen, N., Schumacher, M., Theunissen, E., Kuypers, K., . . . Ramaekers, J. G. (2017). A pooled analysis of on-the-road highway driving studies in actual traffic measuring standard deviation of lateral position (ie,“weaving”) while driving at a blood alcohol concentration of 0.5 g/L. *Psychopharmacology, 234*(5), 837-844.

Keskinen, E. (1994). Why do young drivers have more accidents? Junge Fahrer und Fahrerinnen. *Referate der Ersten Interdiziplinaren Fachkonferenz, December 12-14, 1994 in Koln*.

Leufkens, T., Ramaekers, J., de Weerd, A., Riedel, W., & Vermeeren, A. (2009). On-the-road driving performance and driving related skills in untreated insomnia patients and chronic users of hypnotics. *HYPNOTICS AND ANXIOLYTICS*, 97.

Lin, Y.-C., Chen, T.-Y., Chien, W.-C., Chung, C.-H., Chang, H.-A., Kao, Y.-C., . . . Tzeng, N.-S. (2020). Stimulants associated with reduced risk of hospitalization for motor vehicle accident injury in patients with obstructive sleep apnea-a nationwide cohort study. *BMC pulmonary medicine, 20*(1), 1-11.

Marottoli, R. A., De Leon, C. F. M., Glass, T. A., Williams, C. S., Cooney Jr, L. M., Berkman, L. F., & Tinetti, M. E. (1997). Driving cessation and increased depressive symptoms: Prospective evidence from the New Haven EPESE. *Journal of the American Geriatrics Society, 45*(2), 202-206.

Michon, J. A. (1985). A critical view of driver behavior models: what do we know, what should we do? *Human behavior and traffic safety* (pp. 485-524): Springer.

Musselwhite, C., & Haddad, H. (2010). Mobility, accessibility and quality of later life. *Quality in Ageing and Older Adults*.

O'Hanlon, J. (1984). Driving performance under the influence of drugs: rationale for, and application of, a new test. *British journal of clinical pharmacology, 18*(S1), 121S-129S.

Orriols, L., Foubert-Samier, A., Gadegbeku, B., Delorme, B., Tricotel, A., Philip, P., . . . Group, C. R. (2013). Prescription of antiepileptics and the risk of road traffic crash. *The Journal of Clinical Pharmacology, 53*(3), 339-344.

Orriols, L., Salmi, L. R., Philip, P., Moore, N., Delorme, B., Castot, A., & Lagarde, E. (2009). The impact of medicinal drugs on traffic safety: a systematic review of epidemiological studies. *Pharmacoepidemiology and drug safety, 18*(8), 647-658.

Owsley, C. (2002). Driving mobility, older adults, and quality of life. *Gerontechnology, 2*(1).

Owsley, C., & McGwin Jr, G. (2010). Vision and driving. *Vision research, 50*(23), 2348-2361.

Pizza, F., Jaussent, I., Lopez, R., Pesenti, C., Plazzi, G., Drouot, X., . . . Dauvilliers, Y. (2015). Car crashes and central disorders of hypersomnolence: a French study. *PLoS One, 10*(6), e0129386.

Ramaekers, J. G. (1998). Behavioural toxicity of medicinal drugs. *Drug safety, 18*(3), 189-208.

Ramaekers, J. G. (2003). Antidepressants and driver impairment: empirical evidence from a standard on-the-road test. *Journal of Clinical Psychiatry, 64*(1), 20-29.

Ramaekers, J. G. (2017). Drugs and driving research in medicinal drug development. *Trends in pharmacological sciences, 38*(4), 319-321.

Ravera, S., van Rein, N., De Gier, J. J., & de Jong-van den Berg, L. T. (2011). Road traffic accidents and psychotropic medication use in the Netherlands: a case–control study. *British journal of clinical pharmacology, 72*(3), 505-513.

Rubin, G. S., Ng, E. S., Bandeen-Roche, K., Keyl, P. M., Freeman, E. E., & West, S. K. (2007). A prospective, population-based study of the role of visual impairment in motor vehicle crashes among older drivers: the SEE study. *Investigative ophthalmology & visual science, 48*(4), 1483-1491.

Timmermans, I., Jongejan, N., Meine, M., Doevendans, P., Tuinenburg, A., & Versteeg, H. (2018). Decreased quality of life due to driving restrictions after cardioverter defibrillator implantation. *Journal of Cardiovascular Nursing, 33*(5), 474-480.

van der Sluiszen, N. N., Vermeeren, A., Jongen, S., Theunissen, E., van Oers, A., Van Leeuwen, C., . . . Ramaekers, J. (2016). On-the-road driving performance after use of the antihistamines mequitazine and l-mequitazine, alone and with alcohol. *Psychopharmacology, 233*(18), 3461-3469.

van der Sluiszen, N. N., Wingen, M., Vermeeren, A., Vinckenbosch, F., Jongen, S., & Ramaekers, J. G. (2017). Driving performance of depressed patients who are untreated or receive long-term antidepressant (SSRI/SNRI) treatment. *Pharmacopsychiatry, 50*(05), 182-188.

Vermeeren, A. (2004). Residual effects of hypnotics. *CNS drugs, 18*(5), 297-328.

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.

Woodward, L. J., Fergusson, D. M., & Horwood, L. J. (2000). Driving outcomes of young people with attentional difficulties in adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(5), 627-634.

Chapter 2: Driving Aptitude and Fitness to Drive

Abstract

The current chapter was written to part of *Handbook of Forensic Medicine.* It aims to introduce forensic experts and students to common empirical methods for the identification of groups of drivers who have a chronically (driving aptitude) or transiently (driver fitness) increased risk of being involved in a road traffic accident. To this end, epidemiological methods for studying differential accident involvement and experimental methods for the assessment of driving performance are discussed. Next, focus is shifted from the group level to the individual level. General considerations regarding the assessment of individual driving aptitude and fitness to drive are discussed.

Vinckenbosch, F. R. J., Vermeeren, A., Theunissen E.L. & Ramaekers, J. G. (2022). Driving aptitude and fitness to drive. *Handbook of forensic medicine 2nd edition.*

Introduction

Relevance of traffic medicine

The number of annual road traffic deaths is steadily increasing. While the World Health Organization reported an annual road traffic death toll of 1.25 million in 2015 (World Health Organization, 2015), which seemed to have leveled off since 2007, in 2018 an annual road traffic death toll of 1.35 million was reported (World Health Organization, 2018). The increase in road traffic deaths is partly attributable to the growing number of motor vehicles that frequent the roads. Corrected for this growing fleet of motor vehicles, an encouraging steady decrease of the relative death toll can be observed. Nevertheless, each fatality is a tragedy for those involved and their social network, especially considering it is the leading cause of death in 5–29 year olds. In addition to fatalities, around 50 million people get injured in road traffic accidents each year (World Health Organization, 2018). Taken together, road traffic injuries and deaths account for about one-third of the disability adjusted life years due to accidental injuries, and therefore put a significant burden on society as a whole (Haagsma et al., 2016).

Curbing these numbers starts with eliminating the causal factors of road traffic accidents. Human error lies at the basis of the vast majority of cases (Singh, Kushwaha, Agarwal, & Sandhu, 2016). To minimize the likelihood of human error, it is necessary to identify causal factors during driving and to be able to make informed decisions about an individual drivers' ability to adequately and responsibly operate a motor vehicle. This is the domain of traffic medicine, a scientific discipline that aims to enhance human safety in traffic through research, and the practical application of medical findings and experiences. Provided the significant burden of road traffic accidents and the increasing number of motor vehicles on the roads, traffic medicine will remain an important discipline until technological advances eliminate the need for a human operator, which is conceivably still many years away.

Traffic medicine and forensic medicine

Traffic medicine and forensic medicine are closely related disciplines that overlap a great deal. Traffic medicine is mainly concerned with driving capability, the ability to safely operate a motor vehicle. Driving capability can be thought of as the sum of *driving aptitude, fitness to drive, and driving skills* (Berghaus & Schnabel, 2014; Staak, Hobi, & Berghaus, 1988). *Driving aptitude* is determined by physical and psychological traits and functions that are stable over time. It is dependent on driver dispositions such as personality traits and psychomotor functioning and can be affected by chronic medical conditions such as epilepsy, sleep apnea, narcolepsy, uncorrectable visual impairments, certain personality disorders, and many more. In traffic medicine, the assessment of driving aptitude is crucial for advising patients on whether to drive or determine circumstances under which driving is possible or discouraged. Unlike driving aptitude, *fitness to drive* is not stable over time. It is defined as the situation- and time-dependent ability to drive a vehicle. Fitness to drive can be influenced by transient factors such as fatigue or the occasional use of psychoactive substances. For traffic medicine, fitness to drive is vital to consider for advising patients on when to avoid operating a vehicle and for how long, e.g. when prescribing sedating medication. Lastly, *driving skills* are the operations required to operate a motor vehicle successfully and can be improved through training and experience.

Whereas traffic medicine is mainly concerned with driving capability, forensic medicine is concerned with driver culpability. It builds on the methods and insights of traffic medicine regarding driving capability to help frame legislation and devise methods that aid in enforcing this legislation. An obvious example is the determination of accident culpability of drivers who tested positive for alcohol or other psychoactive substances in biological samples (breath, urine, saliva, or blood). In addition, forensic medicine experts often serve as expert witnesses in court to help answer the question of culpability and risk of recidivism, which is important for determining penalties. In this capacity, the forensic expert is faced with questions regarding driver aptitude, e.g. advising on whether to suspend a driver's license of an older driver, and fitness to drive, e.g. determining the likelihood that a driver was significantly intoxicated at a given time. Thus, the forensic expert will

translate the findings of traffic medicine into practice where the circumstances often permit little room for error. Therefore, it is important that the forensic expert is familiar with the insights and methods of traffic medicine in order to formulate appropriate advice.

Outline and purpose of the chapter

The purpose of this chapter is to familiarize the reader with different empirical methods for the identification of groups of drivers who have a chronically (driving aptitude) or transiently (driver fitness) increased risk of being involved in a road traffic accident. To this end, epidemiological methods for studying differential accident involvement and experimental methods for the assessment of driving performance are discussed. Next, focus is shifted from the group level to the individual level. General considerations regarding the assessment of individual driving aptitude and fitness to drive are discussed.

Empirical methods for identifying risk factors for road traffic accidents

Epidemiological research: differential accident involvement

In practical terms, traffic and forensic medicine are mainly concerned with identifying drivers at increased risk of being involved in traffic accidents. During most of the previous century, the main method for identifying factors associated with increased risk of traffic accident involvement has been the study of differential accident involvement (Ranney, 1994). This approach investigates which characteristics and factors are more prevalent or pronounced in drivers involved in a road traffic accident. Because of obvious ethical objections, the study of differential accident involvement cannot rely on experimental study designs. The approach is therefore limited to epidemiological research methods.

The most widely adopted study design for investigating contributing factors to road traffic accidents and at-risk drivers is the *case–control study*. Case–control studies start with identifying cases, i.e. drivers who have been involved in a traffic accident, and controls, e.g. a random sample of

drivers recruited at roadside checkpoints (Houwing et al. 2009; Vandenbroucke and Pearce 2012). It is then determined whether the odds of being exposed to a certain external factor (e.g. alcohol consumption) prior to the accident or the odds of having an inherent quality (e.g. impulsivity) are higher for the group of drivers involved in a road traffic accident. Over the years, case–control studies have helped identify countless factors and characteristics that are associated with road traffic accidents, ranging from stable traits, such as demographics (e.g. age, gender, etc.) and personality characteristics (e.g. neuroticism, aggression, etc.) (Dumais et al., 2005; Mullin, Jackson, Langley, & Norton, 2000; Wang et al., 2012), to more transient states such as drug intoxications (Gjerde et al. 2011; Ravera et al. 2011). One of the most noteworthy results of case–control research has been providing the rationale for per se limits of alcohol in drivers. The first case–control study investigating the association between blood alcohol concentrations (BACs) and road traffic accidents is known as the Grand Rapids study (Borkenstein, Crowther, & Shumate, 1974; Hurst, Harte, & Frith, 1994). This study reported an exponential increase in crash risk from a BAC > 0.4 g/L onward. Together with more recent findings confirming that relative crash risk increase at a BAC > 0.4–0.5 g/L, the Grand Rapids study laid the foundation for the per se BAC limit of 0.5 g/L adhered to in many countries worldwide (Jones, 2010).

Another common type of epidemiological study is the *retrospective cohort study*. This type of study is a useful alternative to the case–control study when the factor or characteristic under investigation is relatively rare in the general population, e.g. suffering from narcolepsy (Tzeng et al., 2019). In retrospective cohort studies, participants who are known to present with a specific feature of interest are grouped and compared to a control group regarding the prior incidence of motor vehicle accidents. Alternatively, in a *prospective cohort study,* groups of presumed at-risk drivers and controls are compared regarding the incidence of motor vehicle accidents during a predetermined period following study enrollment, e.g. Nabi et al. (2006).

Although the epidemiological approach has proven valuable for identifying at-risk groups and framing legislation, these types of studies are subject to some practical and theoretical limitations. A

major practical concern of epidemiological research is the general requirement for large sample sizes. Especially retro- and prospective cohort studies face this obstacle because of the relatively low frequency of traffic accidents. Case–control studies start with identifying cases of drivers involved in traffic accidents and therefore circumvent the methodological problem of infrequent accident occurrences. However, like all epidemiological study designs, case–control studies face the problem that traffic accidents happen because of a multitude of reasons. This dilutes the relative prevalence of the characteristic or factor of interest. Again, this results in the need for large sample sizes.

The need for large samples is not an insurmountable obstacle as evidenced by the many epidemiological studies that have been conducted successfully. However, this type of research is limited in a more fundamental way. Epidemiological research methods are associative in nature. Because of the lack of randomization and manipulation, epidemiological studies are prone to confounding. A straightforward way of dealing with confounders is to measure or document likely confounders and then statistically correct for them. However, the many possible confounders make it challenging and laborious to deal with this issue. Furthermore, adding explanatory variables to statistical analyses generally decreases the statistical power to detect a true relation. In conclusion, epidemiological research designs are not optimal for demonstrating causality. In order to do that, epidemiological findings require support from experimental research. For further reading on types, strengths, and weaknesses of epidemiological study designs, the reader is referred to DiPietro (2010).

Experimental research: assessment of driving performance

In contrast to epidemiological studies, experimental studies are the preferred research method for studying mechanisms and demonstrating causality. However, in the field of traffic medicine, the benefit of demonstrating causality comes at a great cost. The actual outcome of interest, i.e. traffic accidents, cannot be studied directly. Instead, experimental researchers have to resort to the assessment of driving performance. It is presumed that traffic accidents are preceded by chronic

or acute impairments in driving performance, and hence, that the assessment of driving performance can predict the likelihood of a road traffic accident occurring.

A central question for the assessment of driving performance is the validity of the applied assessment techniques. Three types of validity should be considered when selecting tests or test batteries to assess driving performance, i.e. construct validity, criterion validity, and content validity (Crossley, Humphris, & Jolly, 2002; Cureton, 1951; Downing, 2003).

Construct validity is the extent to which a specific test (battery) assesses the construct of interest, i.e. (a specific aspect of) driving performance. In practice, the demonstration of construct validity relies on showing an association between test performances on different tasks hypothesized to assess the same underlying construct, a.k.a. convergent validity, which can be considered to be a part of construct validity (Carlson & Herdman, 2012). Another aspect of construct validity is ecological validity. In terms of driving performance assessment, ecological validity is the extent to which test outcomes translate to actual "real-world" driving performance. Hence, a test that closely resembles actual driving has high *ecological validity*. There is no established approach for determining ecological validity other than the careful consideration of the characteristics of the respective test and construct under investigation (Schmuckler, 2001). Lastly, for a test to have construct validity, it should be reliable. *Reliability* refers to the repeatability of a test outcome when administered under similar conditions (Golafshani, 2003).

Criterion validity captures how well performance on a specific task (e.g. reaction times in a visual search task) predicts a certain outcome (e.g. pass/fail during a driving examination or future accident risk). This form of validity is more related to performance thresholds than the test itself. Performance thresholds are necessary for making unequivocal decisions regarding fitness to drive and driving aptitude on both the group and individual levels. Hence, if a performance threshold is defined for a particular test, it should be supported by proof of criterion validity in order to be useful for making decisions regarding driver (un)fitness. Criterion validity can be demonstrated by comparing a

tests' outcome to that of a test with high ecological validity, such as on-road driving tests (see section on "On-road driving tests" below).

Lastly, *content validity* is the extent to which a test (battery) assesses all relevant functions that underlie (a specific aspect of) driving performance. As a heuristic, the hierarchical models of the driving task by Michon (1985) and Keskinen (1994) can be applied. The hierarchical model of the driving task proposed by Michon (1985) organized the driving task into three hierarchical levels. *The operational level* (1) at the bottom of the hierarchy consists of highly automated tasks required for basic vehicle control, such as road tracking and braking. Directly above the operational level is *the tactical level* (2). This level describes operations that allow the driver to negotiate the traffic environment. For example, overtaking and turning on an intersection depends on actions at the maneuvering level. The highest level of the hierarchy is *the strategic level* (3). On this level, route planning and a priori risk assessment (e.g. assessing the risk of going for a drive in bad weather) takes place. It is important to note that the different levels are interdependent, or as Michon (1985) puts it, exist in a "nested hierarchy." It is clear that the goals set at the strategic level, e.g. arriving at work on time, can only be accomplished if the necessary actions on the maneuvering level are performed, e.g. overtaking slower traffic or driving faster. In turn, the actions planned on the maneuvering level require the execution of the corresponding action patterns on the operational level to steer the vehicle along the envisioned course. Keskinen (1994) extended Michon's (1985) model with a fourth hierarchical level (Keskinen 1994; Hatakka et al. 2002). The three lowest levels, i.e. (1) *vehicle maneuvering*, (2) *mastering traffic situations*, and (3) *goals and context of driving,* correspond to Michon's (1985) operational, tactical, and strategic levels, respectively. The fourth level, labeled *goals for life and skills for living,* can be understood as reflecting a driver's personal attitudes relevant to driving, such as risk-taking behavior. The fourth level is an important addition because personal attitudes can potentially compensate or undermine functioning at the lower levels.

A comprehensive assessment of driving performance preferably covers all levels described in Keskinen's (1994) model. However, there is arguably no single assessment method that adequately

covers all the hierarchical levels of the driving task. Also, which functional levels need to be assessed depends on the suspected functional impairments. For example, elderly drivers are arguably more likely to be impaired on the lower hierarchical levels i.e. operational functions/vehicle maneuvering and tactical functions/mastering of traffic situations, than younger drivers who are more likely to exhibit more risk-taking behavior (Ball et al. 1998; McGwin Jr. and Brown 1999; Karthaus and Falkenstein 2016). Hence, in the context of driving performance, content validity is the extent to which all aspects of driving performance relevant to the specific population and research question at hand are assessed. It follows that the assessment of driving performance requires the careful assembly of a test battery consisting of on-road and/or simulated driving tasks combined with neurocognitive and psychological testing, depending on the driving behaviors or functional domains of interest. The rest of the section provides a general overview of the methods for the assessment of driving performance in experimental research.

On-road driving tests

On-road driving tests are often considered as the most ecologically valid assessments of driving performance. The tests are similar to a classic driving examination. The participant drives a predetermined route through actual traffic while a licensed instructor rates performance on items listed on a standardized checklist (for an example, see Table 1). Several on-road driving test procedures have been developed over the years. A recent systematic review by Sawada et al. (2019) investigated the validity and reliability of commonly used driving tests and concluded that the Washington University Road Test (Hunt et al., 1997), the Rhode Island Road Test (Brown et al., 2005), the Test Ride for Investigating Practical (TRIP) fitness to drive (Devos et al., 2007; Tant, Brouwer, Cornelissen, & Kooijman, 2002), and the Performance Analysis for Driving Ability (Patomella, Caneman, Kottorp, & Tham, 2004) were highly reliable assessment methods. Proof of criterion or construct validity was gathered by comparing the test outcomes to those of other on-road or neurocognitive tests, respectively. It should be noted that these tests were specifically designed to

assess the driving performance of rehabilitation patients. Nevertheless, the listed tests are good examples of high-quality on-road driving tests.

Item	Score range
Lateral position on the road at speed below 50 km/h	2–8
Lateral position on the road at speed above 50 km/h	2–8
Mechanical operations	3–12
Speed adaptations at speed below 50 km/h	2–8
Speed adaptations at speed above 50 km/h	2–8
Gap distance at speed below 50 km/h	2–8
Gap distance at speed above 50 km/h	2–8
Lane position change	5–20
Anticipation and perception of traffic light and signs	4–16
Visual behavior and communication	8–32
Understanding, insight, and quality of traffic perception	2–8
Turning left	9–36
Joining the traffic stream	6–24
Total score:	49–196

Table 1. Example of items and weights (obtainable scores) of the Test Ride for Investigating Practical (TRIP) fitness to drive. Source: Adapted from Devos et al. (2013).

On-road driving tests are considered the "golden standard" (Sawada et al., 2019) because the tests' high degree of ecological validity cannot be matched by any other assessment technique. These tests take place in actual traffic and usually cover a wide range of traffic environments (city centers, motorways, various types of intersections, etc.). However, several conceivable shortcomings deserve consideration. First, exposing a potentially impaired driver to the complexities of real-world traffic situations introduces safety concerns for the participant, examiner, and other road users. Second, a glance at the example checklist in Table 1 shows that on-road driving tests mainly assess operations at the two lower levels of Michon's (1985) and Keskinen's (1996) hierarchical models, i.e. skills directly related to operating a vehicle in traffic. The higher levels, where, for instance, risk-taking behavior and route planning reside, are not assessed despite their likely considerable impact on lower levels of functioning. Lastly, no two drives are the same because of the ever-changing traffic environment. Therefore, the applied checklists for rating driving performance must be limited to skills and

maneuvers that can be assessed during every drive. Hence, reactions to unexpected and/or near-crash events, arguably better predictors of crash risk, cannot be assessed.

The standardized on-the-road highway driving test

The standardized on-the-road highway driving test is a special case of on-road driving tests (Ramaekers, 2017; Verster & Roth, 2011). The test requires participants to complete a 100 km course on a public highway. The driver is instructed to maintain a steady lateral position in the right traffic lane and to keep a steady velocity of 95 km/h. It is allowed to ignore these instructions to overtake slower traffic, leaving the highway halfway to turn back, or because of immediate safety concerns. A licensed driving instructor who has access to dual controls, i.e. brake, accelerator, and clutch, accompanies the driver. Both driver and instructor can decide to abort the test if safety concerns arise. During the drive, an optic device mounted on top of the test vehicle continuously monitors the vehicle's lateral position relative to the traffic lane demarcation on the left, which is documented on the on-board computer drive. Also, velocity is continuously monitored. The data are later edited offline to mark events such as overtake maneuvers. The remaining data are then used to determine the mean speed, mean lateral position, standard deviation of speed, and the standard deviation of the lateral position (SDLP).

The SDLP is the primary outcome measure of the standardized on-road highway driving test. It is essentially a quantification of road-tracking performance or *lane weaving*, with higher values reflecting poorer performance. The measure has proven to be a very sensitive parameter for detecting the impairing effects of licit and illicit sedating intoxicants such as benzodiazepine receptor agonists, antihistamines, tri- and tetracyclic antidepressants, cannabis, alcohol, but also sleepiness and fatigue (Ramaekers et al. 2000; Vermeeren 2004; Theunissen et al. 2006; Bosker et al. 2012; Jongen et al. 2015; van der Sluiszen et al. 2016; Jongen et al. 2017). Randomized, placebo-controlled, crossover trials have reliably demonstrated a 2.4 cm increase in SDLP at a blood–alcohol concentration of 0.5 g/L (Jongen et al., 2017). Considering the finding of the aforementioned Grand Rapids study that

demonstrated an exponentially increased accident risk at this BAC (Borkenstein et al., 1974; Hurst et al., 1994), a 2.4 cm increase in SDLP serves as the benchmark for clinically or functionally relevant impairment and is considered to reflect the threshold for increased accident risk (Ramaekers, 2017). See figure 1 for an illustration of the standardized on the road driving test (Ramaekers, 2017).

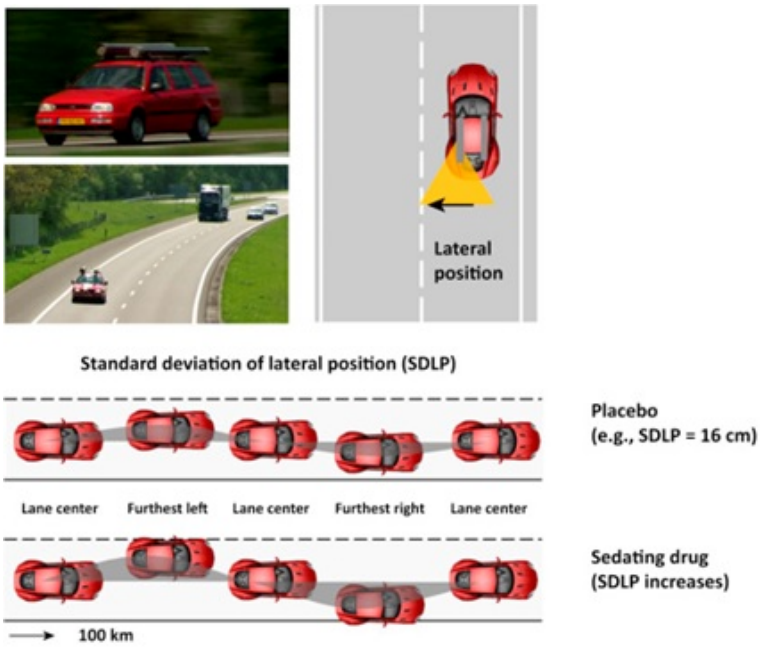

Figure 1. Illustration of the standardized on-road highway driving test and its main outcome measure, the standard deviation of lateral position (SDLP). Source: After Ramaekers (2017).

The standardized on-the-road highway driving test is subject to the same limitations as other on-road driving tests. Even more, it arguably only assesses functioning on the lowest level of Michon's (1985) and Keskinen's (1996) hierarchical models of driving as road tracking is a highly automated function in licensed/experienced drivers. However, because of the model's hierarchical nature, it can be reasoned that if performance is impaired at the lowest level, functioning is impaired on all levels. The standardized on-road highway driving test is a valuable tool for the assessment of sustained attention, because of its rather long duration (approximately 60 minutes) and the monotonous nature, in an ecologically valid setting. This makes the test a superior method for the assessment of the impairing effects of sedative and dissociative drugs (Leufkens, Ramaekers, de Weerd, Riedel, &

Vermeeren, 2009; Ramaekers, Robbe, & O'Hanlon, 2000; Theunissen, Vermeeren, & Ramaekers, 2006), as well as sleep and attentional disorders (Jongen, Perrier, Vuurman, Ramaekers, & Vermeeren, 2015; Verster et al., 2008).

Driving simulator tests

Driving simulators are becoming increasingly popular as more advanced setups become available. Driving simulator setups vary from simple setups consisting of a steering wheel and a screen to a complete vehicle, mounted on a moving base to mimic the kinetic input perceived during driving, with 360° projection of the simulated traffic environment. An example of such a highly advanced driving simulator is the National Advanced Driving Simulator 1 (NADS-1) at the University of Iowa (Chen, Papelis, Waston, & Solis, 2001) (Fig. 2, top right and left). Despite the superior ecological validity, there are few simulators of the latter type, likely because of the high costs associated with purchase, construction, maintenance, and operation. More commonly used are simpler setups consisting of a driver seat, pedals (accelerator, brake, and clutch), steering wheel, and a screen displaying the traffic environment and dashboard, all mounted on a fixed base (Cuenen et al., 2019; Mets et al., 2011). Other, more advanced fixed-base setups involve the body of an actual vehicle mounted in front of a large projection screen (Fig. 2, bottom) (Hussain, Pirdavani, Ariën, Brijs, & Alhajyaseen, 2018).

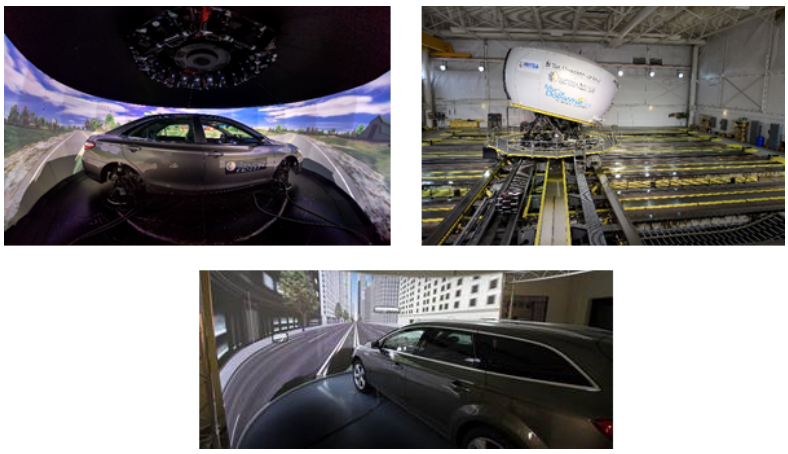

Figure 2. Photographs taken from the inside (top left) and outside (top right) the National Advanced Driving Simulator (NADS-1), a highly advanced moving base driving simulator at the University of Iowa, USA. The bottom image displays a more advanced fixed-base simulator that features the cabin of an actual vehicle, as setup at the Institute for Mobility (IMOB) at Hasselt University, Belgium.

Driving simulator testing circumvents many problems associated with on-road driving tests. First, there are minimal, if any, safety concerns associated with simulator testing. It is therefore possible to assess driving performance of impaired drivers without introducing potential hazards. Driving simulators often come equipped with standard driving scenario's, but most simulation software also allows for the "de novo" design or tweaking of existing scenario's (Fisher, Rizzo, Caird, & Lee, 2011). As such, virtually every aspect of the traffic environment can be manipulated by the experimenter. This makes driving simulators a potent tool for investigating driving behavior under specific (rare or dangerous) circumstances. Also, by manipulating the simulated environment, tests of executive functioning (e.g. working memory, selective attention, etc.) can be embedded in the scenarios (e.g. counting the number of red vehicles encountered) (Fisher et al., 2011). Lastly, an advantage of driving simulator testing is the abundance of outcome measures that can be readily documented, ranging from parameters of direct vehicle control (e.g. steering wheel angle, braking force, accelerator release, etc.) to more dynamic outcome measures (e.g. distance keeping, lateral position, velocity, accelerator release in response to an event, etc.) (Fisher et al., 2011).

However, driving simulators have their own unique limitations. The most prominent one is the possibility of simulator sickness (Helland et al., 2016; Kennedy, Lane, Berbaum, & Lilienthal, 1993). Simulator sickness is similar to motion sickness and likely arises from the mismatch between visual, vestibular, and proprioceptive input (Groen & Bos, 2008). Common complaints are oculomotor strain, feelings of disorientation, and nausea (Kennedy et al., 1993). It is clear that these complaints have the potential to distort the measurements during simulator testing (Helland et al., 2016). Even if there are no apparent signs of simulator sickness, driving performance might still be influenced by the absence of vestibular and proprioceptive input that plays an important role in perceiving acceleration, deceleration, and centrifugal forces during turns (Reymond, Kemeny, Droulez, & Berthoz, 2001). In order to prevent these limitations, highly advanced moving base simulators can be used to mimic actual vehicle movements during driving. However, as stated above, advanced moving base simulators like the NADS-1 are rare. Lastly, risk-taking behavior is not easily assessed in a naturalistic way during

driving simulator testing because of the driver's awareness of the artificial nature of the simulated drive, which might prompt more risk taking because of the absence of real danger. Conversely, the driver might be inclined to drive more responsibly because of test instruction or the social desire to appear as a capable driver to the experimenter.

Neurocognitive and psychological testing

In addition to on-road and driving simulator testing, neurocognitive and psychological test batteries are frequently used to assess driving-related functioning. The US Food and Drug Administration (FDA) provided guidelines for evaluating drug effects on fitness to drive in which they list the minimal required functional domains to be assessed for evaluating driving-related functioning, i.e. (1) alertness and arousal, (2) attention and processing speed, (3) reaction time and psychomotor functioning, (4) sensory perceptual functioning, and (5) executive functioning (Food and Drug Administration, 2017). These functions are rather broad and require deconstruction into subfunctions in order to be assessable. Table 2 shows examples of subfunctions and applicable neurocognitive and psychomotor tests to assess the functional domains listed by the FDA. These functional domains cover functioning/operating at the first three levels of Michon's (1985) and Keskinen's (1994) models, but disregard the highest level of Keskinen's (1994) model which covers relevant personal attitudes and characteristics such as risk taking, impulsivity, aggression, and conscientiousness. Hence, a sixth functional domain with examples of relevant tests was added to Table 2.

Functional domain	Examples of relevant functions	Examples of relevant tests
Alertness and arousal	Alerting, orienting, sustained attention	ANT[a], UFOV[b], PVT[c]
Attention and processing speed	Stimulus response matching, divided attention, selective attention, top-down controlled visual search	ANT[a], SART[d], DSST[e], TMT[f], Stroop test[g], DAT[h]
Reaction time and psychomotor functioning	Fine motor skills, reaction speed	CTT[i], PRT[j], FTT[k], PVT[c]
sensory perceptual functioning	Visual acuity, contrast sensitivity, movement perception, object recognition	ATTPT[l], UFOV[b], CCF[m], motion perception[n], dynamic visual acuity[o]
Executive functioning	mapping dynamic traffic scene, evaluation of performance, planning, working memory	Spatial working memory test[p], WCST[q], Tower of London[r], n-back task[s]
Personal attitudes and characteristics	Impulsivity, risk taking, aggression, respecting traffic rules	RT-18[t], BIS[u], DBQ[v]</TB>

Table 2. Overview of the five functional domains proposed by the FDA (Food and Drug Administration, 2017) which need to be assessed at a minimum for the investigation of driver fitness, complemented with a sixth domain comprising personal attitudes and characteristics, with examples of relevant functions and applicable tests. [a]Attention Network Test (Fan, McCandliss, Sommer, Raz, & Posner, 2002); [b]Useful Field of View test (Ball & Owsley, 1993); [c]Psychomotor Vigilance Task (Dinges & Powell, 1985); [d]Sustained Attention to Response Test (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997); [e]Digit-Symbol Substitution Test (McLeod, Griffiths, Bigelow, & Yingling, 1982); [f]Trailmaking Test (Tombaugh, 2004); [g]Stroop Color and Word Test (Stroop, 1935); [h]Divided Attention Test (Moskowitz, 1973); [i]Critical Tracking Task (Jex, McDonnell, & Phatak, 1966); [j]Pursuit Rotor Task (Ammons, 1947); [k]Finger Tapping Test (Morrison, Gregory, & Paul, 1979); [l]Adaptive Tachistoscopic Traffic Perception Test (Schuhfried, 2009); [m]Critical Flicker Fusion Threshold (Landis, 1954); [n](Lacherez, Au, & Wood, 2014; Wood, 2002); [o](Herdman et al., 1998; Vital et al., 2010); [p](Owen, Downes, Sahakian, Polkey, & Robbins, 1990); [q]Wisconsin Card Sorting Test (Berg, 1948); [r](Krikorian, Bartok, & Gay, 1994); [s](Owen, McMillan, Laird, & Bullmore, 2005); [t](De Haan et al., 2011); [u]Barratt Impulsiveness Scale (Patton, Stanford, & Barratt, 1995; Stanford et al., 2009); [v]Driver Behavior Questionnaire (Martinussen, Hakamies-Blomqvist, Møller, Özkan, & Lajunen, 2013).

The list in Table 2. only serves to present examples of relevant functions and tests and is by no means intended to define the necessary features of a comprehensive and universal test battery. Neurocognitive and psychological test batteries are often assembled with specific driver groups in mind, e.g. stroke rehabilitation or traumatic brain injury patients (McKenna, Jefferies, Dobson, &

Frude, 2004; Nouri & Lincoln, 1993), and validated through comparison with on-road driving test results (pass/fail) with participants from the same population. Not only test batteries but also individual tests are often investigated for their suitability for the assessment of driving performance under specific circumstances, e.g. impairing effects of CNS depressants (Jongen, Vuurman, Ramaekers, & Vermeeren, 2016) or sleep deprivation (Jongen et al., 2015). Thus, when assessing driving performance, it should be considered whether the selected test (battery) is suitable for the demonstration of impairment in the population or context under investigation.

The main advantage of neurocognitive and psychological testing is that the tests can be administered in laboratory or clinic settings, which is convenient, affordable, and safe. In addition, the approach allows for the detailed assessment of functioning relevant at all levels of Michon's and Keskinen's hierarchical models (Keskinen, 1994; Michon, 1985). Unlike the aforementioned testing methods, neurocognitive and psychological test batteries can be constructed to thoroughly assess higher order functions (executive functioning), such as route planning and risk assessment, and personal attitudes, such as consciousness and risk-taking behavior. The clear disadvantage of neurocognitive and psychological testing is the lack of ecological validity. Unless supported by proof of convergent validity from on-road and simulator testing, it is uncertain what role specific neurocognitive functions or psychological characteristics play in overall driving performance.

Observational research of driving performance

Besides epidemiological and experimental research designs, there is also a third option, which is a hybrid form of the former study types. Observational research of driving performance selects pre-existing groups, e.g. long-term hypnotic or anxiolytic users (van der Sluiszen et al., 2019), and applies testing procedures commonly used in experimental studies for the assessment of driving performance. The performance of the group of interest is then compared to that of a control group. These between-subject research designs suffer from the same limitations regarding demonstrating causality as epidemiological studies due to the lack of randomization. They also lack the potential to

demonstrate predictive validity because of the inability to directly assess actual road traffic accidents as an outcome. Nevertheless, observational studies can contribute to proof of convergent validity together with epidemiological findings whenever experimental manipulation is not possible, e.g. diseases and prolonged or irreversible treatments.

Individual assessment of driving aptitude and fitness to drive

Epidemiological and experimental research attempts to uncover risk factors for road traffic accidents. Although these empirical methods allow for identifying groups of drivers who have an increased risk of being involved in a road traffic accident, the findings cannot be extrapolated directly to individual drivers. Also, not all individual cases permit a thorough assessment of driving performance because of circumstantial or practical constraints. The current section discusses some important considerations in the individual assessment of fitness to drive and driving aptitude.

Assessment of fitness to drive

As described above, fitness to drive is defined as the situation- and time-dependent ability to drive a vehicle. Clinicians most often deal with the need for anticipation of unfitness to drive, e.g. impairing effects of prescribed medications, while forensic experts will generally deal with determining (un)fitness to drive retrospectively based on evidence collected at the time of suspicion. Examples of situations where fitness to drive can be impaired are acute drug intoxications, certain medical conditions, such as a transient ischemic attack, temporary mood, or emotional states such as depression or a fit of anger, and many more. These conditions' transient nature introduces the necessity to determine fitness to drive soon after a suspicion of unfitness to drive arises. However, the resources for assessing driving performance and driving-related functioning are usually not available when a suspicion of unfitness to drive presents itself. Therefore, the temporary (in)ability to adequately operate a motor vehicle should be predicted or evaluated after the fact based on available knowledge and evidence.

Hence, the forensic expert will usually provide an expert opinion regarding fitness to drive in the context of a police investigation. The police investigation starts when suspicion of unfitness to drive occurs. This suspicion usually results from the observation of deviating driving behavior by either police officials or eyewitnesses. Examples of relevant observations are handling of the car, behavior of driver and passengers, weaving, driving in the middle of the road, maneuvering (skidding on a curve or curb), speed (too fast and too slow), tailgating, reactions (slow and belated), starting or breaking abruptly or belatedly, and crashes or near-crashes (Berghaus & Schnabel, 2014).

Especially when the police were the observers of the suspicious driving behavior, the observation itself already provides valuable information about the driver's potential functional impairment at the time of the observation. However, it is not sufficient for the determination of (un)fitness to drive. It is crucial to establish whether the deviating driving behavior persists for some time, as every driver occasionally commits a driving error. Continued observation of driving behavior can be performed under circumstances where there is no immediate danger to the suspected driver or other road users. However, for safety reasons, the further collection of information by the police is usually carried out at the roadside after the driver has been stopped.

The main source of information that might help identify drivers who are unfit to drive at roadside is the driver's behavior and appearance. Examples of relevant observations are behavior and mood as well as changes in behavior and mood (tiredness, aggressiveness, etc.), physical abnormalities (speech, walking, agitation, tremor, etc.), orientation, responsiveness, reaction, logical reasoning, abnormalities concerning clothes or smell (in the car and in the respiratory air), and pupils (extension and reaction of the pupil) (Berghaus & Schnabel, 2014). In response to observations of driving behavior and appearance, the police can choose to collect evidence of acute intoxications affecting fitness to drive. Depending on the legal possibilities within a given jurisdiction, common test procedures involve detecting or measuring drug concentrations in biological samples, such as breath (alcohol breathalyzer), saliva, urine, or blood plasma, or behavioral testing, such as the standardized field sobriety test (SFST).

A few points regarding the collection of biological samples to detect substances known to impair driving deserve consideration. First, it is important to consider that detecting a drug known to impair driving does not necessarily indicate functional impairment. For instance, two studies on the effects of smoked cannabis on psychomotor and neurocognitive functioning in heavy cannabis users failed to demonstrate significant functional impairment, despite the fact that the tetrahydrocannabinol (THC) concentrations in blood plasma corresponded to oral fluid THC concentrations that would provide a positive result on some roadside saliva drug tests (Huestis and Cone 2004; Ramaekers et al. 2009; Wille et al. 2010; Ramaekers et al. 2011; Dobri et al. 2019). Roadside saliva and urine drug tests usually provide a binary outcome, i.e. positive or negative result for the presence of the drug (or metabolite) of interest. However, depending on the legislation (some jurisdiction apply per se limits while others have a zero tolerance policy), or specific circumstances (e.g. investigating whether a driver who tested positive for ketamine ran over his neighbor by accident or not), it might be necessary to follow up a positive result by the determination of actual drug concentrations to establish whether the drug was likely to cause the presumed driving impairment (Raes & Verstraete, 2005). Drug concentrations are usually determined in blood plasma samples. Most often, the collection of a blood plasma sample takes place some time after the positive drug test. Hence, the outcome of interest for the forensic expert is the estimated drug concentration at that time that aberrant driving was observed or a road traffic accident occurred. The estimated drug plasma concentration can then be compared to findings from epidemiological and experimental research so that the plausibility of the drug being a causal factor for driving impairment can be considered. However, in the absence of information on drug dosing and time or route of administration, it is not possible to retrospectively determine blood plasma concentrations with certainty. Furthermore, findings from epidemiological and experimental studies indicate drug concentrations at which an increased crash risk or impairment of driving (related functions) becomes apparent at the group level. However, group-level findings from empirical research cannot be extrapolated to the individual without further consideration. Whether a given person is functionally

impaired at a certain drug concentration arguably depends on many factors such as tolerance, concomitant drug use, or personality (e.g. tendency for aggression or impulsivity).

Hence, the police's observations regarding driving and driver behavior are valuable pieces of information that can help determine whether a driver was likely impaired due to drug use at a specific time in the past. In addition to general observations, an SFST (Burns & Moskowitz, 1977) is used in some jurisdictions, e.g. USA and Canada. The SFST consists of three parts, i.e. horizontal gaze nystagmus (HGN), one leg stand (OLS), and walk-and-turn (WAT) test. The HGN test requires the driver to follow an object (e.g. a pen or fingertip) with the eyes. The OLS assesses whether the driver is capable of standing on one leg for 30 seconds with the other foot approximately 15 cm above the ground. Lastly, the WAT consists of nine heel-to-toe steps along a straight line, after which the driver must turn back on one foot and take another nine heel-to-toe steps along the straight line. The SFST was developed for the detection of alcohol-induced impairment but has also been demonstrated to pick up stimulant and CNS depressant induced impairment (Porath-Waller & Beirness, 2014). However, it is unclear to what extent impaired performance during the SFTS is indicative of actual driving impairment. Also, the sensitivity of the SFST might be suboptimal. For instance, a study of the effects of THC (dronabinol) on driving performance demonstrated impaired performance in the on-the-road highway driving test in both occasional and heavy cannabis users, while the SFST was not able to discriminate between placebo and dronabinol conditions (Bosker et al., 2012). Nevertheless, it appears highly likely that severe impairment during the SFST is indicative of actual driving impairment.

In conclusion, the transient nature of (un)fitness to drive makes a thorough assessment of driving performance and driving-related functioning practically impossible. Hence, the forensic expert is required to carefully consider all the pieces of information that are available, i.e. police reports of observed driving behavior or traffic accident circumstances, behavior and appearance of the driver as observed by police, results of toxicological screening for the presence of driving impairing drugs, and

information yielded during potential systematic assessments such as the SFST, in order to formulate proper advice.

Assessment of driving aptitude

After a driver has been judged to be unfit to drive, an assessment of driving aptitude may be necessary. Also, when there is doubt regarding the (un)fitness to drive at some point in the past, it might be advisable to assess driving aptitude. Examples of other reasons for ordering the assessment of driving aptitude can be traffic-related criminal offenses or criminal offenses with a high aggression potential, subsequent to a medical assessment if deemed necessary, renewal of the driving license after license withdrawal due to penalty points, renewal of driving license at a certain age (required for elderly drivers in some jurisdictions), and repeated or severe offenses (Berghaus and Schnabel 2014).

The methods for assessing driving aptitude consist largely of the same techniques used in experimental research (see Section 53.2.2). As discussed, there is a wide variety of on-road driving, simulated driving, and neurocognitive and psychological tests available to assess driving performance and driving-related functioning. The selected assessment methods must specifically consider the concrete information leading to the doubts about the driver's aptitude. Often validated test batteries have been constructed for the assessment of driving aptitude in specific populations, e.g. patients with dementia or mild cognitive impairment (Fuermaier et al., 2017; Piersma et al., 2018; Piersma et al., 2016), or traumatic brain injury (McKenna et al., 2004).

For certain medical conditions, the experimental methods discussed in Section 53.2.2 are inadequate to assess driving aptitude. Driving *inaptitude* can be conceptualized as a constant inability to safely operate a motor vehicle in traffic (e.g. patients with traumatic brain injury with lasting functional consequences) or as recurring bouts of unfitness to drive (e.g. epilepsy patients). Especially, medical conditions that cause driving inaptitude in the latter sense are not easily assessed with the methods described in Section 53.2.2. Instead, a thorough medical examination should be conducted, which also considers general physical and psychological condition, controllability of the disease by

therapy (including medication), side effects of the medicines used, duration, intensity and severity of the disease, a combination of diseases, kind of participation in traffic of the driver (personal activities, commercial vehicles, and professional drivers), and the possibility to compensate for deficits (Berghaus and Schnabel 2014). An example of an alternative method for determining driver aptitude in epilepsy patients was proposed by Somerville et al. (2019), who introduced a decision tree incorporating the recent medical history of the patient.

The individual assessment of driving aptitude has to be coherent and reviewable. This means that the inducing facts and resulting questions must be included, i.e. the applied assessment methods, and, if necessary, an explanation and discussion of the findings and their meaning. The report has to clearly distinguish between the previous history and the current findings. Additionally, the assessment report must be comprehensible, so that the assessed person and the driving license agency, court, and other relevant parties are able to comprehend the report. If the assessment leads to a negative result concerning the driver's aptitude, recommendations should be given on how the preconditions for meeting the requirements can be improved (e.g. by substance abuse counseling or a therapeutic measure). A positive assessment may be made possible by naming impositions (e.g. wearing glasses) or restrictions (e.g. speed limitations).

Conclusion

The current chapter aimed to provide a general overview of the techniques for identifying risk factors for road traffic accidents and the assessment of driving performance and related functioning, as well as their strengths and weaknesses. In addition, the available techniques and difficulties in assessing individual fitness to drive and driving aptitude were considered. It is clear that the selected sources of knowledge and methods of assessment greatly depend on the question at hand.

References

Ammons, R. B. (1947). Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *Journal of Experimental Psychology, 37*(5), 393.

Ball, K., & Owsley, C. (1993). The useful field of view test: a new technique for evaluating age-related declines in visual function. *Journal of the American Optometric Association, 64*(1), 71-79.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of general psychology, 39*(1), 15-22.

Berghaus, G., & Schnabel, E. (2014). Driving Aptitude and Fitness to Drive. *Handbook of Forensic Medicine*, 995-1010.

Borkenstein, R. F., Crowther, R. F., & Shumate, R. P. (1974). The role of the drinking driver in traffic accidents (the Grand Rapids study). *Blutalkohol, 11*(Suppl), 1-131.

Bosker, W., Kuypers, K., Theunissen, E., Surinx, A., Blankespoor, R., Skopp, G., . . . Ramaekers, J. G. (2012). Medicinal THC (dronabinol) impairs on-the-road driving performance of occasional and heavy cannabis users but is not detected in Standardized Field Sobriety Tests. *Addiction, 107*(10), 1837-1844.

Brown, L. B., Ott, B. R., Papandonatos, G. D., Sui, Y., Ready, R. E., & Morris, J. C. (2005). Prediction of on-road driving performance in patients with early Alzheimer's disease. *Journal of the American Geriatrics Society, 53*(1), 94-98.

Burns, M., & Moskowitz, H. (1977). *Psychophysical tests for DWI arrest*. Retrieved from

Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods, 15*(1), 17-32.

Chen, L., Papelis, Y., Waston, G., & Solis, D. (2001). *NADS at the University of IOWA: A tool for driving safety research.* Paper presented at the Proceedings of the 1st human-centered transportation simulation conference.

Crossley, J., Humphris, G., & Jolly, B. (2002). Assessing health professionals. *Medical education, 36*(9), 800-804.

Cuenen, A., Jongen, E. M., Brijs, T., Brijs, K., Van Vlierden, K., & Wets, G. (2019). The effect of a simulator based training on specific measures of driving ability in older drivers. *Transportation research part F: traffic psychology and behaviour, 64*, 38-46.

Cureton, E. (1951). Educational measurement. *American Council on Education*.

De Haan, L., Kuipers, E., Kuerten, Y., van Laar, M., Olivier, B., & Verster, J. C. (2011). The RT-18: a new screening tool to assess young adult risk-taking behavior. *International journal of general medicine, 4*, 575.

Devos, H., Nieuwboer, A., Vandenberghe, W., Tant, M., De Weerdt, W., & Uc, E. (2013). Validation of driving simulation to assess on-road performance in Huntington disease.

Devos, H., Vandenberghe, W., Nieuwboer, A., Tant, M., Baten, G., & De Weerdt, W. (2007). Predictors of fitness to drive in people with Parkinson disease. *Neurology, 69*(14), 1434-1441.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior research methods, instruments, & computers, 17*(6), 652-655.

Downing, S. M. (2003). Validity: on the meaningful interpretation of assessment data. *Medical education, 37*(9), 830-837.

Dumais, A., Lesage, A. D., Boyer, R., Lalovic, A., Chawky, N., Ménard-Buteau, C., . . . Turecki, G. (2005). Psychiatric risk factors for motor vehicle fatalities in young men. *The Canadian Journal of Psychiatry, 50*(13), 838-844.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of cognitive neuroscience, 14*(3), 340-347.

Fisher, D. L., Rizzo, M., Caird, J., & Lee, J. D. (2011). *Handbook of driving simulation for engineering, medicine, and psychology*: CRC Press.

Food and Drug Administration. (2017). Evaluating Drug Effects on the Ability to Operate a Motor Vehicle, Guidance for Industry. 2017.

Fuermaier, A. B., Piersma, D., de Waard, D., Davidse, R. J., de Groot, J., Doumen, M. J., . . . Scheltens, P. (2017). Assessing fitness to drive—a validation study on patients with mild cognitive impairment. *Traffic injury prevention, 18*(2), 145-149.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The qualitative report, 8*(4), 597-607.

Groen, E. L., & Bos, J. E. (2008). Simulator sickness depends on frequency of the simulator motion mismatch: An observation. *Presence: Teleoperators and Virtual Environments, 17*(6), 584-593.

Haagsma, J. A., Graetz, N., Bolliger, I., Naghavi, M., Higashi, H., Mullany, E. C., . . . Phillips, M. R. (2016). The global burden of injury: incidence, mortality, disability-adjusted life years and time trends from the Global Burden of Disease study 2013. *Injury Prevention, 22*(1), 3. doi:10.1136/injuryprev-2015-041616

Helland, A., Lydersen, S., Lervåg, L.-E., Jenssen, G. D., Mørland, J., & Slørdal, L. (2016). Driving simulator sickness: impact on driving performance, influence of blood alcohol concentration, and effect of repeated simulator exposures. *Accident Analysis & Prevention, 94*, 180-187.

Herdman, S. J., Tusa, R. J., Blatt, P., Suzuki, A., Venuto, P. J., & Roberts, D. (1998). Computerized dynamic visual acuity test in the assessment of vestibular deficits. *Otology & Neurotology, 19*(6), 790-796.

Hunt, L. A., Murphy, C. F., Carr, D., Duchek, J. M., Buckles, V., & Morris, J. C. (1997). Reliability of the Washington University Road Test: A performance-based assessment for drivers with dementia of the Alzheimer type. *Archives of neurology, 54*(6), 707-712.

Hurst, P. M., Harte, D., & Frith, W. J. (1994). The grand rapids dip revisited. *Accident Analysis & Prevention, 26*(5), 647-654.

Hussain, Q., Pirdavani, A., Ariën, C., Brijs, T., & Alhajyaseen, W. (2018). The impact of perceptual countermeasures on driving behavior in rural-urban transition road segments: a driving simulator study. *Advances in transportation studies, 46*.

Jex, H. R., McDonnell, J. D., & Phatak, A. V. (1966). A``Critical''Tracking Task for Manual Control Research. *IEEE Transactions on Human Factors in Electronics*(4), 138-145.

Jones, A. W. (2010). The relationship between blood alcohol concentration (BAC) and breath alcohol concentration (BrAC): a review of the evidence. *Road safety web publication, 15*, 1-43.

Jongen, S., Perrier, J., Vuurman, E., Ramaekers, J., & Vermeeren, A. (2015). Sensitivity and validity of psychometric tests for assessing driving impairment: effects of sleep deprivation. *PLoS one, 10*(2), e0117045.

Jongen, S., Vermeeren, A., van der Sluiszen, N., Schumacher, M., Theunissen, E., Kuypers, K., . . . Ramaekers, J. G. (2017). A pooled analysis of on-the-road highway driving studies in actual traffic measuring standard deviation of lateral position (ie,“weaving”) while driving at a blood alcohol concentration of 0.5 g/L. *Psychopharmacology, 234*(5), 837-844.

Jongen, S., Vuurman, E., Ramaekers, J., & Vermeeren, A. (2016). The sensitivity of laboratory tests assessing driving related skills to dose-related impairment of alcohol: A literature review. *Accident Analysis & Prevention, 89*, 31-48.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology, 3*(3), 203-220.

Keskinen, E. (1994). Why do young drivers have more accidents? Junge Fahrer und Fahrerinnen. *Referate der Ersten Interdiziplinaren Fachkonferenz, December 12-14, 1994 in Koln*.

Krikorian, R., Bartok, J., & Gay, N. (1994). Tower of London procedure: a standard method and developmental data. *Journal of clinical and Experimental Neuropsychology, 16*(6), 840-850.

Lacherez, P., Au, S., & Wood, J. M. (2014). Visual motion perception predicts driving hazard perception ability. *Acta ophthalmologica, 92*(1), 88-93.

Landis, C. (1954). Determinants of the critical flicker-fusion threshold. *Physiological Reviews, 34*(2), 259-286.
Leufkens, T., Ramaekers, J., de Weerd, A., Riedel, W., & Vermeeren, A. (2009). On-the-road driving performance and driving related skills in untreated insomnia patients and chronic users of hypnotics. *HYPNOTICS AND ANXIOLYTICS*, 97.
Martinussen, L. M., Hakamies-Blomqvist, L., Møller, M., Özkan, T., & Lajunen, T. (2013). Age, gender, mileage and the DBQ: The validity of the Driver Behavior Questionnaire in different driver groups. *Accident Analysis & Prevention, 52*, 228-236.
McKenna, P., Jefferies, L., Dobson, A., & Frude, N. (2004). The use of a cognitive battery to predict who will fail an on-road driving test. *British Journal of Clinical Psychology, 43*(3), 325-336.
McLeod, D. R., Griffiths, R. R., Bigelow, G. E., & Yingling, J. (1982). An automated version of the digit symbol substitution test (DSST). *Behavior Research Methods & Instrumentation, 14*(5), 463-466.
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. *Human Psychopharmacology: Clinical and Experimental, 26*(6), 434-439.
Michon, J. A. (1985). A critical view of driver behavior models: what do we know, what should we do? *Human behavior and traffic safety* (pp. 485-524): Springer.
Morrison, M. W., Gregory, R. J., & Paul, J. J. (1979). Reliability of the Finger Tapping Test and a note on sex differences. *Perceptual and Motor Skills, 48*(1), 139-142.
Moskowitz, H. (1973). Laboratory studies of the effects of alcohol on some variables related to driving. *Journal of Safety Research*.
Mullin, B., Jackson, R., Langley, J., & Norton, R. (2000). Increasing age and experience: are both protective against motorcycle injury? A case-control study. *Injury Prevention, 6*(1), 32-35.
Nabi, H., Guéguen, A., Chiron, M., Lafont, S., Zins, M., & Lagarde, E. (2006). Awareness of driving while sleepy and road traffic accidents: prospective study in GAZEL cohort. *Bmj, 333*(7558), 75.
Nouri, F. M., & Lincoln, N. B. (1993). Predicting driving performance after stroke. *BMJ: British Medical Journal, 307*(6902), 482.
Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia, 28*(10), 1021-1034.
Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human brain mapping, 25*(1), 46-59.
Patomella, A.-H., Caneman, G., Kottorp, A., & Tham, K. (2004). Identifying scale and person response validity of a new assessment of driving ability. *Scandinavian journal of occupational therapy, 11*(2), 70-77.
Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of clinical psychology, 51*(6), 768-774.
Piersma, D., Fuermaier, A. B., De Waard, D., Davidse, R. J., De Groot, J., Doumen, M. J., . . . Scheltens, P. (2018). Assessing fitness to drive in patients with different types of dementia. *Alzheimer disease and associated disorders, 32*(1), 70.
Piersma, D., Fuermaier, A. B., De Waard, D., Davidse, R. J., de Groot, J., Doumen, M. J., . . . Vermeeren, A. (2016). Prediction of fitness to drive in patients with Alzheimer's dementia. *PLoS One, 11*(2), e0149566.
Porath-Waller, A. J., & Beirness, D. J. (2014). An examination of the validity of the standardized field sobriety test in detecting drug impairment using data from the drug evaluation and classification program. *Traffic injury prevention, 15*(2), 125-131.

Raes, E., & Verstraete, A. G. (2005). Usefulness of roadside urine drug screening in drivers suspected of driving under the influence of drugs (DUID). *Journal of analytical toxicology, 29*(7), 632-636.

Ramaekers, J. G. (2017). Drugs and driving research in medicinal drug development. *Trends in pharmacological sciences, 38*(4), 319-321.

Ramaekers, J. G., Robbe, H., & O'Hanlon, J. (2000). Marijuana, alcohol and actual driving performance. *Human Psychopharmacology: Clinical and Experimental, 15*(7), 551-558.

Ranney, T. A. (1994). Models of driving behavior: a review of their evolution. *Accident Analysis & Prevention, 26*(6), 733-750.

Reymond, G., Kemeny, A., Droulez, J., & Berthoz, A. (2001). Role of lateral acceleration in curve driving: Driver model and experiments on a real vehicle and a driving simulator. *Human factors, 43*(3), 483-495.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35*(6), 747-758.

Sawada, T., Tomori, K., Hamana, H., Ohno, K., Seike, Y., Igari, Y., & Fujita, Y. (2019). Reliability and validity of on-road driving tests in vulnerable adults: a systematic review. *International Journal of Rehabilitation Research, 42*(4).

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy, 2*(4), 419-436.

Schuhfried, G. (2009). Adaptive Tachistoscopic Traffic Perception Test. *ATAVT Test Manual version, 22*.

Singh, H., Kushwaha, V., Agarwal, A., & Sandhu, S. (2016). Fatal road traffic accidents: Causes and factors responsible. *Journal of Indian Academy of Forensic Medicine, 38*(1), 52-54.

Somerville, E. R., Australia, D. C. o. t. E. S. o., Somerville, E., Black, A., Lander, C., Jones, D., . . . Archer, J. (2019). A decision tree to determine fitness to drive in epilepsy: results of a pilot in two Australian states. *Epilepsia, 60*(7), 1445-1452.

Staak, M., Hobi, V., & Berghaus, G. (1988). *Testverfahren zur Überprüfung des Einflusses von Arzneimitteln auf die Verkehrssicherheit: medizinischer Schwerpunkt: Bericht zum Forschungsprojekt 8305/2 der Bundesanstalt für Strassenwesen, Bereich Unfallforschung*: Bundesanstalt für Strassenwesen. Bereich Unfallforschung.

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and individual differences, 47*(5), 385-395.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18*(6), 643.

Tant, M., Brouwer, W., Cornelissen, F., & Kooijman, A. (2002). Driving and visuospatial performance in people with hemianopia. *Neuropsychological rehabilitation, 12*(5), 419-437.

Theunissen, E. L., Vermeeren, A., & Ramaekers, J. G. (2006). Repeated-dose effects of mequitazine, cetirizine and dexchlorpheniramine on driving and psychomotor performance. *British journal of clinical pharmacology, 61*(1), 79-86.

Tombaugh, T. N. (2004). Trail Making Test A and B: normative data stratified by age and education. *Archives of clinical neuropsychology, 19*(2), 203-214.

Tzeng, N.-S., Hsing, S.-C., Chung, C.-H., Chang, H.-A., Kao, Y.-C., Mao, W.-C., . . . Chien, W.-C. (2019). The risk of hospitalization for motor vehicle accident injury in narcolepsy and the benefits of stimulant use: a nationwide cohort study in Taiwan. *Journal of clinical sleep medicine, 15*(6), 881-889.

van der Sluiszen, N. N., Vermeeren, A., Verster, J. C., van de Loo, A. J., van Dijken, J. H., Veldstra, J. L., . . . Ramaekers, J. G. (2019). Driving performance and neurocognitive skills of long-term users of benzodiazepine anxiolytics and hypnotics. *Human Psychopharmacology: Clinical and Experimental, 34*(6), e2715.

Verster, J. C., Bekker, E. M., de Roos, M., Minova, A., Eijken, E. J., Kooij, J. S., . . . Olivier, B. (2008). Methylphenidate significantly improves driving performance of adults with attention-deficit hyperactivity disorder: a randomized crossover trial. *Journal of Psychopharmacology, 22*(3), 230-237.

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.

Vital, D., Hegemann, S. C., Straumann, D., Bergamin, O., Bockisch, C. J., Angehrn, D., . . . Probst, R. (2010). A new dynamic visual acuity test to assess peripheral vestibular function. *Archives of Otolaryngology–Head & Neck Surgery, 136*(7), 686-691.

Wang, X., Yang, L., Zhang, H., Liu, Y., Tang, Z., Wang, J., . . . Dong, A. (2012). Association of type-A behavior and personality with accident proneness among drivers. *Chin J Public Health, 28*(03), 282-284.

Wood, J. M. (2002). Age and visual impairment decrease driving performance as measured on a closed-road circuit. *Human factors, 44*(3), 482-494.

World Health Organization. (2015). *Global status report on road safety 2015* (9241565063). Retrieved from

World Health Organization. (2018). *Global status report on road safety 2018: Summary*. Retrieved from

Chapter 3: Effects of Solriamfetol on On-the-Road Driving Performance in Participants with Excessive Daytime Sleepiness Associated with Obstructive Sleep Apnea

Abstract

Excessive daytime sleepiness occurs commonly in patients with obstructive sleep apnea, which increases risk for motor vehicle accidents. This study evaluated the impact of solriamfetol, a dopamine and norepinephrine reuptake inhibitor, on on-the-road driving performance in participants with excessive daytime sleepiness associated with obstructive sleep apnea. Eligible participants were aged 21-75 years with obstructive sleep apnea and excessive daytime sleepiness (Maintenance of Wakefulness Test mean sleep latency <30 minutes and Epworth Sleepiness Scale score ≥10). Participants were randomized 1:1 to receive solriamfetol (150 mg/day for 3 days, then 300 mg/day for 4 days) or placebo for 7 days, before crossover to the other treatment for 7 days. On Day 7 of each treatment period, a standardized on-road driving test at 2 and 6 hours postdose assessed driving performance. Standard deviation of lateral position was the primary endpoint. Solriamfetol significantly reduced standard deviation of lateral position at 2 (n=34; least squares mean difference, –1.1 cm; 95% confidence interval, –1.85, –0.32; p = 0.006), and 6 hours postdose (n=32; least squares mean difference, –0.8 cm; 95% confidence interval, –1.58, –0.03; p = 0.043). At 2 hours postdose, 4 placebo-treated and 1 solriamfetol-treated participants failed to complete the driving test; at 6 hours postdose, 7 and 3 participants, respectively, failed to complete the test. The most common treatment-emergent adverse events were headache, nausea, insomnia, dizziness, and agitation. Solriamfetol 300 mg/day significantly improved on-the-road driving performance in participants with excessive daytime sleepiness associated with obstructive sleep apnea. Safety was consistent with larger clinical trials.

Vinckenbosch F.R.J., Asin J., de Vries N., Vonk P.E. , Donjacour C.E.H.M., Lammers G., Overeem S., Janssen H., Wang G., Chen D., Carter L.P., Zhou K., Vermeeren A. & Ramaekers J.G. (2022). Effects of Solriamfetol on On-the-Road Driving Performance in Participants with Excessive Daytime Sleepiness Associated with Obstructive Sleep Apnea. Human psychopharmacology: clinical and experimental

Introduction

Obstructive sleep apnea (OSA) is a sleep-related breathing disorder that is estimated to affect nearly 1 billion adults worldwide (Benjafield et al., 2019). Excessive daytime sleepiness (EDS) is a common symptom of OSA (Dongol & Williams, 2016; Pagel, 2009) and can severely impact patients' lives, causing impairments in mood, quality of life (QoL), cognitive function, work productivity, and safety (Garbarino, Guglielmi, Sanna, Mancardi, & Magnavita, 2016; Gasa et al., 2013; Mulgrew et al., 2007; Pepin et al., 2009; Stepnowsky et al., 2019; Zhou, Camacho, Tang, & Kushida, 2016). In addition, patients with OSA and EDS have ~2.5 times the risk of motor vehicle accidents compared with healthy controls (Tregear, Reston, Schoelles, & Phillips, 2009), and shorter sleep latency as measured with the Maintenance of Wakefulness Test (MWT) is significantly correlated with sleepiness-related motor vehicle accidents and near misses in patients with sleep disorders (Philip et al., 2021).

Primary OSA therapy, such as continuous positive airway pressure (CPAP), can reduce symptoms of EDS; nevertheless, persistence of EDS has been reported in 9% to 22% of patients, despite their use of CPAP (Gasa et al., 2013; Pepin et al., 2009). Pharmacologic treatment can complement primary airway therapy for the alleviation of residual EDS associated with OSA (Marra, Arnaldi, & Nobili, 2019). In laboratory studies, the wake-promoting agents (WPAs) modafinil and armodafinil (approved in the United States, but not the European Union, for the treatment of persistent EDS in patients with OSA (European Medicines Agency, 2011; Nuvigil , 2018; Provigil, 2018) have been shown to improve measures of simulated driving performance (Chapman et al., 2014; Kay & Feldman, 2013; Williams et al., 2010). In a retrospective cohort study, use of methylphenidate or modafinil was associated with a 20% reduction in the risk of hospitalization attributable to a motor vehicle accident in patients with OSA (Lin et al., 2020). However, studies demonstrating that a pharmacologic treatment can result in specific improvements in *on-the-road* driving performance in sleepy patients with OSA are lacking.

Solriamfetol (SUNOSI™) is a dopamine and norepinephrine reuptake inhibitor approved in the United States and European Union to improve wakefulness in adult patients with EDS associated with OSA (approved dose range, 37.5–150 mg/day) (Sunosi™ (solriamfetol) tablets Prescribing Information, 2019; Sunosi™ (solriamfetol) tablets Summary of Product Characteristics, 2020). Solriamfetol was investigated in participants with EDS associated with OSA in short (12 weeks) and longer-term (up to 52 weeks) clinical trials, where treatment with solriamfetol at doses ranging from 37.5 to 300 mg/day was associated with reduced EDS and improvements on measures of daily functioning, work productivity, and QoL (Malhotra et al., 2020; Schweitzer et al., 2019; Weaver et al., 2020; Weaver et al., 2019).

In parallel with the 12-week phase 3 trial, the current study was conducted to evaluate the effects of solriamfetol on on-the-road driving performance in participants with EDS associated with OSA.

Methods

This study (NCT02806895; EudraCT 2015-003930-28) was conducted from July 5, 2016 to May 28, 2019 at 4 clinical sites and 1 driving test site in the Netherlands. The study protocol was approved by the medical ethics committee of University Hospital Maastricht and Maastricht University (www.toetsingonline.nl, NL56214.068.16), and all participants provided written informed consent.

Participants

Participants were recruited from sleep clinics or clinical sites. Eligible participants were men and women aged 21 to 75 years with a diagnosis of OSA, per the *International Classification of Sleep Disorders – Third Edition* (American Academy of Sleep Medicine, 2014) and EDS, based on mean sleep latency <30 minutes over 4 trials of the MWT at screening, as well as Epworth Sleepiness Scale (ESS) score ≥10 at baseline. Other study inclusion criteria were average total nightly sleep ≥6 hours (assessed

via actigraphy and sleep diary), body mass index (BMI) 18 to <40 kg/m^2, and one of the following: use of a primary OSA therapy (e.g., positive airway pressure or oral appliance) ≥1 night/week, history of ≥1 month's attempt to use a primary OSA therapy, or history of surgical intervention for OSA. Additional criteria for inclusion were normal vision (corrected or uncorrected), possession of a valid driver's license for ≥1 year, history of driving on a regular basis, and ability to operate a vehicle with a manual transmission.

Key study exclusion criteria included an unwillingness to try to use a primary OSA therapy, occupational nighttime shift work, usual bedtime after 1:00 A.M., a clinically relevant medical or psychiatric disorder (other than OSA) associated with EDS, a history or presence of an unstable medical or psychiatric condition, or pregnancy. Additional exclusion criteria were excessive caffeine use (>8 cups of coffee/day), smoking >10 cigarettes/day, use of medication that could affect sleep–wake functions within 7 days before screening, use of a monoamine oxidase inhibitor within 14 days or 5 half-lives before screening, use of an investigational drug within 30 days or 5 half-lives before baseline, anticipated use of any of these substances during the study, or previous use of solriamfetol.

Design

A randomized, double-blind, placebo-controlled, 2-period crossover study design was used. Eligible participants were randomly assigned 1:1 to receive either solriamfetol (150 mg/day for 3 days, followed by 300 mg/day for 4 days) or placebo for 7 days (Period 1) and then cross over to the other treatment for 7 days (Period 2); there was no washout between periods. Solriamfetol 150- and 300-mg tablets and placebo tablets were supplied in identical opaque gelatin capsules to ensure adequate blinding. This study was initiated before regulatory approval or dosing recommendations were finalized. Therefore, the 300-mg/day dose used was based on prior phase 2 study data (Bogan et al., 2015; Ruoff

et al., 2016), consistent with the maximum dose used in pivotal trials of solriamfetol for patients with OSA (Malhotra et al., 2020; Schweitzer et al., 2019).

Procedures

The study included a screening/washout period of ≤28 days prior to the first dose of study treatment: Eligibility was assessed (including general safety assessments; in addition, MWT and ESS were assessed at visit 2), prohibited medications were washed out, and participants completed a practice driving test (at baseline/visit 3). On Day 7 and 14 (ie, Day 7 of each period), visits were conducted to evaluate driving performance. A safety follow-up visit was conducted approximately 1 week after completion of Period 2 (Figure 1).

Participants were instructed to take a single capsule once daily, within 1 hour of waking in the morning, on an empty stomach, and then to wait ≥30 minutes before having breakfast. On driving test days, the capsule for that day was administered at the driving test site in the presence of an investigator at 8:45 A.M. (2 hours before the start of the first drive); 30 minutes after administration, participants received a light breakfast. Throughout the study, caffeine users were instructed to not increase their use during the study, and nicotine users were instructed to maintain a consistent level of use. In addition, on driving test days, 1 cup of black coffee was permitted prior to arrival at the test site, with no additional consumption until after the second driving test; nicotine use was restricted to 1 cigarette in the morning ≥1 hour before the first MWT trial and 1 cigarette on waking on driving test days, with no other use until after the study procedures were completed on those days.

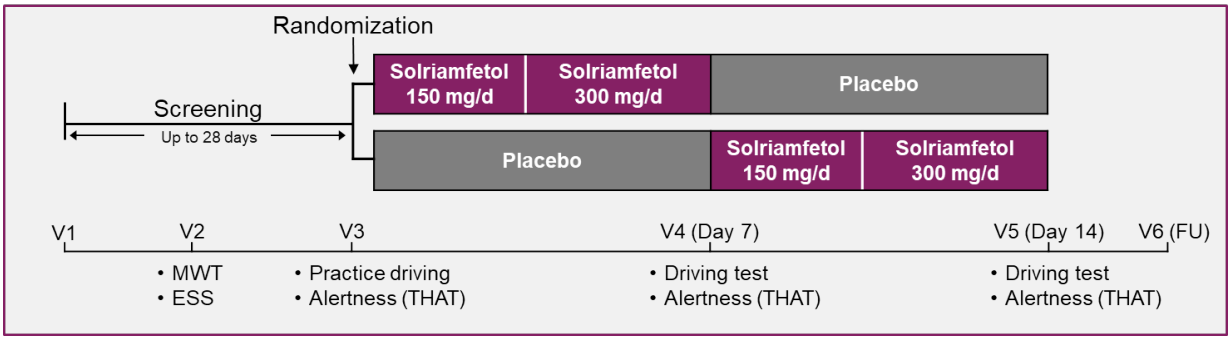

Figure 1. Study design. ESS, Epworth Sleepiness Scale; FU, follow-up; MWT, Maintenance of Wakefulness Test; THAT, Toronto Hospital Alertness Test; V, visit.

At the end of each treatment period, a standardized on-road driving test (Verster & Roth, 2011) was conducted at 2 hours and 6 hours after administration of study treatment (Figure 2). For each test (~1 hour in duration), participants drove a specially instrumented vehicle over a 100 km (~62 miles) primary highway circuit; they were accompanied by a licensed driving instructor with access to dual controls (brakes, clutch, accelerator). Participants were instructed to maintain both a steady lateral position between the delineated boundaries of the slower (right) traffic lane and a constant speed of 95 km/h (~59 mph). Participants were permitted to deviate from these instructions only to pass a slower vehicle, to respond to slower traffic ahead, or to exit and reenter the highway at the turnaround point (these events were later removed for the purposes of the analysis of driving parameters by 2 experienced raters). Vehicle speed and lateral distance to the left-lane line were continuously recorded, and the data stored on an onboard computer. The driving test could be stopped by the participant or by the accompanying driving instructor if either considered it unsafe to continue.

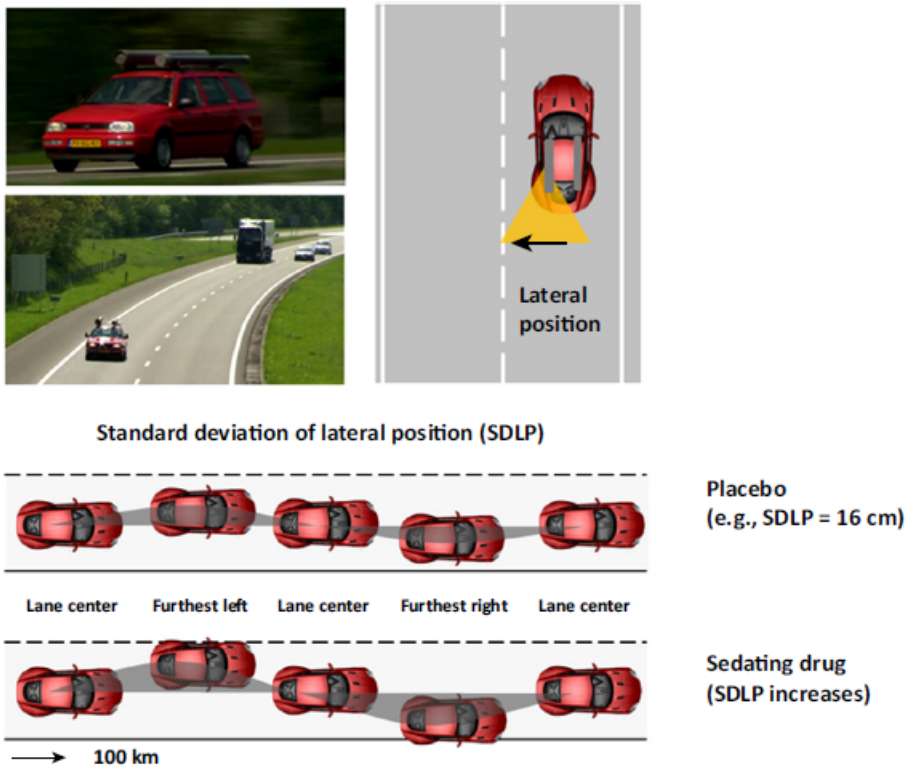

Figure 2. Standardized on-the-road driving test. The left upper panel shows the instrumented vehicle during the driving test in actual traffic on a primary highway. The lateral position of the car relative to the white middle line is continuously measured during a 1-h drive by means of a camera that is mounted on the roof of the car (right upper panel). The mean SDLP over the entire ride is calculated offline after completion of the test using signal editing software. SDLP is a measure of weaving and indicates road-tracking control of the driver. Drugs that induce sleepiness and sedation cause significant increments in weaving motion and thus loss of vehicular control. SDLP, standard deviation of lateral position.

Assessments and outcomes

The primary outcome assessment from the driving tests was standard deviation of lateral position (SDLP) in centimeters—a measure of "weaving" or road-tracking control (Ramaekers, 2017; Verster & Roth, 2011). Data were analyzed for all driving tests (completed or incomplete) with data available; for incomplete driving tests, SDLP data from the part of the test that was completed were analyzed. Standard deviation of speed and number of lane drifts (defined as deviations > 100 cm from the absolute lateral position within an 8-second window) were also determined from driving test data.

The Toronto Hospital Alertness Test (THAT) is a 10-item self-report questionnaire that measures perceived alertness over the previous week; scores can range from 0 to 50, with higher scores indicating greater alertness (Shapiro et al., 2006). This assessment was administered at baseline and on driving test days, prior to administration of study treatment.

Safety assessments included a physical examination, electrocardiogram, clinical laboratory tests, and assessment of adverse events (AEs).

Participants using a primary OSA therapy (PAP or oral appliance) at screening recorded their primary OSA therapy usage and the estimated duration of use (more than half of the night, less than half of the night, or don't know) on a daily basis.

Statistical analyses

The primary efficacy endpoint was SDLP at 2 hours postdose, and the secondary efficacy endpoints included SDLP at 6 hours postdose, percentage of participants with improved or impaired driving on solriamfetol compared with placebo, standard deviation of speed, lane drifts, and THAT score.

For the primary endpoint, the null hypothesis was that at 2 hours postdose the mean SDLP values for solriamfetol and placebo were equal; the alternative hypothesis was that they were not equal. The treatment difference in mean SDLP between solriamfetol and placebo at 2 hours postdose was tested; a 5% type I error rate ($p < 0.05$) was considered statistically significant. A sample size of 36 participants would provide 90% power to detect a mean difference of 2.0 cm on the primary outcome measure, SDLP (Ramaekers, Kuypers, & Samyn, 2006; Verster et al., 2008), assuming a standard deviation (of SDLP) of 3.25 cm and a 2-sided 0.05 significance level using paired *t* test. A study enrollment of 40 participants was planned in order to allow for dropouts. Because of logistical

challenges (eg, long distances between clinical and driving test sites), the study was completed with 34 enrollees, with an estimated power of 88.9%.

Efficacy analyses were performed with data from the modified intent-to-treat analysis population, which comprised all randomized participants who received ≥1 dose of study drug and had evaluable SDLP data at 2 hours postdose.

Mean change in SDLP was analyzed with a repeated mixed effect analysis of variance (ANOVA) model with treatment (solriamfetol, placebo), time (2 hours postdose, 6 hours postdose), treatment period, treatment sequence, and treatment × time interaction as fixed effects and participant as a random effect. The 2-sided 95% CIs for changes in SDLP with solriamfetol and placebo, based on the repeated mixed effect ANOVA model, were calculated for each driving test. The assumption of normal data was examined on the residuals from the mixed effect model using the Shapiro-Wilk normality test.

Maximum McNemar symmetry analyses (Laska, Meisner, & Wanderling, 2012) were used to detect an asymmetry in the distribution of the change in driving performance at 2 hours and 6 hours postdose. The test examined the differences in the proportions of impaired drivers and improved drivers following treatment using a generalized sign test over all relevant thresholds. Single McNemar tests were used to analyze the difference in proportions of participants taking solriamfetol with improved or impaired driving performance compared with placebo at each relevant threshold. Thresholds of 1.0, 1.5, 2.0, 2.5, and 3.5 cm were tested (Ramaekers et al., 2006; Verster et al., 2008). In comparisons of solriamfetol and placebo, improvement was defined as a decrease in SDLP in participants treated with solriamfetol compared with placebo at the threshold, and impairment was defined as an increase in SDLP at the threshold, or failure to complete the driving test while on solriamfetol because of sleepiness or safety concerns regardless of their performance while on placebo (participants who failed to

complete the driving test while on placebo but completed the test while on solriamfetol were not counted as impaired or improved).

The number of participants who failed to complete the driving test was summarized descriptively, as was the duration of the drive before stopping. Additional secondary efficacy measures (standard deviation of speed, number of lane drifts) were analyzed with an ANOVA method similar to that used for SDLP. THAT scores were analyzed using a mixed effect analysis of covariance (ANCOVA) model. No multiplicity adjustments were made in the efficacy analyses for multiple endpoints, and all *p* values are therefore nominal.

Demographic, OSA history, and safety data were summarized descriptively for the safety population, which included all participants who received ≥1 dose of study drug. No formal statistical testing was performed.

Results

Of the 59 participants who were screened, 34 met the study inclusion criteria and were enrolled. All participants received ≥1 dose of study treatment and comprised the safety population; 1 participant was withdrawn after study Period 1 and did not receive the study treatment (placebo) for Period 2. All enrolled participants were white, of non-Hispanic/Latino ethnicity, and located in the Netherlands. Participants had a mean (standard deviation [SD]) ESS score of 14.4 (3.5) and a mean (SD) MWT sleep latency of 14.3 (7.3) minutes. Actigraphy and sleep diary data showed no differences in total sleep time between placebo and solriamfetol treatment (data not shown). Twenty-nine participants were using primary OSA therapy; the remaining 5 participants had attempted CPAP use but ultimately discontinued. Use of primary OSA therapy was stable throughout the study. The mean percentage of nights that participants used primary OSA therapy for more than half the night was 95.7% at baseline,

94.6% at the end of the placebo treatment period, and 92.7% at the end of the solriamfetol treatment period (n=28 at each time point).

Characteristic	**Total (N=34)**
Age, years, mean (SD)	51.6 (12.3)
Male, n (%)	30 (88)
BMI, kg/m^2, mean (SD)	29.3 (3.9)
Use of a primary OSA therapy, n (%)	29 (85)
History of surgical intervention for OSA, n (%)	10 (29)
MWT sleep latency, min, mean (SD)	14.3 (7.3)
ESS total score, mean (SD)	14.4 (3.5)
THAT total score, mean (SD) [n=29]	26.0 (7.6)

Table 1. Demographic and Baseline Clinical Characteristics. [a]For each participant, MWT sleep latency is the average of 4 trials with nonmissing values. BMI, body mass index; ESS, Epworth Sleepiness Scale; MWT, Maintenance of Wakefulness Test; OSA, obstructive sleep apnea; SD, standard deviation; THAT, Toronto Hospital Alertness Test.

On the primary outcome measure, SDLP at 2 hours postdose, there was a statistically significant reduction with solriamfetol compared with placebo (least squares [LS] mean difference, –1.1 cm; p = 0.006; Table 2). An improvement with solriamfetol versus placebo was also observed at 6 hours postdose (LS mean difference, –0.8 cm; p = 0.043). Individual driving performance with solriamfetol versus placebo is shown in figure 2.

Time Point	Standard Deviation of Lateral Position					
	Placebo		Solriamfetol		LS Mean Difference[a]	
	n	LS Mean (SE), cm	n	LS Mean (SE), cm	(95% CI), cm	*p*[b]
2 hours postdose[c]	33	19.9 (0.63)	34	18.8 (0.63)	–1.1 (–1.85, –0.32)	0.006
6 hours postdose	32	20.0 (0.63)	32	19.2 (0.63)	–0.8 (–1.58, –0.03)	0.043

Table 2. Analysis of Standard Deviation of Lateral Position. [a]Solriamfetol – placebo. [b]Repeated mixed effect analysis of variance (ANOVA). [c]Primary endpoint. CI, confidence interval; LS, least squares; SE, standard error.

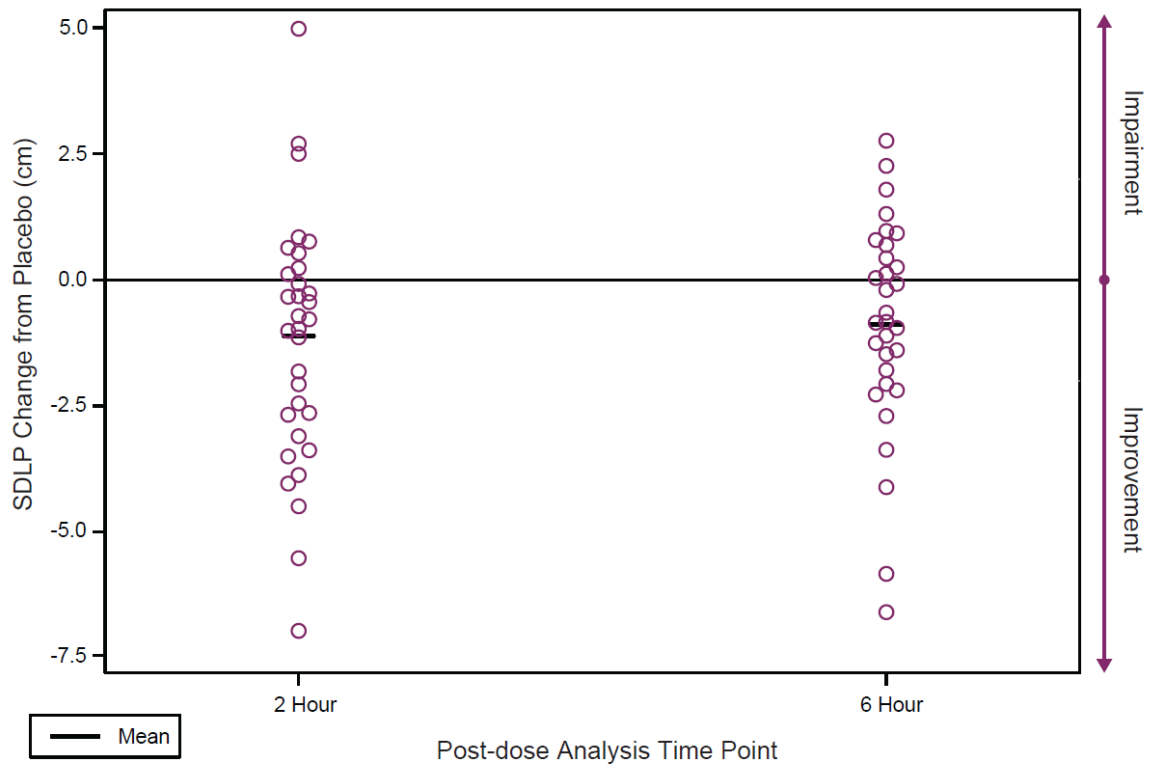

Figure 2. Individual driving performance with solriamfetol compared with placebo. Data for all participants, including those with incomplete driving tests. For participants who did not complete the driving test, data for the part of the drive that was completed were used to calculated SDLP. SDLP, standard deviation of lateral position.

Eight participants stopped ≥1 driving test prematurely. More participants had incomplete tests when receiving placebo compared with solriamfetol at 2 hours postdose and 6 hours postdose (Table 3). Specifically, 7 participants failed to complete ≥1 test while on placebo, and 3 failed to complete ≥1 test while on solriamfetol; 2 participants failed to complete ≥1 test on both treatments. The duration of incomplete drives ranged from 11 to 53 minutes on placebo and 28 to 51 minutes on solriamfetol. None of the participants receiving solriamfetol had their driving test halted by the instructor, compared with 2 participants receiving placebo at each time point. Overall numerically higher percentages of participants had improvements on solriamfetol at all thresholds examined at both time points. However, the maximum McNemar test did not show asymmetry at either 2 hours (Figure 3) or 6 hours postdose.

	Placebo	**Solriamfetol**
Incomplete tests, N (n stopped by participant, n stopped by instructor)		
2 hours	4 (2, 2)	1 (1, 0)
6 hours	7 (5, 2)	3 (3, 0)
Participants, n, with incomplete tests[a]	7[b]	3[b]
Duration of drive before stopping, min		
Mean (SD)	33.5 (14.8)	39.0 (9.8)
Median (interquartile range)	32.0 (21, 45)	38.5 (31.5, 46.5)

Table 3. Incomplete Driving Tests. [a]Eight participants had incomplete tests; 5 of these participants had multiple incomplete tests (ie, on both treatments and/or at multiple time points); of the 3 who had a single incomplete test, 2 had an incomplete test on placebo (both at the 6-hour time point), and 1 had an incomplete test on solriamfetol (at the 6-hour time point). [b]Two of these participants had ≥1 incomplete test on solriamfetol and ≥1 on placebo.

Threshold, Difference in SDLP	Improved, n (%)	Impaired, n (%)	% Improved – % Impaired (95% CI)[a]
1.0 cm	15 (44.1)	5 (14.7)	29.4
1.5 cm	13 (38.2)	5 (14.7)	23.5
2.0 cm	12 (35.3)	5 (14.7)	20.6
2.5 cm	10 (29.4)	4 (11.8)	17.6
3.0 cm	8 (23.5)	2 (5.9)	17.6
3.5 cm	5 (14.7)	2 (5.9)	8.8

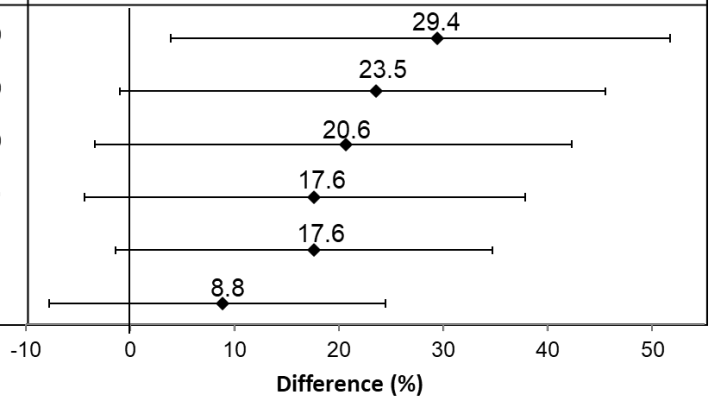

Figure 3. Percentage of participants with improved versus impaired driving performance with solriamfetol treatment compared with placebo at 2 hours postdose. Improvement and impairment based on difference in SDLP for solriamfetol versus placebo at each threshold. Percentages of participants with improvement and impairment were compared using a McNemar test to detect an asymmetry in the distribution of the change in driving performance. [a]All nominal $p > 0.05$, except nominal $p = 0.041$ at 1.0 cm at 2 hours postdose. CI, confidence interval; SDLP, standard deviation of lateral position.

Secondary measures of driving performance—standard deviation of speed and lane drifts—were not different between solriamfetol and placebo at either time point (Table 4). THAT scores at the end of the treatment period were higher (indicating greater alertness) for participants receiving solriamfetol than for participants receiving placebo (27.5 vs 23.9; LS mean difference, 3.6; $p = 0.024$).

Post hoc analyses were performed to examine the relationship between baseline ESS scores and MWT sleep latency and SDLP at 2 and 6 hours postdose with either treatment. Pearson correlations ranged from –0.22 to 0.14 (all $p > 0.05$), indicating no correlation between either measure of sleepiness at baseline and SDLP at the 2- or 6-hour time point for solriamfetol or placebo.

Time Point	Placebo		Solriamfetol		LS Mean Difference[a] (95% CI)	*p*[b]
	n	LS Mean (SE)	n	LS Mean (SE)		
Standard deviation of speed, km/h						
2 hours postdose	33	2.55 (0.10)	34	2.62 (0.10)	0.1 (–0.10, 0.23)	0.412
6 hours postdose	32	2.84 (0.10)	32	2.73 (0.10)	–0.1 (–0.28, 0.06)	0.199
Number of lane drifts						
2 hours postdose	33	2.89 (0.57)	34	1.76 (0.56)	–1.1 (–2.40, 0.14)	0.081
6 hours postdose	32	2.07 (0.57)	32	2.12 (0.57)	0.050 (–1.25, 1.35)	0.939
THAT (higher total score indicates greater alertness)						
End of treatment period	33	23.9 (1.2)	34	27.5 (1.2)	3.6 (0.50, 6.66)	0.024

Table 4. Additional Secondary Endpoints. [a]Solriamfetol – placebo. [b]*p* values are nominal; standard deviation of speed and number of lane drifts were analyzed using a repeated mixed effect analysis of variance (ANOVA) model; THAT scores were analyzed using a mixed effect analysis of covariance (ANCOVA) model. CI, confidence interval; LS, least squares; SE, standard error; THAT, Toronto Hospital Alertness Test.

Treatment-emergent adverse events (TEAEs) were reported in approximately two-thirds of participants overall. The majority of TEAEs were mild to moderate in severity, and none led to study drug interruption or withdrawal. There were no serious TEAEs or deaths. The most common TEAEs were headache, nausea, insomnia, dizziness, and agitation (Table 5).

TEAE, n (%)	**Placebo (n=33)**	**Solriamfetol (n=34)**
Participants with any TEAE	11 (33.3)	17 (50.0)
TEAEs leading to discontinuation	0	0
Common TEAEs[a]		
Headache	4 (12.1)	5 (14.7)
Nausea	2 (6.1)	4 (11.8)
Insomnia	0	4 (11.8)
Dizziness	2 (6.1)	3 (8.8)
Agitation	1 (3.0)	1 (2.9)

Table 5. Treatment-Emergent Adverse Events [a]Incidence ≥5% overall. TEAE, treatment-emergent adverse event.

Discussion

This double-blind, crossover study evaluated the effect of solriamfetol treatment on driving performance in participants with EDS associated with OSA. Participants received 7 days of treatment and undertook an on-road driving performance test 2 and 6 hours after dosing. Solriamfetol (150 mg/day for 3 days followed by 300 mg/day for 4 days) significantly improved SDLP, an important measure of driving performance, at both time points compared with placebo. Fewer participants

completed the driving test on placebo than solriamfetol at both time points. Additionally, a numerically greater percentage of participants had improved SDLP than impaired SDLP with solriamfetol compared with placebo at 2 hours postdose.

Fifteen (11.5%) of 131 tests were stopped because the instructor or participant considered it unsafe to continue. This happened more frequently than in comparable studies assessing sedating drugs in healthy volunteers (3.1%) (Verster & Roth, 2012). Most incomplete tests in this study were stopped under placebo treatment (n=11/15, 73.3%) and at the participant's request (n=11/15, 73.3%; **Table 3**). In contrast, during the aforementioned studies, 3 to 4 times more tests were stopped by the instructor than the participant (Verster & Roth, 2012). This suggests that participants in our study were often aware of their potential impairment and careful to avoid further risks.

The clinical relevance of the SDLP improvement can be interpreted by comparing observed SDLP values with normative data. A study of 74 healthy participants yielded a mean (SE) SDLP of 18.19 (0.46) cm with a 2-sided 95% CI upper bound of 19.09 cm (Vinckenbosch et al., 2021). LS mean SDLP values with placebo at 2 and 6 hours postdose in our study were 19.9 cm and 20.0 cm, respectively, indicating that sleepy participants with OSA receiving placebo were impaired relative to a healthy population. LS mean SDLP values 2 hours post-solriamfetol administration (18.8 cm) fell within the aforementioned 95% CI, suggesting group-level normal road-tracking performance, while LS mean SDLP 6 hours post-solriamfetol (19.2 cm) remained outside the 95% CI.

The on-road driving test is the gold standard for assessing drug-induced changes in driving (Jongen et al., 2017). However, studies with other WPAs in participants with OSA have examined only simulated driving (Chapman et al., 2014; Kay & Feldman, 2013; Williams et al., 2010). One such study in participants with OSA before CPAP initiation (Kay & Feldman, 2013) showed greater improvement with armodafinil (150 mg/day) than placebo on the Driving Safety Score (mean z-score derived from

predefined safety elements, including out-of-lane driving and lane position deviation [ie, SDLP]). The absence of on-road driving data with WPAs limits comparisons of the current results to previous studies.

Consistent with its established impact on EDS, solriamfetol treatment was associated with higher THAT scores compared with placebo—indicating greater alertness. In a 12-week phase 3 study of participants with OSA, solriamfetol at doses up to 300 mg/day significantly improved MWT sleep latency, decreased ESS scores, increased the percentage of participants reporting overall improvement on the Patient Global Impression of Change scale at 12 weeks, and improved measures of daily functioning, work productivity, and QoL compared with placebo (Schweitzer et al., 2019; Weaver et al., 2020). These improvements in wakefulness and functional outcomes were maintained for up to 52 weeks in an open-label long-term extension study (Malhotra et al., 2020; Weaver et al., 2019). Inclusion criteria for the current study were similar to those used in phase 3 studies; likewise, mean baseline MWT latency (14.3 minutes) and ESS scores (14.4) in this study were similar to those reported in the 12-week phase 3 study (MWT, 12.0-13.6 minutes; ESS score, 14.8-15.6) (Schweitzer et al., 2019). Here, the MWT and ESS were assessed only at screening/baseline to confirm eligibility; treatment effects were not examined. Baseline MWT and ESS scores did not correlate with SDLP at any time point, indicating solriamfetol's beneficial effect did not depend on baseline levels of sleepiness.

Solriamfetol's safety profile in this study was consistent with the larger 12-week and 52-week studies (Malhotra et al., 2020; Schweitzer et al., 2019). Most TEAEs were mild or moderate in severity. Headache, nausea, insomnia, and dizziness occurred more frequently with solriamfetol than placebo. No TEAEs were serious or led to treatment/study discontinuation.

Limitations include the fact that the tested dose of solriamfetol (300 mg/day) exceeds the highest recommended dose (150 mg/day). As previously noted, this study was conducted before regulatory approvals of solriamfetol in OSA, and dosing was based on previous and ongoing studies at

the time this study was designed (Bogan et al., 2015; Ruoff et al., 2016; Schweitzer et al., 2019). Thus, it is unknown how the magnitude of functional improvements observed at 300 mg/day would translate to the highest approved dose (150 mg/day) in clinical practice. However, efficacy of the 150-mg and 300-mg doses in the overall OSA population in the phase 3 study was similar (Schweitzer et al., 2019). While SDLP is linked to accident risk (Ramaekers, 2017), how the functional improvements observed here might affect accident risk is unknown, as the study was not designed to directly assess this. Additionally, long-term effects on driving performance were not assessed. An open-label extension study indicated that solriamfetol's wake-promoting effects are maintained for up to 1 year (Malhotra et al., 2020); it is reasonable to expect improved driving performance would also be maintained.

Strengths of the study include the fact that participants' baseline characteristics reflected real-world OSA populations (eg, primarily male; mean age, ~51 years; mean BMI, ~29 kg/m^2) (Bailly et al., 2016; Tkacova et al., 2014). Additionally, the driving test was conducted in on-the-road traffic. The crossover design eliminated between-group differences in participant characteristics (eg, BMI, apnea-hypopnea index, hypoxemia) that predict motor vehicle accidents in drivers with OSA (Tregear et al., 2009).

For participants with EDS associated with OSA, solriamfetol treatment was associated with significant improvement in on-the-road driving performance, as assessed by SDLP at 2 hours and 6 hours postdose; additional secondary outcome measures (THAT scores) also indicated greater alertness compared with placebo. These findings demonstrate that solriamfetol's wake-promoting efficacy observed across multiple clinical trials (Malhotra et al., 2020; Schweitzer et al., 2019) is also associated with improved real-world functional performance in this study.

References

American Academy of Sleep Medicine. (2014). *International Classification of Sleep Disorders* (3rd ed.). Darien, IL: American Academy of Sleep Medicine.

Bailly, S., Destors, M., Grillet, Y., Richard, P., Stach, B., Vivodtzev, I., . . . Pepin, J. L. (2016). Obstructive sleep apnea: a cluster analysis at time of diagnosis. *PloS One, 11*(6), e0157318. doi:10.1371/journal.pone.0157318

Benjafield, A. V., Ayas, N. T., Eastwood, P. R., Heinzer, R., Ip, M. S. M., Morrell, M. J., . . . Malhotra, A. (2019). Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet. Respir. Med., 7*(8), 687-698. doi:10.1016/s2213-2600(19)30198-5

Bogan, R. K., Feldman, N., Emsellem, H. A., Rosenberg, R., Lu, Y., Bream, G., . . . Lankford, D. A. (2015). Effect of oral JZP-110 (ADX-N05) treatment on wakefulness and sleepiness in adults with narcolepsy. *Sleep Medicine, 16*(9), 1102-1108. doi:10.1016/j.sleep.2015.05.013

Chapman, J. L., Kempler, L., Chang, C. L., Williams, S. C., Sivam, S., Wong, K. K., . . . Marshall, N. S. (2014). Modafinil improves daytime sleepiness in patients with mild to moderate obstructive sleep apnoea not using standard treatments: a randomised placebo-controlled crossover trial. *Thorax, 69*(3), 274-279. doi:10.1136/thoraxjnl-2013-203796

Dongol, E. M., & Williams, A. J. (2016). Residual excessive sleepiness in patients with obstructive sleep apnea on treatment with continuous positive airway pressure. *Current Opinion in Pulmonary Medicine, 22*(6), 589-594. doi:10.1097/mcp.0000000000000324

European Medicines Agency. (2011). Modafinil - Article 31 referral - Annex I, II, III, IV. Retrieved from https://www.ema.europa.eu/documents/referral/modafinil-article-31-referral-annex-i-ii-iii-iv_en.pdf

Garbarino, S., Guglielmi, O., Sanna, A., Mancardi, G. L., & Magnavita, N. (2016). Risk of occupational accidents in workers with obstructive sleep apnea: systematic review and meta-analysis. *Sleep, 39*(6), 1211-1218. doi:10.5665/sleep.5834

Gasa, M., Tamisier, R., Launois, S. H., Sapene, M., Martin, F., Stach, B., . . . Pneumology-FFP, S. C. o. t. S. R. o. t. F. F. o. (2013). Residual sleepiness in sleep apnea patients treated by continuous positive airway pressure. *Journal of Sleep Research, 22*(4), 389-397. doi:10.1111/jsr.12039

Jongen, S., Vermeeren, A., van der Sluiszen, N. N., Schumacher, M. B., Theunissen, E. L., Kuypers, K. P., . . . Ramaekers, J. G. (2017). A pooled analysis of on-the-road highway driving studies in actual traffic measuring standard deviation of lateral position (i.e., "weaving") while driving at a blood alcohol concentration of 0.5 g/L. *Psychopharmacology, 234*(5), 837-844. doi:10.1007/s00213-016-4519-z

Kay, G. G., & Feldman, N. (2013). Effects of armodafinil on simulated driving and self-report measures in obstructive sleep apnea patients prior to treatment with continuous positive airway pressure. *Journal of Clinical Sleep Medicine, 9*(5), 445-454. doi:10.5664/jcsm.2662

Laska, E., Meisner, M., & Wanderling, J. (2012). A maximally selected test of symmetry about zero. *Statistics in Medicine, 31*(26), 3178-3191. doi:10.1002/sim.5384

Lin, Y. C., Chen, T. Y., Chien, W. C., Chung, C. H., Chang, H. A., Kao, Y. C., . . . Tzeng, N. S. (2020). Stimulants associated with reduced risk of hospitalization for motor vehicle accident injury in patients with obstructive sleep apnea-a nationwide cohort study. *BMC Pulmonary Medicine, 20*(1), 28. doi:10.1186/s12890-019-1041-1

Malhotra, A., Shapiro, C., Pepin, J. L., Hedner, J., Ahmed, M., Foldvary-Schaefer, N., . . . Schwab, R. (2020). Long-term study of the safety and maintenance of efficacy of solriamfetol (JZP-110) in

the treatment of excessive sleepiness in participants with narcolepsy or obstructive sleep apnea. *Sleep., 43*(2), zsz220. doi:10.1093/sleep/zsz220

Marra, S., Arnaldi, D., & Nobili, L. (2019). The pharmacotherapeutic management of obstructive sleep apnea. *Expert Opinion on Pharmacotherapy, 20*(16), 1981-1991. doi:10.1080/14656566.2019.1652271

Mulgrew, A. T., Ryan, C. F., Fleetham, J. A., Cheema, R., Fox, N., Koehoorn, M., . . . Ayas, N. T. (2007). The impact of obstructive sleep apnea and daytime sleepiness on work limitation. *Sleep Medicine, 9*(1), 42-53. doi:10.1016/j.sleep.2007.01.009

Nuvigil [package insert]. (2018). North Wales, PA: Teva Pharmaceuticals.

Pagel, J. F. (2009). Excessive daytime sleepiness. *American Family Physician, 79*(5), 391-396. Retrieved from https://www.aafp.org/afp/2009/0301/p391.pdf

Pepin, J. L., Viot-Blanc, V., Escourrou, P., Racineux, J. L., Sapene, M., Levy, P., . . . Mallart, A. (2009). Prevalence of residual excessive sleepiness in CPAP-treated sleep apnoea patients: the French multicentre study. *European Respiratory Journal, 33*(5), 1062-1067. doi:10.1183/09031936.00016808

Philip, P., Guichard, K., Strauss, M., Léger, D., Pepin, E., Arnulf, I., . . . Dauvilliers, Y. (2021). Maintenance of Wakefulness Test: how does it predict accident risk in patients with sleep disorders? *Sleep Medicine, 77*249-255. doi:10.1016/j.sleep.2020.04.007

Provigil [package insert]. (2018). North Wales, PA: Teva Pharmaceuticals.

Ramaekers, J. G. (2017). Drugs and driving research in medicinal drug development. *Trends in Pharmacological Sciences, 38*(4), 319-321. doi:10.1016/j.tips.2017.01.006

Ramaekers, J. G., Kuypers, K. P., & Samyn, N. (2006). Stimulant effects of 3,4-methylenedioxymethamphetamine (MDMA) 75 mg and methylphenidate 20 mg on actual driving during intoxication and withdrawal. *Addiction, 101*(11), 1614-1621. doi:10.1111/j.1360-0443.2006.01566.x

Ruoff, C., Swick, T. J., Doekel, R., Emsellem, H. A., Feldman, N. T., Rosenberg, R., . . . Black, J. (2016). Effect of oral JZP-110 (ADX-N05) on wakefulness and sleepiness in adults with narcolepsy: a phase 2b study. *Sleep, 39*(7), 1379-1387. doi:10.5665/sleep.5968

Schweitzer, P. K., Rosenberg, R., Zammit, G. K., Gotfried, M., Chen, D., Carter, L. P., . . . Strohl, K. P. (2019). Solriamfetol for excessive sleepiness in obstructive sleep apnea (TONES 3): a randomized controlled trial. *American Journal of Respiratory and Critical Care Medicine, 199*(11), 1421-1431. doi:10.1164/rccm.201806-1100OC

Shapiro, C. M., Auch, C., Reimer, M., Kayumov, L., Heslegrave, R., Huterer, N., . . . Devins, G. M. (2006). A new approach to the construct of alertness. *Journal of Psychosomatic Research, 60*(6), 595-603. doi:10.1016/j.jpsychores.2006.04.012

Stepnowsky, C., Sarmiento, K. F., Bujanover, S., Villa, K. F., Li, V. W., & Flores, N. M. (2019). Comorbidities, health-related quality of life, and work productivity among people with obstructive sleep apnea with excessive sleepiness: findings from the 2016 US National Health and Wellness Survey. *Journal of Clinical Sleep Medicine, 15*(2), 235-243. doi:10.5664/jcsm.7624

Sunosi™ (solriamfetol) tablets Prescribing Information. (2019). Palo Alto, CA: Jazz Pharmaceuticals, Inc.

Sunosi™ (solriamfetol) tablets Summary of Product Characteristics. (2020). Dublin, Ireland: Jazz Pharmaceuticals Ireland Ltd.

Tkacova, R., McNicholas, W. T., Javorsky, M., Fietze, I., Sliwinski, P., Parati, G., . . . Hedner, J. (2014). Nocturnal intermittent hypoxia predicts prevalent hypertension in the European Sleep Apnoea Database cohort study. *European Respiratory Journal, 44*(4), 931-941. doi:10.1183/09031936.00225113

Tregear, S., Reston, J., Schoelles, K., & Phillips, B. (2009). Obstructive sleep apnea and risk of motor vehicle crash: systematic review and meta-analysis. *Journal of Clinical Sleep Medicine, 5*(6), 573-581. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2792976/pdf/jcsm.5.6.573.pdf

Verster, J. C., Bekker, E. M., de Roos, M., Minova, A., Eijken, E. J., Kooij, J. J., . . . Volkerts, E. R. (2008). Methylphenidate significantly improves driving performance of adults with attention-deficit hyperactivity disorder: a randomized crossover trial. *J Psychopharmacol., 22*(3), 230-237. doi:10.1177/0269881107082946

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. doi:10.2147/ijgm.s19639

Verster, J. C., & Roth, T. (2012). The prevalence and nature of stopped on-the-road driving tests and the relationship with objective performance impairment. *Accident Analysis and Prevention, 45*498-506. doi:10.1016/j.aap.2011.09.003

Vinckenbosch, F. R. J., Vermeeren, A., Vuurman, E. F. P. M., van der Sluiszen, N. N. J. J. M., Verster, J. C., E., v. d. L. A. J. A., . . . Ramaekers, J. G. (2021). An explorative approach to understanding individual differences in driving performance and neurocognition in long-term benzodiazepine users. *Hum Psychopharmacol., 36*(4), e2778. doi:10.1002/hup.2778

Weaver, T. E., Drake, C. L., Benes, H., Stern, T., Maynard, J., Thein, S. G., . . . Bogan, R. K. (2020). Effects of solriamfetol on quality of life measures from a 12-week phase 3 randomized trial. *Annals of the American Thoracic Society., 17*(8), 998-1007. doi:10.1513/AnnalsATS.202002-136OC

Weaver, T. E., Pepin, J. L., Schwab, R., Shapiro, C., Hedner, J., Ahmed, M., . . . Malhotra, A. (2019). Long-term effects of solriamfetol on quality of life in participants with excessive daytime sleepiness associated with narcolepsy or obstructive sleep apnea [abstract 0601]. *Sleep., 42(suppl 1)*A239.

Williams, S. C., Marshall, N. S., Kennerson, M., Rogers, N. L., Liu, P. Y., & Grunstein, R. R. (2010). Modafinil effects during acute continuous positive airway pressure withdrawal: a randomized crossover double-blind placebo-controlled trial. *American Journal of Respiratory and Critical Care Medicine, 181*(8), 825-831. doi:10.1164/rccm.200908-1307OC

Zhou, J., Camacho, M., Tang, X., & Kushida, C. A. (2016). A review of neurocognitive function and obstructive sleep apnea with or without daytime sleepiness. *Sleep Medicine, 23*99-108. doi:10.1016/j.sleep.2016.02.008

Effects of solriamfetol on on-the-road driving in participants with narcolepsy: A randomised crossover trial

Abstract

The goal of the current investigation was to evaluate the impact of solriamfetol, a dopamine and norepinephrine reuptake inhibitor, on on-the-road driving performance in participants with narcolepsy. In this randomised, double-blind, placebo-controlled, crossover study, driving performance during a 1-hour on-road driving test was assessed at 2 hours and 6 hours post-dose following 7 days of treatment with solriamfetol (150 mg/day × 3, then 300 mg/day × 4) or placebo. The primary endpoint was standard deviation of lateral position (SDLP) at 2 hours post-dose. The study included 24 participants (54% male; mean age, 40 years); 22 had evaluable SDLP data. At 2 hours post-dose, median SDLP was significantly lower (improved) with solriamfetol compared with placebo (19.08 vs. 20.46 cm [median difference, −1.9 cm], $p = 0.002$). Four participants on solriamfetol and 7 on placebo had incomplete driving tests. At 6 hours post-dose, median SDLP was not statistically significantly different with solriamfetol compared with placebo (19.59 vs. 19.78 cm [median difference, −1.1 cm], $p = 0.125$). Three participants on solriamfetol and 10 on placebo had incomplete driving tests. Common adverse events (≥5%) included headache, decreased appetite, and somnolence.

Vinckenbosch F., Lammers G., Overeem S., Chen D., Wang G., Carter L.P., Zhou K., Ramaekers J.G. & Vermeeren A. (under review) Effects of solriamfetol on on-the-road driving in participants with narcolepsy: A randomised crossover trial. Human psychopharmacology: clinical and experimental

Introduction

Narcolepsy is a chronic neurological disorder characterised by excessive daytime sleepiness (EDS) (Kornum et al., 2017; Szabo, Thorpy, Mayer, Peever, & Kilduff, 2019). Patients with narcolepsy often experience negative effects on daily functioning (Flores, Villa, Black, Chervin, & Witt, 2016), including impaired driving performance (Findley et al., 1995; Kotterba et al., 2004). Patients with narcolepsy are also at higher risk for motor vehicle accidents (MVAs) and resulting hospitalisations (Liu, Perez, & Lau, 2018; Philip et al., 2010; Pizza et al., 2015; Tzeng et al., 2019). For example, in a case–control study of MVAs occurring during the preceding year, the odds of having any MVA were ~3 times greater (and the odds of sleepiness-related MVA >8 times greater) in drivers with narcolepsy or hypersomnia compared with controls (Philip et al., 2010). Experimental evidence suggests that treatment with modafinil improves some measures of on-road (Philip et al., 2014) and simulated driving (Kotterba et al., 2004; Sagaspe et al., 2019) performance in patients with narcolepsy or hypersomnia. In addition, two epidemiologic studies showed that long-term treatment with modafinil or psychostimulants reduced the risk for MVAs (Pizza et al., 2015; Tzeng et al., 2019). While reduced sleep latency, as measured with the Maintenance of Wakefulness Test, has been shown to be significantly correlated with sleepiness-related MVAs and near misses in a population of patients with diverse sleep disorders (Philip et al., 2021), a reliable predictor of fitness to drive in patients with narcolepsy specifically is still lacking.

Solriamfetol (SUNOSI™, Jazz Pharmaceuticals, Palo Alto, CA) is a dopamine and norepinephrine reuptake inhibitor approved in the US and EU to improve wakefulness in adults with EDS associated with narcolepsy (75–150 mg/day) or obstructive sleep apnoea (OSA; 37.5–150 mg/day) (Sunosi™ (solriamfetol) tablets Prescribing Information, 2019; Sunosi™ (solriamfetol) tablets Summary of Product Characteristics, 2020). In short- (12 weeks) and long- (up to 52 weeks) term clinical trials in participants with narcolepsy, solriamfetol at doses ranging from 75 to 300 mg/day reduced EDS and

improved measures of daily functioning, work productivity, and quality of life (Emsellem et al., 2020; Malhotra et al., 2020; Thorpy et al., 2019; Weaver et al., 2019).

As few randomised controlled trials have evaluated on-the-road driving performance in this population, this study was conducted to evaluate the effects of solriamfetol on on-the-road driving performance in participants with narcolepsy.

Methods

Participants

Participants were recruited from sleep clinics or clinical sites. Eligible participants were men and women aged 21 to 75 years with a diagnosis of narcolepsy, per the *International Classification of Sleep Disorders – Third Edition* (American Academy of Sleep Medicine, 2014) or the *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition* (American Psychiatric Association, 2013). Other study inclusion criteria were average total nightly sleep ≥6 hours (as verified through actigraphy and sleep diaries), body mass index (BMI) 18 to <40 kg/m^2, normal vision (corrected or uncorrected), possession of a valid driver's license for ≥1 year, history of driving on a regular basis, and ability to operate a vehicle with a manual transmission.

Key exclusion criteria included occupational nighttime shift work, usual bedtime after 1:00 A.M., clinically relevant medical or psychiatric disorders (other than narcolepsy) associated with EDS, history or presence of unstable medical or psychiatric conditions, pregnancy, previous use of solriamfetol, excessive caffeine use (>8 cups of coffee/day), or smoking >10 cigarettes/day. Use of medications that affect sleep-wake functions was prohibited during the study and required a washout period prior to the first dose of study treatment (stimulants or alerting agents, 3 days; sodium oxybate, 7 days; and other medications that could affect sleep-wake functions or monoamine oxidase inhibitors, 14 days or 5 half-lives).

Study design

This was a randomised, double-blind, placebo-controlled, 2-period crossover study of solriamfetol in participants with narcolepsy. Treatment periods consisted of 7 days of placebo or 7 days of solriamfetol (150 mg/day for 3 days, then 300 mg/day for 4 days); there was no washout between periods. This study was initiated before regulatory approval of solriamfetol or dosing recommendations were finalised; therefore, the 300-mg/day dose used here was based on prior phase 2 studies (Bogan et al., 2015; Ruoff et al., 2016) and is consistent with the maximum dose used in phase 3 trials of solriamfetol (Malhotra et al., 2020; Thorpy et al., 2019).

Randomisation and blinding

Eligible participants were randomly assigned 1:1 to one of 2 treatment sequences: solriamfetol followed by placebo (solriamfetol/placebo) or placebo followed by solriamfetol (placebo/solriamfetol) (Figure 1). Randomisation was performed by the investigator with an interactive response technology system; assignment to one treatment sequence or the other followed a blocked randomisation schedule generated by a statistician (not involved in the analysis of the study data) before the start of the study.

Solriamfetol 150- and 300-mg tablets and placebo tablets were supplied in identical opaque gelatin capsules to ensure adequate blinding. All study personnel were blinded to study treatments.

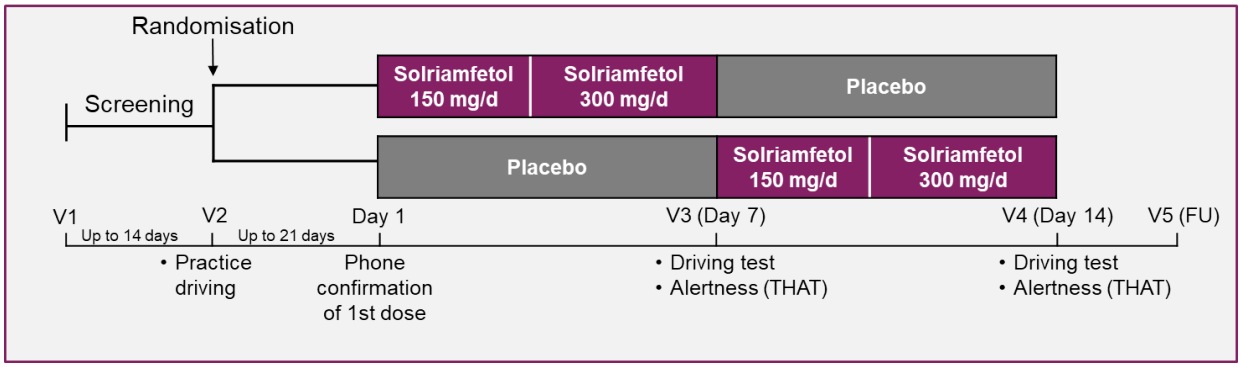

Figure 1. Study design. FU, follow-up; THAT, Toronto Hospital Alertness Test; V, visit.

Procedures

The study included a screening/washout period of ≤5 weeks prior to the first dose of study treatment, during which eligibility was assessed (including general safety assessments), prohibited medications were washed out, and participants completed a practice driving test. Eligible participants were randomized and started taking study drug at home. Participants were contacted by telephone 2 days prior to starting study treatment and on Day 1 of Period 1 to confirm their first dose of study treatment. On Day 7 and 14 (i.e. Day 7 of each period), visits were conducted to evaluate driving performance. A safety follow-up visit was conducted approximately 1 week after completion of Period 2.

Participants were instructed to take a single capsule orally once daily, within 1 hour of waking in the morning, on an empty stomach, and then to wait ≥30 minutes before having breakfast. On driving test days, the capsule for that day was administered at the driving test site in the presence of an investigator at 8:45 A.M. (2 hours before the start of the first drive); 30 minutes after administration, participants received a light breakfast. Throughout the study, caffeine users were instructed to not increase their use during the study, and nicotine users were instructed to maintain a consistent level of use. In addition, on driving test days, 1 cup of black coffee was permitted prior to arrival at the test site, with no additional consumption until after the second driving test, and nicotine use was restricted to 1 cigarette on waking, with no other use until after the study procedures were completed on those days.

At the end of each treatment period, a standardized on-road driving test (Verster & Roth, 2011) was conducted at 2 hours and at 6 hours after administration of drug or placebo (**Figure 1**). For each test (~1 hour in duration), participants drove a specially instrumented vehicle over a 100 km (~62 miles) primary highway circuit; they were accompanied by a licensed driving instructor with access to dual controls (brakes, clutch, accelerator). Participants were instructed to maintain both a steady lateral position between the delineated boundaries of the slower (right) traffic lane and a

constant speed of 95 km/h (~59 mph). Participants were permitted to deviate from these instructions only to pass a slower vehicle, to respond to slower traffic ahead, or to exit and reenter the highway at the turnaround point (these events were removed for the purpose of data analysis by 2 experienced editors of the driving data). Vehicle speed and lateral distance to the left-lane line were continuously recorded, and the data were stored on an onboard computer. The driving test could be stopped by the participant or by the accompanying driving instructor if either considered it unsafe to continue.

Assessments and outcomes

The primary outcome assessment from the driving tests was standard deviation of lateral position (SDLP) in centimeters—a measure of "weaving" or road-tracking control (Ramaekers, 2017; Verster & Roth, 2011). For participants who did not complete the driving test, SDLP data from the part of the test that was completed were analysed. Standard deviation of speed and number of lane drifts (defined as deviations >100 cm from the absolute lateral position within an 8-second window) were also determined from driving test data.

The Toronto Hospital Alertness Test (THAT) is a 10-item self-report questionnaire that measures perceived alertness over the previous week; scores can range from 0 to 50, with higher scores indicating greater alertness (Shapiro et al., 2006). This assessment was administered on driving test days, prior to administration of study treatment.

Safety assessments included a physical examination, ECG, clinical laboratory tests, and assessment of adverse events (AEs).

Statistical analyses

The primary efficacy endpoint was SDLP at 2 hours post-dose; secondary efficacy endpoints included SDLP at 6 hours post-dose, percentages of participants with improved or impaired driving on solriamfetol compared with placebo, standard deviation of speed, lane drifts, and THAT score.

For the primary endpoint, the null hypothesis was that mean SDLP with solriamfetol and mean SDLP with placebo were equal; the alternative hypothesis was that they were not equal. The treatment difference in mean SDLP between solriamfetol and placebo at 2 hours post-dose was tested; a 5% type I error rate ($p < 0.05$) was considered statistically significant. A sample size of 30 participants would provide 90% power to detect a mean difference of 2.0 cm on the primary outcome measure of SDLP (Ramaekers, Kuypers, & Samyn, 2006; Verster et al., 2008), assuming an SD of 3.0 cm (Verster et al., 2008) and a 2-sided 0.05 significance level using a paired *t*-test. To account for 10% dropouts without evaluable SDLP data, a sample size of 33 participants was planned. Due to slow recruitment, the study was ended prior to reaching the planned enrollment (post hoc calculations based on number of enrolled participants indicated an estimated power of ~78%).

Efficacy analyses were performed with data from the modified intent-to-treat analysis population, which comprised all randomised participants who received ≥1 dose of study drug and had evaluable SDLP data at 2 hours post-dose.

Change in SDLP was analysed with a repeated mixed-effects analysis of variance (ANOVA). Normality assumption was examined on the mixed effect model residuals using the Shapiro-Wilk normality test; it was observed that change in SDLP did not meet the normality assumption, and therefore the Wilcoxon signed rank test was used to compare the pairwise treatment differences.

Maximally selected McNemar symmetry analyses (Laska, Meisner, & Wanderling, 2012) were used to detect asymmetry in the distribution of the change in driving performance at 2 hours and 6 hours post-dose. Single McNemar tests were used to analyse the difference in proportions of

participants with improved or impaired driving performance at relevant thresholds. Thresholds of 1.0, 1.5, 2.0, 2.5, 3.0, and 3.5 cm were used. In comparisons of solriamfetol and placebo, improvement was defined as a decrease in SDLP in participants treated with solriamfetol compared to placebo at the threshold, and impairment was defined as an increase in SDLP at the threshold or failure to complete the driving test while on solriamfetol because of sleepiness or safety concerns (regardless of their performance on placebo; participants who failed to complete the driving test while on placebo but who completed the test while on solriamfetol were not counted as impaired or improved).

The number of participants who failed to complete the driving test and the duration of the drive before stopping were summarised descriptively. Additional secondary efficacy measures (standard deviation of speed, number of lane drifts, and THAT scores) were analysed using a similar ANOVA method as described for SDLP. No multiplicity adjustments were made in the efficacy analyses for multiple endpoints, and all *p* values are therefore nominal.

Demographic, narcolepsy history, and safety data were summarised for the safety population, which included all participants who received ≥1 dose of study drug. No formal statistical testing was performed on these parameters.

Results

A total of 29 participants were screened; of these 4 failed screening and 25 were enrolled. One participant withdrew consent prior to dosing; therefore 24 participants comprised the safety population. Two participants withdrew from the study and did not have evaluable SDLP data at 2 hours post-dose; therefore, the modified intention-to-treat (mITT) population comprised 22 participants, all of whom completed the study. The safety population was 54% male, with a mean age of 40.4 years; demographic and clinical characteristics (obtained from medical history) are listed in Table 1.

Characteristic	Participants (*N* = 24)
Age, years, mean (SD)	40.4 (11.8)
Male, *n* (%)	13 (54.2)
BMI, kg/m^2, mean (SD)	26.7 (5.2)
Narcolepsy history[a]	
Mean MWT sleep latency, min, mean (SD)	(*n* = 22)
	4.0 (2.5)
Presence of daily irresistible need to sleep, *n* (%)	23 (95.8)
Hypnagogic hallucinations, *n* (%)	14 (58.3)
Sleep paralysis and disruptive nighttime sleep, *n* (%)	20 (83.3)
Number of SOREM periods, mean (SD)	(*n* = 23)
	3.1 (1.1)
HLA DQB1*0602 positive, *n* (%)	21 (87.5)[b]

Table 1. Demographic and baseline clinical characteristics. [a]Data from medical history. [b]Data not available for 3 participants.
Abbreviations: BMI, body mass index; HLA DQB1*0602, human leukocyte antigen DQB1*0602 allele; MWT, Maintenance of Wakefulness Test; SOREM, sleep-onset rapid eye movement.

The observed mean (SD) SDLP at 2 hours post-dose was 20.9 (3.6) cm with placebo and 19.0 (3.6) cm with solriamfetol (mean [SD] difference, –1.91 [2.5] cm) and at 6 hours post-dose was 21.6 (5.8) cm and 19.8 (3.5) cm, respectively (mean [SD] difference, –1.62 [4.4] cm).

	Standard deviation of lateral position					
	Placebo		Solriamfetol		Median difference[a]	
Time point	*n*	Median, cm	*n*	Median, cm	(range), cm	*p*[b]
2 hours post-dose[c]	22	20.46	22	19.08	−1.9 (−6.7 to 2.6)	0.002
6 hours post-dose	21	19.78	22	19.59	−1.1 (−12.1 to 6.0)	0.125

Table 2. Analysis of standard deviation of lateral position. [a]Solriamfetol -placebo. [b]Wilcoxon rank sum test. [c]Primary endpoint.

On the primary endpoint of SDLP at 2 hours post-dose, the median SDLP was significantly lower with solriamfetol compared with placebo (median difference, −1.90 cm [range, −6.7 to 2.6]; *p* = 0.002); the median difference in SDLP at 6 hours post-dose was −1.1 cm (range, −12.1 to 6.0; *p* = 0.125) (Table 2). SDLP differences from placebo for individual participants' data are illustrated in Figure 3A.

A total of 12 participants had ≥1 incomplete driving test. The number of incomplete driving tests was greater with placebo compared with solriamfetol at both 2 hours post-dose and 6 hours post-dose (Table 3). Specifically, 11 participants had ≥1 incomplete test while on placebo (6 on both tests and 5 on a single test [1 at 2 hours; 4 at 6 hours]) and 5 participants had ≥1 incomplete test while on solriamfetol (2 on both tests and 3 on a single test [2 at 2 hours; 1 at 6 hours]). For both placebo and solriamfetol, at 2 hours post-dose, more tests were stopped by the instructor than by the participant; at 6 hours post-dose, more tests were stopped by the participant.

	Placebo	**Solriamfetol**
Number of incomplete tests		
2 hours post-dose	7	4
Stopped by participant	3	1
Stopped by instructor	4	3
6 hours post-dose	10	3
Stopped by participant	7	2
Stopped by instructor	3	1
Number of participants with incomplete tests[a]	11[b]	5[b]
Duration of drive before stopping, min[c]		
Mean (SD)	27.5 (14.56)	25.9 (11.45)
Median (IQR)	26.0 (21, 35)	27.0 (16, 33)
[range]	[6, 54]	[12, 44]

Table 3. Incomplete driving tests. [a]A total of 12 participants had incomplete tests; 7 of these participants had multiple incomplete tests (ie, on both treatments and/or at multiple timepoints); of the 5 who had a single incomplete test, 4 had an incomplete test on placebo (1 at the 2-hour time point and 3 at the 6-hour time point) and 1 had an incomplete test on solriamfetol (at the 2-hour time point).
[b]4 of these participants had at least 1 incomplete test on solriamfetol and at least 1 on placebo. [c]Each driving test was scheduled to be ~60 minutes in duration. Abbreviations: IQR, interquartile range; SD, standard deviation.

Overall higher percentages of participants had improvement (vs. impairment) on solriamfetol at all thresholds (from 1.0 to 3.5 cm, except 3.5 cm at 2 hours); however, single McNemar tests at each threshold did not demonstrate differences at either time point (all $p > 0.05$), and the maximum McNemar test did not show asymmetry at either 2 hours (Figure 3B) or 6 hours post-dose (data not shown). Individual participant data for SDLP by treatment at 2 hours post-dose and 6 hours post-dose are illustrated in Figure 3C and Figure 3D, respectively.

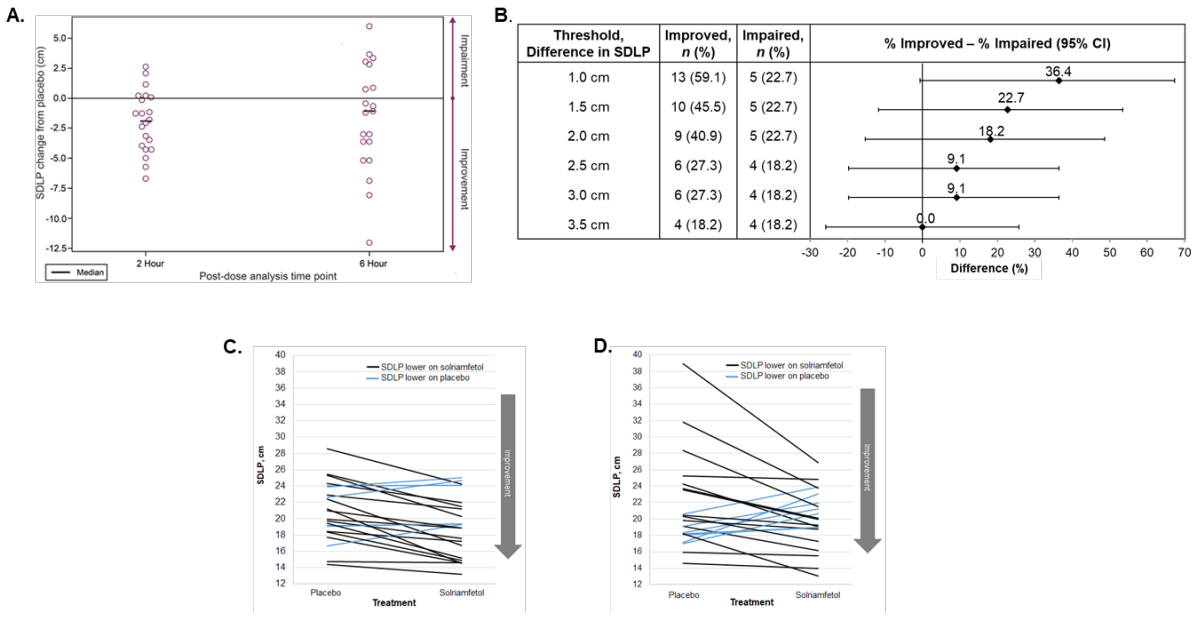

Figure 3. Individual driving performance and symmetry analysis. **(A)** SDLP difference from placebo by participant at 2 hours (n = 22) and 6 hours (n = 21) post-dose. **(B)** Symmetry analysis of SDLP difference scores at 2 hours post-dose: percentage (out of n = 22) with improvement vs. impairment of driving performance with solriamfetol compared with placebo, at thresholds increasing from 1.0 cm to 3.5 cm. **(C)** SDLP at 2 hours post-dose by participant (n = 22). **(D)** SDLP at 6 hours post-dose by participant (n = 21). SDLP, standard deviation of lateral position.

On the additional secondary endpoints of standard deviation of speed and number of lane drifts, no differences were observed between solriamfetol and placebo at 2 or 6 hours post-dose; however, THAT scores were higher (indicating greater alertness) with solriamfetol compared with placebo. The least squares (LS) mean (standard error [SE]) standard deviation of speed at 2 hours post-dose was 2.8 (0.2) km/h with solriamfetol and 3.0 (0.2) km/h with placebo (LS mean difference, −0.22 [95% CI: −0.48, 0.05]) and at 6 hours was 3.1 (0.2) with solriamfetol and 3.2 (0.2) with placebo (LS mean difference, −0.11 [95% CI: −0.38, 0.17]). The LS mean (SE) number of lane drifts at 2 hours

was 2.3 (0.8) with solriamfetol and 3.3 (0.8) with placebo (LS mean difference, −0.98 [95% CI: −3.1, 1.1]) and at 6 hours post-dose was 3.6 (0.8) with solriamfetol and 3.7 (0.84) with placebo (LS mean difference, −0.08 [95% CI: −2.2, 2.0]). The LS mean (SE) THAT score with placebo was 26.8 (1.4) and with solriamfetol was 34.0 (1.4), and the LS mean difference between solriamfetol and placebo was 7.1 (95% CI: 4.1, 10.2).

Treatment-emergent AEs (TEAEs) were reported for 20 (83%) participants; 6 (26%) participants experienced a TEAE while on placebo and 17 (74%) while on solriamfetol. One participant discontinued due to AEs (nausea and vomiting, which occurred while on placebo). All TEAEs were mild or moderate in severity. The most common TEAEs reported while participants were taking solriamfetol were headache and decreased appetite (*n* = 4 each). There were no serious or fatal TEAEs. Changes from baseline in systolic and diastolic blood pressure and pulse rate were generally small, and their occurrence was proportionately similar between the 2 treatment groups and across treatment periods/visits (data not shown).

Discussion

This study demonstrates that solriamfetol treatment at 150 mg/day for 3 days followed by 300 mg/day for 4 days significantly improved driving performance compared with placebo in participants with narcolepsy, as determined by the primary endpoint of SDLP at 2 hours post-dose. SDLP at 6 hours post-dose reflected some improvement with solriamfetol, although to a lesser extent.

While clear thresholds for clinically relevant improvement in SDLP have not been established, the clinical meaningfulness of the primary finding may be considered in the context of normative data. In an analysis of data from 74 healthy participants, the mean (SE) SDLP was 18.19 (0.46) cm with an upper limit of the 2-sided 95% CI of 19.09 cm (Vinckenbosch et al., 2021). The mean and median SDLP with placebo at 2 hours post-dose in the present study (20.88 and 20.46,

respectively) exceeded this threshold, suggesting impairment in this population while on placebo, whereas the mean and median SDLP with solriamfetol (18.97 and 19.08, respectively) was within the CI of the aforementioned population of healthy participants, suggesting weaving and road-tracking ability within a healthy population norm while treated with solriamfetol.

Although this study was not designed to directly assess the risk of traffic accidents, studies of the effects of blood alcohol concentration and use of benzodiazepines on driving performance suggest that change in SDLP and crash risk are highly correlated and that SDLP is a valid predictor of alcohol- or drug-induced crash risk (Owens & Ramaekers, 2009). Data are lacking to confirm the predictive validity of SDLP in the context of wake-promoting agents and the potential for reducing the risk for traffic accidents. However, epidemiologic studies suggest stimulant and modafinil use reduces crash risk in patients with narcolepsy (Pizza et al., 2015; Tzeng et al., 2019). Nonetheless, the on-road driving test is generally regarded as the gold standard for assessing drug-induced changes in driving performance (Jongen et al., 2017). Further, the ecological validity of the on-road driving test is one of its strengths. On-road driving might differ in important ways from simulators in terms of motivational factors that might have an effect on the validity of simulator testing. These factors may include the feeling of playing a video game versus actually driving next to real cars and people, feelings of sickness that can accompany simulator testing, and additional factors such as the lack of kinetic input (turning, speed adjustments) and graphics (often lacking the dimension of depth).

Few studies of narcolepsy treatments have evaluated functional outcomes such as driving. In particular, studies of the effects of wake-promoting agents on measures of on-the-road driving performance, and specifically SDLP, in patients with narcolepsy are limited (Philip et al., 2014; Sagaspe et al., 2019). In a study of modafinil in patients with narcolepsy or idiopathic hypersomnia, the reduction in mean SDLP in an on-road driving test (conducted ~1.5 hours post-dose) with modafinil (400 mg/day) compared with placebo was not statistically significant (23.6±0.6 vs. 24.9±0.9 cm; p =

0.06) (Philip et al., 2014). This is in contrast to the findings of this study, which showed a statistically significant improvement in SDLP at 2 hours post-dose with solriamfetol.

Several participants were unable to complete one or more driving tests. The greater number of incomplete driving tests with placebo, particularly at the 6-hour post-dose time point, suggests that participants had less driving difficulty while on solriamfetol treatment. This finding supports the primary endpoint as it also reflects improvement in driving performance with solriamfetol. Considering these findings in the context of data from healthy participants, the overall percentage of driving tests stopped in this study was ~28% (24/87), which is nearly 9 times higher than in previous studies with healthy volunteers (3.1%) (Verster & Roth, 2012). In this study, 40% (17/43) of tests on placebo and 16% (7/44) on solriamfetol were stopped, whereas less than 1% and ~4% of the driving tests in unmedicated healthy volunteers and patients on various potentially sedating drug treatments, respectively, were stopped in previous studies (Verster & Roth, 2012). No participants stopped driving tests in the aforementioned modafinil study (Philip et al., 2014), despite the fact that those tests covered more than twice the distance, though participants in that study were allowed to remain on anticataplectic medication in contrast to the current study. This shows that, with and without medication, a significant percentage of participants in the current study had problems maintaining alertness for up to an hour during prolonged highway driving. Interestingly, more tests were stopped by the participant than by the instructor (13 vs. 11; Table 3). In contrast, in studies with healthy volunteers the decision to stop was 3 to 4 times more often made by the instructor than by the participant (Verster & Roth, 2012). This suggests that participants with narcolepsy in this study seemed aware of potential impairment and were careful to avoid further risks. If participants decided to stop before effects on SDLP were detectable, the observed treatment effect on SDLP may be an underestimation of the ability of solriamfetol to improve performance in this setting. For example, if participants had not stopped their tests while on placebo, SDLP likely would have reflected greater impairment.

The SD of SDLP has been reported to range from 2.6 to 4.2 cm in healthy participants or in participants with ADHD with or without stimulant or hypnotic treatment (Vermeeren et al., 2014; Verster et al., 2008; Verster & Roth, 2011). The power calculation performed to determine the sample size required for the present study therefore assumed an SD of 3.0 cm, in line with the estimated SD for power estimation in a study of methylphenidate use in participants with attention deficit hyperactivity disorder (Verster et al., 2008). However, the observed SD of SDLP in this study ranged from 3.5 to 5.8 cm, suggesting the study may have been underpowered to detect a difference in SDLP. Although an improvement was still detected at 2 hours post-dose in participants treated with solriamfetol, it was not maintained at 6 hours post-dose.

Other secondary driving outcomes (standard deviation of speed and lane drifts) showed minimal differences between solriamfetol and placebo, which may be due to a relative lack of sensitivity or statistical power. Standard deviation of speed is less sensitive to changes in driving performance parameters compared with SDLP (Irwin, Iudakhina, Desbrow, & McCartney, 2017; Verster & Roth, 2014). In contrast, the difference between treatments in THAT scores was more substantial and suggested greater alertness with solriamfetol. This improvement is consistent with the established wake-promoting effects of solriamfetol on other measures, such as the Epworth Sleepiness Scale and the Maintenance of Wakefulness Test, which showed treatment differences from placebo (least squares mean) of –2.2 to –4.7 points and 2.6 to 10.1 minutes, respectively, after 12 weeks of treatment with solriamfetol at doses of 75 to 300 mg/day in the phase 3 trial of solriamfetol in participants with narcolepsy (Malhotra et al., 2020; Thorpy et al., 2019). These wake-promoting effects have been shown to be maintained for up to 12 months in an open-label extension study (Malhotra et al., 2020).

The tolerability profile of solriamfetol in this study is consistent with those observed in other clinical trials in participants with narcolepsy (Ruoff et al., 2016; Thorpy et al., 2019). All TEAEs were

mild or moderate in severity. No participant discontinued the study due to AEs while taking solriamfetol.

One limitation of this study is the use of solriamfetol at a dose of 300 mg/day, which exceeds the maximum recommended dose of 150 mg/day. The 300-mg/day dose was selected on the basis of prior phase 2 study data (Bogan et al., 2015; Ruoff et al., 2016) and was the highest dose used in the pivotal trials of solriamfetol in participants with narcolepsy, which demonstrated efficacy at 75 mg, 150 mg, and 300 mg (Malhotra et al., 2020; Thorpy et al., 2019). It could also be argued that the current study population was not wholly representative of clinical populations because of the prohibition against using other narcolepsy treatments during the study. Although this limitation may hamper generalisability in a population that often requires polypharmacy (Thorpy & Hiller, 2017), other treatments that affect sleepiness were prohibited to isolate the effects of solriamfetol on driving performance. Solriamfetol may also be used as monotherapy for some patients with narcolepsy (Abad, 2021). Finally, there was no active comparator in this study, limiting the ability to draw comparisons with other wake-promoting agents.

Conclusion

In participants with narcolepsy, solriamfetol at 300 mg/day significantly improved SDLP, an important measure of driving performance, at 2 hours post-dose. These findings indicate that the robust wake-promoting efficacy of solriamfetol demonstrated in clinical trials resulted in improved real-world functional performance in participants with narcolepsy.

References

Abad, V. C. (2021). Profile of solriamfetol in the management of excessive daytime sleepiness associated with narcolepsy or obstructive sleep apnea: Focus on patient selection and perspectives. *Nature and Science of Sleep, 13,* 75-91.

American Academy of Sleep Medicine. (2014). *International classification of sleep disorders* (3rd ed.). Darien, IL: American Academy of Sleep Medicine.

American Psychiatric Association. (2013). Narcolepsy. In *Diagnostic and statistical manual of mental disorders* (5th ed., pp. 372-378). Washington, DC: American Psychiatric Association.

Bogan, R. K., Feldman, N., Emsellem, H. A., Rosenberg, R., Lu, Y., Bream, G., . . . Lankford, D. A. (2015). Effect of oral JZP-110 (ADX-N05) treatment on wakefulness and sleepiness in adults with narcolepsy. *Sleep Medicine, 16*(9), 1102-1108. doi:10.1016/j.sleep.2015.05.013

Emsellem, H. A., Thorpy, M. J., Lammers, G. J., Shapiro, C. M., Mayer, G., Plazzi, G., . . . Dauvilliers, Y. (2020). Measures of functional outcomes, work productivity, and quality of life from a randomized, phase 3 study of solriamfetol in participants with narcolepsy. *Sleep Medicine, 67,* 128-136. doi:10.1016/j.sleep.2019.11.1250

Findley, L., Unverzagt, M., Guchu, R., Fabrizio, M., Buckner, J., & Suratt, P. (1995). Vigilance and automobile accidents in patients with sleep apnea or narcolepsy. *Chest, 108*(3), 619-624. doi:10.1378/chest.108.3.619

Flores, N. M., Villa, K. F., Black, J., Chervin, R. D., & Witt, E. A. (2016). The humanistic and economic burden of narcolepsy. *Journal of Clinical Sleep Medicine, 12*(3), 401-407. doi:10.5664/jcsm.5594

Irwin, C., Iudakhina, E., Desbrow, B., & McCartney, D. (2017). Effects of acute alcohol consumption on measures of simulated driving: A systematic review and meta-analysis. *Accident Analysis and Prevention, 102,* 248-266. doi:10.1016/j.aap.2017.03.001

Jongen, S., Vermeeren, A., van der Sluiszen, N. N., Schumacher, M. B., Theunissen, E. L., Kuypers, K. P., . . . Ramaekers, J. G. (2017). A pooled analysis of on-the-road highway driving studies in actual traffic measuring standard deviation of lateral position (i.e., "weaving") while driving at a blood alcohol concentration of 0.5 g/L. *Psychopharmacology, 234*(5), 837-844. doi:10.1007/s00213-016-4519-z

Kornum, B. R., Knudsen, S., Ollila, H. M., Pizza, F., Jennum, P. J., Dauvilliers, Y., & Overeem, S. (2017). Narcolepsy. *Nature Reviews Disease Primers, 3,* 16100. doi:10.1038/nrdp.2016.100

Kotterba, S., Mueller, N., Leidag, M., Widdig, W., Rasche, K., Malin, J. P., . . . Orth, M. (2004). Comparison of driving simulator performance and neuropsychological testing in narcolepsy. *Clinical Neurology and Neurosurgery, 106*(4), 275-279. doi:10.1016/j.clineuro.2003.12.003

Laska, E., Meisner, M., & Wanderling, J. (2012). A maximally selected test of symmetry about zero. *Statistics in Medicine, 31*(26), 3178-3191. doi:10.1002/sim.5384

Liu, S. Y., Perez, M. A., & Lau, N. (2018). The impact of sleep disorders on driving safety--findings from the Second Strategic Highway Research Program naturalistic driving study. *Sleep, 41*(4), 1-11. doi:10.1093/sleep/zsy023

Malhotra, A., Shapiro, C., Pepin, J. L., Hedner, J., Ahmed, M., Foldvary-Schaefer, N., . . . Schwab, R. (2020). Long-term study of the safety and maintenance of efficacy of solriamfetol (JZP-110) in the treatment of excessive sleepiness in participants with narcolepsy or obstructive sleep apnea. *Sleep, 43*(2), zsz220. doi:10.1093/sleep/zsz220

Owens, K., & Ramaekers, J. G. (2009). Drugs, driving, and models to measure driving impairment. In J. C. Verster, S. R. Pandi-Perumal, J. G. Ramaekers, & J. J. de Gier (Eds.), *Drugs, driving and traffic safety* (pp. 43-58). Basel, Switzerland: Birkhauser Verlag.

Philip, P., Chaufton, C., Taillard, J., Capelli, A., Coste, O., Leger, D., . . . Sagaspe, P. (2014). Modafinil improves real driving performance in patients with hypersomnia: a randomized double-blind placebo-controlled crossover clinical trial. *Sleep, 37*(3), 483-487. doi:10.5665/sleep.3480

Philip, P., Guichard, K., Strauss, M., Léger, D., Pepin, E., Arnulf, I., . . . Dauvilliers, Y. (2021). Maintenance of Wakefulness Test: how does it predict accident risk in patients with sleep disorders? *Sleep Medicine, 77,* 249-255. doi:10.1016/j.sleep.2020.04.007

Philip, P., Sagaspe, P., Lagarde, E., Leger, D., Ohayon, M. M., Bioulac, B., . . . Taillard, J. (2010). Sleep disorders and accidental risk in a large group of regular registered highway drivers. *Sleep Medicine, 11*(10), 973-979. doi:10.1016/j.sleep.2010.07.010

Pizza, F., Jaussent, I., Lopez, R., Pesenti, C., Plazzi, G., Drouot, X., . . . Dauvilliers, Y. (2015). Car crashes and central disorders of hypersomnolence: a French study. *PloS One, 10*(6), e0129386. doi:10.1371/journal.pone.0129386

Ramaekers, J. G. (2017). Drugs and driving research in medicinal drug development. *Trends in Pharmacological Sciences, 38*(4), 319-321. doi:10.1016/j.tips.2017.01.006

Ramaekers, J. G., Kuypers, K. P., & Samyn, N. (2006). Stimulant effects of 3,4-methylenedioxymethamphetamine (MDMA) 75 mg and methylphenidate 20 mg on actual driving during intoxication and withdrawal. *Addiction, 101*(11), 1614-1621. doi:10.1111/j.1360-0443.2006.01566.x

Ruoff, C., Swick, T. J., Doekel, R., Emsellem, H. A., Feldman, N. T., Rosenberg, R., . . . Black, J. (2016). Effect of oral JZP-110 (ADX-N05) on wakefulness and sleepiness in adults with narcolepsy: a phase 2b study. *Sleep, 39*(7), 1379-1387. doi:10.5665/sleep.5968

Sagaspe, P., Micoulaud-Franchi, J. A., Coste, O., Leger, D., Espie, S., Davenne, D., . . . Philip, P. (2019). Maintenance of Wakefulness Test, real and simulated driving in patients with narcolepsy/hypersomnia. *Sleep Medicine, 55,* 1-5. doi:10.1016/j.sleep.2018.02.009

Shapiro, C. M., Auch, C., Reimer, M., Kayumov, L., Heslegrave, R., Huterer, N., . . . Devins, G. M. (2006). A new approach to the construct of alertness. *Journal of Psychosomatic Research, 60*(6), 595-603. doi:10.1016/j.jpsychores.2006.04.012

Sunosi™ (solriamfetol) tablets Prescribing Information. (2019). Palo Alto, CA: Jazz Pharmaceuticals, Inc.

Sunosi™ (solriamfetol) tablets Summary of Product Characteristics. (2020). Dublin, Ireland: Jazz Pharmaceuticals Ireland Ltd.

Szabo, S. T., Thorpy, M. J., Mayer, G., Peever, J. H., & Kilduff, T. S. (2019). Neurobiological and immunogenetic aspects of narcolepsy: implications for pharmacotherapy. *Sleep Medicine Reviews, 43,* 23-36. doi:10.1016/j.smrv.2018.09.006

Thorpy, M. J., & Hiller, G. (2017). The medical and economic burden of narcolepsy: Implications for managed care. *American Health & Drug Benefits, 10*(5), 233-241. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5620503/pdf/ahdb-10-233.pdf

Thorpy, M. J., Shapiro, C., Mayer, G., Corser, B. C., Emsellem, H., Plazzi, G., . . . Dauvilliers, Y. (2019). A randomized study of solriamfetol for excessive sleepiness in narcolepsy. *Annals of Neurology, 85*(3), 359-370. doi:10.1002/ana.25423

Tzeng, N. S., Hsing, S. C., Chung, C. H., Chang, H. A., Kao, Y. C., Mao, W. C., . . . Chien, W. C. (2019). The risk of hospitalization for motor vehicle accident injury in narcolepsy and the benefits of stimulant use: A nationwide cohort study in Taiwan. *Journal of Clinical Sleep Medicine, 15*(6), 881-889. doi:10.5664/jcsm.7842

Vermeeren, A., Vuurman, E. F., Leufkens, T. R., Van Leeuwen, C. J., Van Oers, A. C., Laska, E., . . . Roth, T. (2014). Residual effects of low-dose sublingual zolpidem on highway driving performance the morning after middle-of-the-night use. *Sleep, 37*(3), 489-496. doi:10.5665/sleep.3482

Verster, J. C., Bekker, E. M., de Roos, M., Minova, A., Eijken, E. J., Kooij, J. J., . . . Volkerts, E. R. (2008). Methylphenidate significantly improves driving performance of adults with attention-deficit

hyperactivity disorder: A randomized crossover trial. *Journal of Psychopharmacology, 22*(3), 230-237. doi:10.1177/0269881107082946

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. doi:10.2147/ijgm.s19639

Verster, J. C., & Roth, T. (2012). The prevalence and nature of stopped on-the-road driving tests and the relationship with objective performance impairment. *Accident Analysis and Prevention, 45,* 498-506. doi:10.1016/j.aap.2011.09.003

Verster, J. C., & Roth, T. (2014). 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. *Human Psychopharmacology, 29*(1), 19-24. doi:10.1002/hup.2377

Vinckenbosch, F. R. J., Vermeeren, A., Vuurman, E. F. P. M., van der Sluiszen, N. N. J. J. M., Verster, J. C., van de Loo, A. J. A. E., . . . Ramaekers, J. G. (2021). An explorative approach to understanding individual differences in driving performance and neurocognition in long-term benzodiazepine users. *Human Psychopharmacology, 36*(4), e2778. doi:10.1002/hup.2778

Weaver, T. E., Pepin, J. L., Schwab, R., Shapiro, C., Hedner, J., Ahmed, M., . . . Malhotra, A. (2019). Long-term effects of solriamfetol on quality of life in participants with excessive daytime sleepiness associated with narcolepsy or obstructive sleep apnea [abstract 0601]. *Sleep, 42*(suppl 1), A239.

Chapter 5: An explorative approach to understanding individual differences in driving performance and neurocognition in long-term benzodiazepine users.

Abstract

Previous research reported cognitive and psychomotor impairments in long-term users of benzodiazepine receptor agonists (BZRAs). This article explores the role of acute intoxication and clinical complaints. Neurocognitive and on-road driving performance of 19 long-term (≥6 months) regular (≥ twice weekly) BZRA users with estimated plasma concentrations, based on self-reported use, exceeding the therapeutic threshold (C_{BZRA}+), and 31 long-term regular BZRA users below (C_{BZRA}-), was compared to that of 76 controls. BZRA users performed worse on tasks of response speed, processing speed, and sustained attention. Age, but not C_{BZRA} or self-reported clinical complaints, was a significant covariate. Road-tracking performance was explained by C_{BZRA} only. The C_{BZRA}+ group exhibited increased mean standard deviation of lateral position comparable to that at blood-alcohol concentrations of 0.5g/L. Functional impairments in long-term BZRA users are not attributable to self-reported clinical complaints or estimated BZRA concentrations, except for road-tracking, which was impaired in CBZRA+ users. Limitations to address are the lack of assessment of objective clinical complaints, acute task related stress, and actual BZRA plasma concentrations. In conclusion, the results confirm previous findings that demonstrate inferior performance across several psychomotor and neurocognitive domains in long-term BZRA users.

Vinckenbosch, F. R. J., Vermeeren, A., Vuurman, E. F. P. M., van der Sluiszen, N. N. J. J. M., Verster, J. C., van de Loo, A. J. A. E., van Dijken, J. H., Veldstra, J. L., Brookhuis, K. A., de Waard, D., & Ramaekers, J. G. (2021). An explorative approach to understanding individual differences in driving performance and neurocognition in long-term benzodiazepine users. *Human Psychopharmacology-Clinical and Experimental*, *36*(4), 1-17. [e2778]. https://doi.org/10.1002/hup.2778

Introduction

Benzodiazepine receptor agonists (BZRAs) are a class of drugs prescribed mainly for the symptomatic treatment of insomnia and anxiety. They act as positive allosteric modulators of the gamma-aminobutyric acid type a (GABAa) receptors in the central nervous system (CNS) where they potentiate the actions of the inhibitory neurotransmitter GABA, thus acting as CNS depressants. Although CNS suppression is the intended therapeutic effect, psychomotor and cognitive side-effects such as unsteady gait, slowed response speed, impaired sustained attention and anterograde amnesia also occur (Jongen, Vuurman, Ramaekers, & Vermeeren, 2018; Uzun, Kozumplik, Jakovljević, & Sedić, 2010). These side-effects can negatively impact daily functioning. This is especially apparent in the elderly where the risks of falling and cognitive decline have been linked to the use of BZRAs (Paterniti, Dufouil, & Alpérovitch, 2002; Sorock & Shimkin, 1988; Uzun et al., 2010). Also, in the context of traffic safety, the use of BZRAs has been linked to increased crash risk and has been observed to impair road tracking during standardized on-the-road driving testing (Jongen et al., 2018; Leufkens & Vermeeren, 2014; Roth, Eklov, Drake, & Verster, 2014; Vermeeren, 2004; Verster, Veldhuijzen, & Volkerts, 2005).

Prolonged and regular (i.e. daily or near daily) BZRA use has the potential to induce physical dependence (Owen & Tyrer, 1983). It is therefore advised that treatment duration is limited to 2-4 weeks, including gradually tapering off the dose (Ashton, 1994). However, in clinical practice the prolonged use of BZRAs is frequently observed. Approximately 12% to 30% of all first time users progress to long-term use (Bushnell, Stürmer, Gaynes, Pate, & Miller, 2017; Gerlach, Maust, Leong, Mavandadi, & Oslin, 2018). Arguably, this phenomenon is in part due to the practice of "*doctor shopping*" by patients who do not wish to discontinue treatment (Cook, Biyanova, Masci, & Coyne, 2007; Peirce, Smith, Abate, & Halverson, 2012). Also, there seems to be some empirical basis for prolonged BZRA treatment of clinical anxiety as it has been reported that, unlike the sedating properties, tolerance does not develop to the anxiolytic effects (Vinkers & Olivier, 2012). Given this

reality of frequently occurring long-term BZRA use, it is important to determine its potential consequences.

Several studies have investigated the effects of long-term BZRA use on performance. A recent study by van der Sluiszen et al. (2019) on neurocognitive and driving performance of long-term (≥ 6 months) BZRA users found that this group exhibited significant impairments on tasks of vigilance, executive functioning, and reaction speed. In line with these findings, a meta-analysis by Barker, Greenwood, Jackson, and Crowe (2004a) reported that long-term BZRA users (≥ 1 year) are impaired on neuropsychological tests of attention, problem solving, visuospatial cognition, general intelligence, psychomotor speed, and non-verbal memory. A follow-up investigation found that impaired performance was still apparent in these domains as well as in verbal memory, motor control, and speed of processing, six months after last use (Barker, Greenwood, Jackson, & Crowe, 2004b). The persistence of the observed impairments led the authors to the suggestion that this might be indicative of a permanently acquired cognitive deficit caused by prolonged BZRA exposure.

In contrast, a prior study by Lucki, Rickels, and Geller (1986) did not find any significant differences in psychomotor or cognitive functioning between patients treated with BZRAs and BZRA-free patients, with the exception of slower visual temporal processing in the former. These findings suggest that prolonged BZRA use has little consequences for cognitive or psychomotor performance. Instead, it suggests that the reported performance decrements in BZRA users relative to healthy controls might be attributable to a systematic difference between BZRA users and controls, other than BZRA use. The clinical condition for which the BZRA is prescribed is the most obvious candidate. The study by van der Sluiszen et al. (2019) showed that BZRA treated individuals still report increased levels of anxiety, depression, and insomnia. It is known that anxiety (Yu et al., 2018), depression (McDermott & Ebmeier, 2009), and insomnia (Fortier-Brochu, Beaulieu-Bonneau, Ivers, & Morin, 2012) can have a negative impact on cognitive and psychomotor test performance. However, Barker et al. (2004a, 2004b) did not consider clinical complaints as a potential confounder. The study by van

der Sluiszen et al. (2019) did assess clinical complaints but did not consider them as covariates in the analysis.

Besides controlling for clinical complaints, it is also important to control for the acute effects of BZRAs on performance. It has been reported that tolerance to the performance impairing effects of benzodiazepines is only partial (Pomara, Tun, DaSilva, & Hernando, 1998). Therefore, it seems likely that the observed impairments in long-term BZRA users are partly attributable to these "residual acute effects". The studies by Lucki et al. (1986) and Barker et al. (2004a) aimed to control for this by excluding participants that used a BZRA within 4 hours of laboratory testing. However, depending on the half-life and dose of the respective BZRA, this approach is arguably inadequate to completely control for residual acute effects. van der Sluiszen et al. (2019) pointed out the problem of heterogeneity of medication use that is inherent to a clinical sample of BZRA users. The patients included in the study used different BZRAs (i.e. different potencies), at different doses, at different frequencies, and at different times relative to the laboratory tests and on-road driving. All of these factors determine the achieved drug plasma concentrations and should therefore be taken into consideration when controlling for potential residual acute BZRA effects.

A previous investigation by Verster and Roth (2013a) concluded that no significant relationship exists between individual BZRA blood plasma concentrations and road tracking performance during on-road driving. It is conceivable that the psychomotor and cognitive effects of BZRAs are subject to a wide range of inter-individual variability. However, it should be noted that all of the reported plasma concentrations fell well within the therapeutic window for the respective BZRA. It can therefore not be excluded that a relationship exists between BZRAs blood plasma concentrations and road-tracking performance all together. The notion that no significant behavioral or cognitive effects are expected below the therapeutic threshold remains unchanged, while effects are expected when blood plasma concentration exceed it, albeit to varying degrees. It is therefore argued that it is important to assess whether or not participants have BZRA plasma concentrations

exceeding the therapeutic threshold at the time of cognitive testing in order to control for potential residual acute BZRA effects in long-term users.

Uncovering what contributes to the observed functional impairments in long-term BZRA users is important since different causal factors imply different clinical management strategies. If the impairments reported in long-term BZRA users are attributable to residual acute effects or acquired functional deficits due to prolonged exposure, cessation or limitation of BZRA use would be advised. However, if clinical complaints account for the observed impairments, continuation could be the best decision (Shinfuku et al., 2019). Also, for the individual assessment of the fitness to drive of long-term BZRA users, it is important to elucidate which factors might impair driving performance to allow for efficient screening. The present study revisited the dataset from van der Sluiszen et al. (2019) that included on-road driving and neurocognitive performance of long-term benzodiazepine users and healthy controls. The current investigation compares on-road driving performance and neurocognitive functioning of long-term regular BZRA users to that of healthy controls in order to elucidate whether clinical complaints and residual acute effects, operationalized as estimated BZRA plasma concentrations, account for the previously reported performance decrements in BZRA users.

Methods

Participants

Medical, driving and neurocognitive data of 55 long-term (> 6months), chronic (≥2 times per week) BZRA users, aged 21 to 75, were retrieved from a previous study into the effects of long-term use of sedative medications on driving ability initiated by the Dutch government (van der Sluiszen et al., 2019; Verster et al., 2016b). In this study, all information regarding eligibility was gathered through completion of an extensive medical questionnaire which was subsequently reviewed by a clinician. Participants were required to be in the possession of a valid driver's license, of which a photocopy was obtained, and to drive at least 500km/year. In addition, normal or corrected to normal vision, and a body mass index between 17 to 35kg/m^2 were necessary prerequisites. In

addition to the medical questionnaire, the requirement regarding visual acuity was checked on site using an eye chart examination. Participants were excluded if they consumed >21 alcoholic beverages per week, smoked >20 cigarettes per day, or used any psychoactive substances recreationally and regularly, or recently prior to testing as determined by an alcohol breathalyzer test and a urine test. Only those BZRA users of whom a complete medication profile could be retrieved, i.e. type of medication, dose, frequency, and time of last use, were included in the current analysis. In addition to the BZRA user group, data of 76 control participants was retrieved. Control participants were free of psychoactive medications and diagnosed psychiatric, neurological, and substance abuse disorders, as determined by the inspection of the medical questionnaire by a clinician.

Estimation and classification of BZRA plasma concentrations

In order to quantify expected residual acute drug effects, the average steady state drug plasma concentration was estimated for each BZRA user from use as reported by the participant (drug, dose, time of dosing) and established pharmacokinetic parameters of the drug (see equation 1, (Wakamatsu, Aoki, Sakiyama, Ohnishi, & Sugita, 2013)). These and other parameters were also used to estimate drug plasma concentration at the start of the testing day (see equation 2, (Wakamatsu et al., 2013)). The pharmacokinetic parameters entered into equation 1 and 2 and their literature references are listed in table 1.

$$\bar{C}ss = \frac{F*D}{CL*\tau}$$

Equation 1. $\bar{C}$ss: Average drug plasma concentration at steady state (ng/ml); **F:** Bioavailability (% absorbed); **D:** dose (mg); **CL:** Clearance rate (ml/min/kg); **τ:**the dosing interval (hours).

$$Css(t) = \frac{F*D*Ka}{Vd*(Ka-Kel)} * \left(\frac{e^{-Kel*t}}{1-e^{-Kel*\tau}} - \frac{e^{-Ka*t}}{1-e^{-Ka*\tau}}\right)$$

Equation 2. Css(t): estimated drug plasma concentration at steady state at time **t** (ng/ml); **t:** time since last use relative to start of testday (hours); **F:** bioavailability (% of drug absorbed); **D:** drug dose (mg); **Ka:** absorption rate constant (h^{-1}); **Vd:** Apparent volume of distribution(L/kg); **Kel:** elimination rate constant (h^{-1}).

BZRA	F	$t_{1/2}$ abs	Ka	Vd	CL	Kel	references
Alprazolam	90	19.2	2.17	0.84	67.2	.080	Greenblatt and Wright (1993); Smith, Kroboth, Vanderlugt, Phillips, and Juhl (1984); Wright (1995)
Brotizolam	70	10.2	4.08	0.66	111	.168	Jochemsen, Wesselman, Hermans, Van Boxtel, and Breimer (1983); Langley and Clissold (1988); Scavone, Greenblatt, Harmatz, and Shader (1986)
Clonazepam	90	24.6	1.69	2.95	42	.014	Berlin and Dahlström (1975); Crevoisier, Delisle, Joseph, and Foletti (2003); Greenblatt et al. (2005); Wishart et al. (2018)
Clorazepate	91	18.6	2.24	1.28	13.2	.010	Ochs, Steinhaus, Locniskar, Knüchel, and Greenblatt (1982); Shader et al. (1981); Wishart et al. (2018)
Diazepam	94	31.8	1.31	1.83	21	.011	Divoll, Greenblatt, Ochs, and Shader (1983); Eatman et al. (1977); Greenblatt, Allen, Harmatz, and Shader (1980); Greenblatt, Harmatz, Friedman, Locniskar, and Shader (1989)
Lorazepam	90	32.4	1.28	1.15	57	.050	Greenblatt (1981); Greenblatt, Divoll, Harmatz, and Shader (1982); Wishart et al. (2018)
Lormetazepam	94	102	0.41	6.8	240	.035	Hildebrand, Hellstern, Hümpel, Hellenbrecht, and Saller (1990); Kampf, Huempel, Lerche, and Kessel (1981); Lombardo, Obach, Shalaeva, and Gao (2002)
Midazolam	50	18	2.31	1.3	330	.253	Greenblatt et al. (1984); Malacrida, Fritz, Suter, and Crevoisier (1992); Wishart et al. (2018)
Nitrazepam	100	16.2	2.57	2.55	54	.021	Greenblatt et al. (1985); Jochemsen et al. (1982)
Oxazepam	93	37.8	1.1	1.5	87	.058	Greenblatt (1981); Sonne et al. (1988)
Temazepam	95	117	0.36	1.4	71.4	.051	Divoll, Greenblatt, Harmatz, and Shader (1981); Schwarz (1979); Wishart et al. (2018)
Zolpidem	70	37.8	2.24	0.54	348	.644	Greenblatt et al. (2013); Langtry and Benfield (1990); Olubodun et al. (2002); Salvà and Costa (1995)
Zopiclone	80	-	3.49	-	228	.172	Caille, Du Souich, Spenard, Lacasse, and Vezina (1984); Fernandez, Martin, Gimenez, and Farinotti (1995); Gaillot, Heusse, Hougton, Aurele, and Dreyfus (1983); Noble, Langtry, and Lamb (1998)

Table 1. pharmacokinetic parameters of encountered BZRA's as applied in equation 1 and 2; **F:** bioavailability (% of drug absorbed); **$t_{1/2}$ abs:** absorption half-life(min); **Ka:** absorption rate constant (h^{-1}), calculated as Ka= Ln(2)/$t_{1/2}$ abs; **Vd:** Apparent volume of distribution(L/kg); **CL:** Clearance rate (ml/h/kg); **Kel:** elimination rate constant (h^{-1}), calculated as Kel=CL/Vd.

For three BZRA users, it was not possible to estimate the drug plasma concentrations because of irregular drug use. For another two users the concentration estimates could not be determined because reliable pharmacokinetic parameters of bromazepam could not be retrieved from the literature. Hence, a total of 50 BZRA users were included for the analysis. Six BZRA users indicated using two BZRAs daily. The doses of both BZRAs were first converted to the diazepam equivalents (Ashton, 1994) before being entered into the equations. The estimated equivalent plasma concentrations were then added for the final estimate.

Estimated plasma concentrations were compared to the therapeutic threshold of the respective drug (Schulz, Iwersen-Bergmann, Andresen, & Schmoldt, 2012) in order to determine whether any CNS depressant effects would be likely. The correspondence rate of the two drug plasma concentration estimates, i.e. $\bar{C}ss$ and Css(t), in this respect was 100%. In other words, BZRA users with estimated $\bar{C}ss$ levels that exceeded the therapeutic threshold were also found to have estimated Css(t) levels exceeding this threshold (C_{BZRA}+, N=19), and vice versa (C_{BZRA}-, N=31). A

summary of group descriptives, self-reported clinical complaint severity, and medication use is provided in table 2. A complete overview of type, dose and frequency of BZRA use, as well as the estimated drug plasma concentrations per BZRA user is provided in appendix 1.

	Control Participants	Patients (C_{BZRA-})	Patients (C_{BZRA+})
N	76	31	19
Gender *[F:M]*	35:41	19:12	9:10
Mean age (SD)	55.6 (12.7)	52.8 (12.0)	56.5 (11.4)
Mean annual distance (SD) *[km]*	13499 km (9276)	12798 km (8886)	15553 km (24189)
BDI (SD) [a]	2.51 (2.69)	12.39 (10.49)	8.89 (6.54)
STAI-T (SD) [b]	27.33 (5.65)	43.23 (11.59)	40.89 (12.55)
PSQI (SD) [c]	2.86 (2.33)	8.9 (4.87)	8.47 (5.35)
GSQS (SD) [d]	1.54 (2.05)	5.19 (4.39)	4.84 (4.44)
Median alcoholic beverages per week (IQR)	4.5 (8.5)	2 (7)	4 (6)
Mean Equivalent BZRA plasma concentration (SD) *[ng/ml]* [e]	-	46.9 (57.3)	257.2 (574.1)
Mean Equivalent BZRA steady state plasma concentration (SD) *[ng/ml]* [f]	-	46.7 (49.7)	245 (529.5)
Duration of use *[years]*	-	8.6 (8.4)	6.9 (7.0)
CNS co-medications [g]:	-	23 (75%)	14 (74%)
Tri- and tetracyclic antidepressants [h]	-	3 (10%)	3 (16%)
Selective serotonin/Serotonin-norepinephrine reuptake inhibitors [i]	-	13 (42%)	9 (47%)
Antipsychotics [j]	-	5 (16%)	1 (5%)
Opioids [k]	-	4 (13%)	1 (5%)
Lamotrigine	-	1 (3%)	-
Lithium	-	2 (6%)	1 (5%)
Paracetamol	-	3 (10%)	4 (21%)
Pramipexole	-	1 (3%)	-
Pregabaline	-	1 (3%)	-
Tranylcypromine	-	1 (3%)	-
Trazodone	-	1 (3%)	-

Table 2. Sample descriptives. **[a]** Beck's depression inventory; **[b]** State-trait anxiety inventory–trait; **[c]** Pittsburgh sleep quality index; **[d]** Groningen sleep quality scale; **[e]** Mean estimated equivalent diazepam plasma concentration at start of test day; **[f]** Mean estimated equivalent diazepam average steady state plasma concentration; **[g]** Number of participants that used at least one other central nervous system medication (classified as N-class by the anatomical, therapeutic, and chemical (ATC) classification system) daily or multiple times per week; **[h]** amitriptyline(n=2), clomipramine(n=1), mirtazapine(n=2) and nortriptyline(n=1); **[i]** Citalopram(n=8), duloxetine(n=2), escitalopram(n=3), fluoxetine(n=3), paroxetine(n=4), sertraline(n=1), and venlafaxine(n=1); **[j]** Olanzapine(n=1), quetiapine (n=7), risperidone(n=1); **[k]** Codeine(n=2), oxycodone(n=2), and tramadol(n=1).

Materials

Self-reported clinical questionnaires

The Beck Depression inventory (BDI): consists of 21 items containing a statement that is related to depressive symptomatology. The participant indicates how relatable each item is on a scale from 0 to 3. Higher scores indicate more severe depressive symptoms (Beck, Steer, & Carbin, 1988).

The Groningen Sleep Quality Scale (GSQS): assesses subjective sleep quality in 14 items that are rated as true or false. Higher scores are indicative of poorer subjective sleep quality (Mulder-Hajonides van der Meulen & Van den Hoofdakker, 1990).

Pittsburgh Sleep Quality Index (PSQI): is a 19-item self-report questionnaire for the assessment of sleep quality. Participant rate the relatability of the presented items on a 0 to 3 scale. Summary scores range from 0 to 21, with higher scores indicating poorer sleep quality (Buysse, Reynolds III, Monk, Berman, & Kupfer, 1989).

State-Trait Anxiety Inventory - Trait (STAI-T): the Trait component of the STAI is a 20-item self-report questionnaire for the assessment of trait anxiety, as opposed to state anxiety which is assessed by the complementary part of the STAI. Respondents rate the relatability of the presented items on a 1 to 4 scale. Summary scores range from 20 to 80, with higher scores indicating higher trait anxiety (Spielberger, 2010).

Psychomotor tasks and visual perception tasks

Trail Making Test (TMT): The TMT is a pen-and-paper psychomotor test consisting of two parts. In part A, participants have to connect the numbers 1 to 25 in ascending order without lifting the pen as fast as possible. In part B, participants have to do the same, except that they have to alternate between the numbers 1 to 13 and the letter A to L (i.e. 1-A-2-B-3-C,…). Completion time serves as the main outcome measure. The maximum allowed completion time for part A and B were five and six minutes respectively (Reitan, 1958) .

Digit Symbol Substitution test (DSST): During the DSST, a list of nine abstract symbols is presented to the participants. Each symbol corresponds a number, i.e. 1 to 9. Below, the participant is presented with a long list with the numbers 1 to 9 in a random order. The participants are asked to fill in as many of the corresponding symbols below the number as they can within 90 seconds. Participants are instructed to complete the list sequentially. The total number of correct answers serves as the main outcome. A pen-and-paper version of the DSST was administered (Wechsler, 1955).

Adaptive Tachistoscopic Traffic Perception Test (ATTPT): The ATTPT was included in order to assess perceptual performance. During the ATTPT, a traffic scene is flashed on a computer screen. After the short scene presentation, participants need to indicate which elements were present, i.e. cyclists, cars, traffic lights and signs, and/or pedestrians. Response accuracy was used as the main outcome measure (Schuhfried, 2009).

Reaction test (RT): The RT is a computerized reaction time test that consists of three parts. In part 1 (RT1), participants need to respond as fast as possible when a circle presented in the middle of the screen lights up yellow. In the second part (RT2), participants need to respond as fast as possible to the presentation of a high pitched tone. The last part (RT3), requires participants to respond whenever the yellow circle and high pitched tone are presented simultaneously, and to withhold a response when this is not the case (e.g. circle lights up red, yellow light and tone are not simultaneously). The main outcome measure of this set of tasks is the mean reaction time (Prieler, 2008) .

Adaptive Determination Test (ADT): The determination test is a complex computerized reaction time test. Participants are presented with a grid of eight empty circles on a computer screen. During the test, the circles will light up in in any of 5 colors. Participants need to press the corresponding colored button as fast as possible on a specially equipped key board. Simultaneously, high and low pitched tones are presented. Each of the two pitches has a corresponding button on the keyboard that needs to be pressed as fast as possible following tone presentation. Furthermore, in

the right and left bottom corners of the screen, two rectangles are presented. When they light up, the participant needs to press the corresponding foot pedal. Stimulus presentation speeds up as performance gets better and slows down if reactions take longer or mistakes are made. Number of correct responses was used as the outcome measure. The task duration was set at 10 minutes (Neuwirth & Benesch, 2003) .

Psychomotor Vigilance Test (PVT): The PVT is a prolonged (10min) simple reaction time test. Participants press the response button when a stimulus is presented. Stimuli are presented at random intervals with an average of 8 seconds between subsequent stimulus presentations. Median reaction time and number of attentional lapses (reaction time > 500ms) serve as the outcome measures of this test. (Dinges & Powell, 1985)

Standardized on-the-road driving test

The standardized on the road driving test consists of a 100km drive on a highway in actual traffic in an instrumented test vehicle (O'Hanlon, 1984; Ramaekers, 2017; Verster & Roth, 2011). Participants are accompanied by a licensed driving instructor who has access to dual controls. During the drive, participants are instructed to keep a steady position in the right traffic lane and to maintain a steady speed of 95km/h. The lateral position of the vehicle relative to the traffic lane demarcation on its left and the velocity are logged every 250ms. From this, the standard deviation of speed (SDS) and the standard deviation of the lateral position (SDLP) are calculated. The SDLP serves as the main outcome measure and is essentially a measure of road tracking ability. Higher SDLP values correspond to more lane weaving, hence worse performance. The SDLP has been shown sensitive to the impairing effects of alcohol and acute BZRA administration on road tracking ability (Roth et al., 2014; Vermeeren, 2004; Verster et al., 2016a; Verster, Veldhuijzen, & Volkerts, 2004; Verster et al., 2005).

Procedure

All participants were informed about the purpose, procedure and associated risks of study participation prior to providing written informed consent. Hereafter, participants were medically screened based on a medical questionnaire that was reviewed by a clinician to ensure the minimal requirements for safely operating a vehicle in actual traffic were met. Next, participants were invited to a training day and a testing day. The training day served to familiarize the participants with the tasks in order to ascertain a smooth course of the testing day during which the data was collected. In addition, at the start of the training day participants completed the BDI, PSQI, and STAI-T, as well as a test of visual acuity, before practicing the test battery. At the start of the test day, medication, drugs, alcohol, nicotine, and caffeine use in the last 24 hours were documented. In addition, recent use of alcohol and common recreational and medicinal drugs was tested using an alcohol breathalyzer test and a urine drugs test respectively. Next, participants completed the GSQS and then completed the test battery in the same order as on the training day, i.e. pen-and-paper psychomotor tests (TMT and DSST), computerized perception and reaction time tests (ATTPT, RT, and ADT), the on-the-road driving test, and lastly the PVT. At the end of the testing day participants also completed a set of driving simulator scenarios. However, these data are not considered here. The study was approved by the medical ethics committee of the Maastricht University Medical Center and was executed in accordance to the declaration of Helsinki (1964) and its most recent amendments (2013). For a more detailed description of the testing procedure, the reader is referred to the study by van der Sluiszen et al. (van der Sluiszen et al., 2019) or the technical report (Verster et al., 2016b).

Statistical analysis

Dimension reduction

In order to simplify the interpretability of the various test outcomes (i.e. TMT, DSST, ATTPT, RT, ADT, PVT, and the on-the-road driving test) and to minimize the multiple testing problem, an ordinary least squares factor analysis (FA) was performed on the various outcome measures of both groups, i.e. BZRA users and controls, combined. The extracted solution was then rotated using Direct Oblimin rotation (δ=0). The factor scores for each participant were estimated using least squares regression and the resulting estimates were saved as a new variable. Missing variable scores were replaced by the mean before being entered in the FA. To determine the number of factors to extract, the parallel analysis method was employed (Franklin, Gibson, Robertson, Pohlmann, & Fralish, 1995). One thousand random permutations of the raw dataset were performed to determine the percentiles of the eigenvalues of random data. Eigenvalues of the raw dataset exceeding the eigenvalues of the permutated datasets marking the border of the 95th percentile were counted and the total number was used as the number of factors to be extracted during the FA.

Depression, insomnia, and anxiety are a cluster of highly comorbid disorders (Taylor, Lichstein, Durrence, Reidel, & Bush, 2005). Therefore, and in order to have parsimonious predictors in the general linear model (see below), the scores on the self-assessment clinical complaint questionnaires, i.e. BDI, STAI, PSQI, and GSQS, were bundled into one composite score using a principal component analysis. A parallel analysis also confirmed that a single component solution was appropriate. Component scores based on regression coefficients were saved as a clinical complaint composite score.

General linear modelling

Next, one-way analyses of covariance (ANCOVA) were performed for each extracted factor. Variables that loaded only moderately or weakly (r= [-0.5, 0.5]) on any of the factors yielded by the factor analysis were also considered separately as dependent variables. Group (i.e. BZRA user or

control) was entered as random factor. The clinical complaint composite score was entered as a continuous covariate. Also, estimated drug plasma concentration was entered as a categorical covariate (C_{BZRA}) with three levels, i.e. (0)no BZRA/control, (1)BZRA levels below the effective concentration (i.e. C_{BZRA}-) and (2)BZRA levels above the effective concentration (i.e. C_{BZRA}+). Finally, considering the wide age range of the participants, age was also included as a covariate. Non-significant covariates were first removed from the model before interpreting the results. Dunett's t-test was used for post hoc testing in case of multiple comparisons, with the control group as the appointed reference. Levene's test for equality of error variances, the White's test for heteroscedasticity, and normality checks were performed in order to assure that the model's assumptions were met. All statistical analyses were conducted using the Statistical Package for Social Sciences (SPSS)(IBM corporation, 2017).

Results

Missing data

The test score of the TMT part B could not be retrieved for one BZRA user of the C_{BZRA}+ subgroup. For another BZRA user in this subgroup, the PVT data could not be recovered. Two control participants and one BZRA user of the C_{BZRA}+ subgroup did not complete the on-the-road driving test because of technical difficulties with the test vehicle. Therefore, no SDLP or SDS data was available for these participants. For another additional three control participants and one BZRA user in the C_{BZRA}+ subgroup no velocity data of the on-the-road driving test was available, hence the SDS could not be determined for these participants.

Dimension reduction

Based on the parallel analysis, it was decided that a three factor solution was appropriate for the factor analysis. The correlation coefficients of the variables and the factors after the Direct Oblimin rotation are shown in Table 3. The first factor was found to correlate highly with

performance on single response reaction time tasks, i.e. RT and PVT, and was labelled *Response latency*. The second factor seems to capture the performance on the TMT, DSST, and ADT, all of which depend on the speed of stimulus processing for the selection of the appropriate response, and was therefore labelled *Processing speed*. The third factor is explained mostly by performance on the ADT, PVT and on-the-road driving test. It is proposed that these tasks share a requirement of *sustained attention*, hence the factor was labeled as such.

	Response latency	**Processing speed**	**Sustained attention**
TMT A			
mean completion time (Sec)	.304**	-.782**	.181**
TMT B			
mean completion time (Sec)	.345**	-.666**	.023
DSST			
correct responses (#)	-.352**	.761**	.279**
ATTPT			
accuracy (%)	-.003	.092	.073
RT1			
mean reaction time (msec)	.941**	-.291**	-.076
RT2			
mean reaction time (msec)	.773**	-.213**	-.052
RT3			
mean reaction time (msec)	.679**	-.427**	-.033
ADT			
correct responses (#)	-.443	.749**	.424**
PVT			
median reaction time (msec)	.632**	-.310**	-.412**
lapses (#)	.604**	-.299**	-.352**
On-the-road driving test			
SDLP (cm)	.184*	-.098	-.467**
SDS (km/h)	.307**	-.193*	-.222**

Table 3. Structure matrix obtained after the factor analysis of the listed outcome measures. The Pearson correlation coefficients quantifying the linear relation between each outcome measure with the three extracted factors are shown; * $p < .05$; ** $p < .01$.

The SDLP was found to correlate only moderately with its main factor, *Sustained attention*, and was therefore also considered separately as a dependent variable in the group comparisons, as was the SDS. The ATTPT response accuracy was also considered separately as dependent variable because it correlated very weakly with any of the extracted factors.

Finally, the parallel analysis confirmed that a single component solution would yield an optimal solution for the principal component analysis of the clinical complaint measures. The component loadings of the BDI (r= .840, p <.01), PSQI (r= .871, p<.01), STAI-T (r= .882, p< .01), and GSQS (r= .693, p<.01) were all indicative of a strong and positive relationships with the principal component.

Group comparisons

The ANCOVA of *Response latency*, found that Age (F(1,125)= 6.3, p= .01, *f*= .22) and Group (F(1,125)= 9.22, p < .01, *f*= .27) significantly contributed to the model. Inspection of the estimated marginal means revealed that controls (M=-0.2, SE= 0.1) seemed to outperform BZRA users (M=0.3, SE= 0.13, p < .003). The clinical complaint composite score (F(1,121)= 0.23, p= .88) and C_{BZRA} (F(1,121)= 0.56, p= .81) did not contribute significantly to the model (Figure 1a).

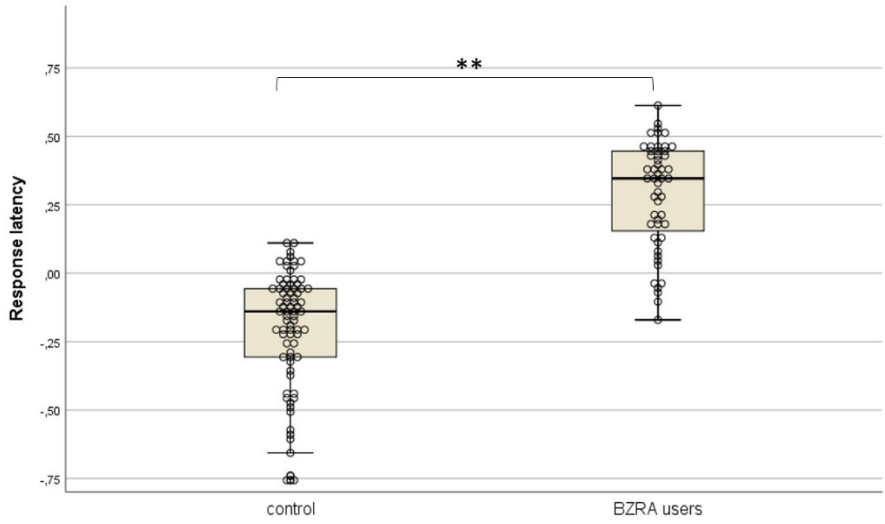

Figure 1. Unstandardized predicted values of factor scores (response latency) corrected for the effect of age. *Estimated marginal mean difference was found to be statistically significant (p<.05).**Estimated marginal mean difference was found to be statistically significant(p<.01).

The analysis of *Processing speed* revealed both factor Group (F(1,125)= 96.97, p < .001, *f*= .88) and the covariate Age (F(1,125)= 48.08, p < .001, *f*= .62) were significant predictors, while C_{BZRA} did not significantly contribute to the model (F(1,121)= 1.29, p= .26). The clinical complaint composite score was found to be marginally significant (F(1,121)= 3.21, p= .08), and was therefore

initially kept in the model. However, after removal of the estimated drug concentration covariate, the contribution of the clinical complaint composite score was clearly not significant (F(1,121)= 2.3, p= .13) and the covariate was therefore removed from the model. Inspection of the estimated marginal means showed that BZRA users (M= -0.48, SE=0.09) performed significantly worse compared to controls (M=0.33, SE=0.08, p < .001) (Figure 2).

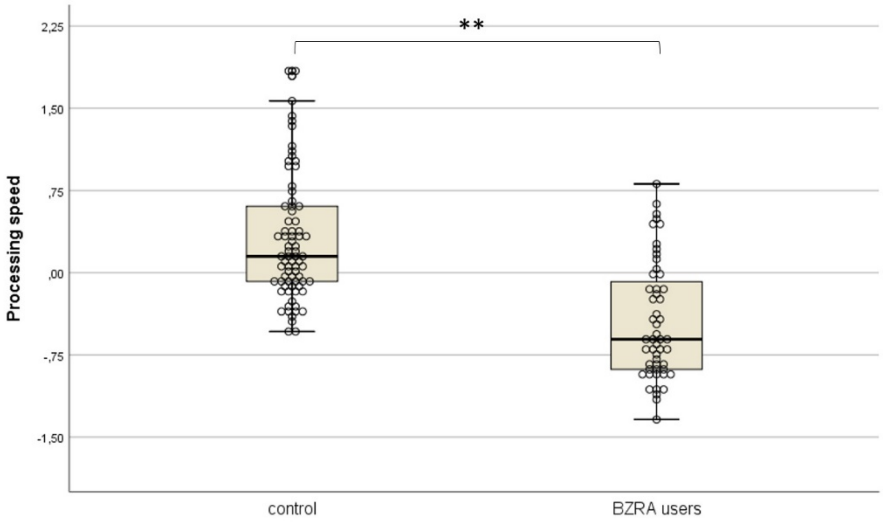

Figure 2. Unstandardized predicted values of factor scores (processing speed) corrected for the effect of age. *Estimated marginal mean difference was found to be statistically significant (p<.05).**Estimated marginal mean difference was found to be statistically significant(p<.01).

For the ANCOVA of *Sustained attention*, the covariates C_{BZRA} (F(1,121)= 0.21, p = .644) and clinical complaint composite score (F(1,121)= 0.88, p = .35) did not contribute significantly to the model and were removed. After their removal, it was found that both the covariate Age (F(1,125)= 5.28, p=.023, *f*= .21) and the random factor Group (F(1,125)= 9.04, p= .003, *f*= .27) significantly predicted performance of this factor, with controls (M=0.16, SE= 0.08) outperforming BZRA users (M= -0.24, SE= 0.13, p= .003)(Figure 3).

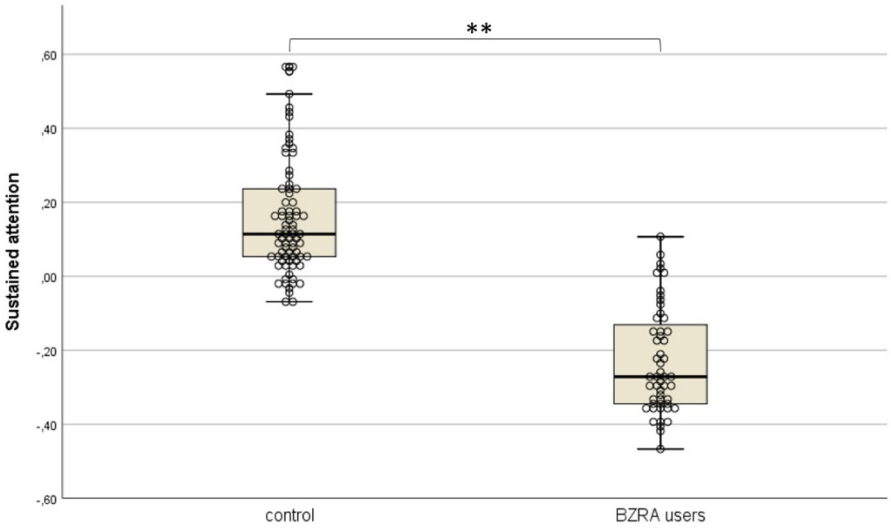

Figure 3. Unstandardized predicted values of factor scores corrected for the effect of age. (A) Response latency: higher scores indicate worse performance. (B) Processing speed: higher scores mean better performance. (C) Sustained attention: higher scores indicate better performance. *Estimated marginal mean difference was found to be statistically significant (p<.05).**Estimated marginal mean difference was found to be statistically significant(p<.01).

The separate analysis of *SDLP* demonstrated that Age (F(1,118)=0.5, p= .48) and clinical complaint composite score (F(1,118)=0.36, p= .55) were non-significant predictors and were removed from the model. After their removal, it was found that C_{BZRA} (F(1,120)= 5.38, p= .022) significantly predicted the SDLP. However, after the correction for this covariate, it was found that the random factor Group was not significant (F(1,120)= 2.48, p= .118). A post hoc one-way ANOVA with C_{BZRA} as the sole factor again confirmed that it significantly predicted SDLP (F(2,120)= 3,52, p= .033). Post hoc, Dunnett's t-test revealed that the C_{BZRA}+ subgroup had significantly higher SDLP values (M=20.74, SE= 0.91) compared to control participants (M=18.19, SE= 0.46, p= .012, *d*= 0.65), while the difference between controls and the C_{BZRA}- subgroup was insignificant (M= 18.14, SE= 0.58, p= .737) (Figure 4a).

The SDS was also analyzed separately because of its weak correlation with any of the factors. It was found that neither clinical complaint composite score (F(1,114)=0.15, p= .697) or C_{BZRA} (F(1,114)= 0.05, p= .812) significantly contributed to the model and were removed. Hereafter it was found that both Age (F(1,118)=6.4, p= .013, *f*= 0.23) and Group (F(1,118)= 9.01 p= .003, *f*= 0.28) were significant predictors. Control participants (M= 2.39, SE= 0.09) outperformed BZRA users (M= 2.81, SE= 0.11, p= .003) (Figure 4b).

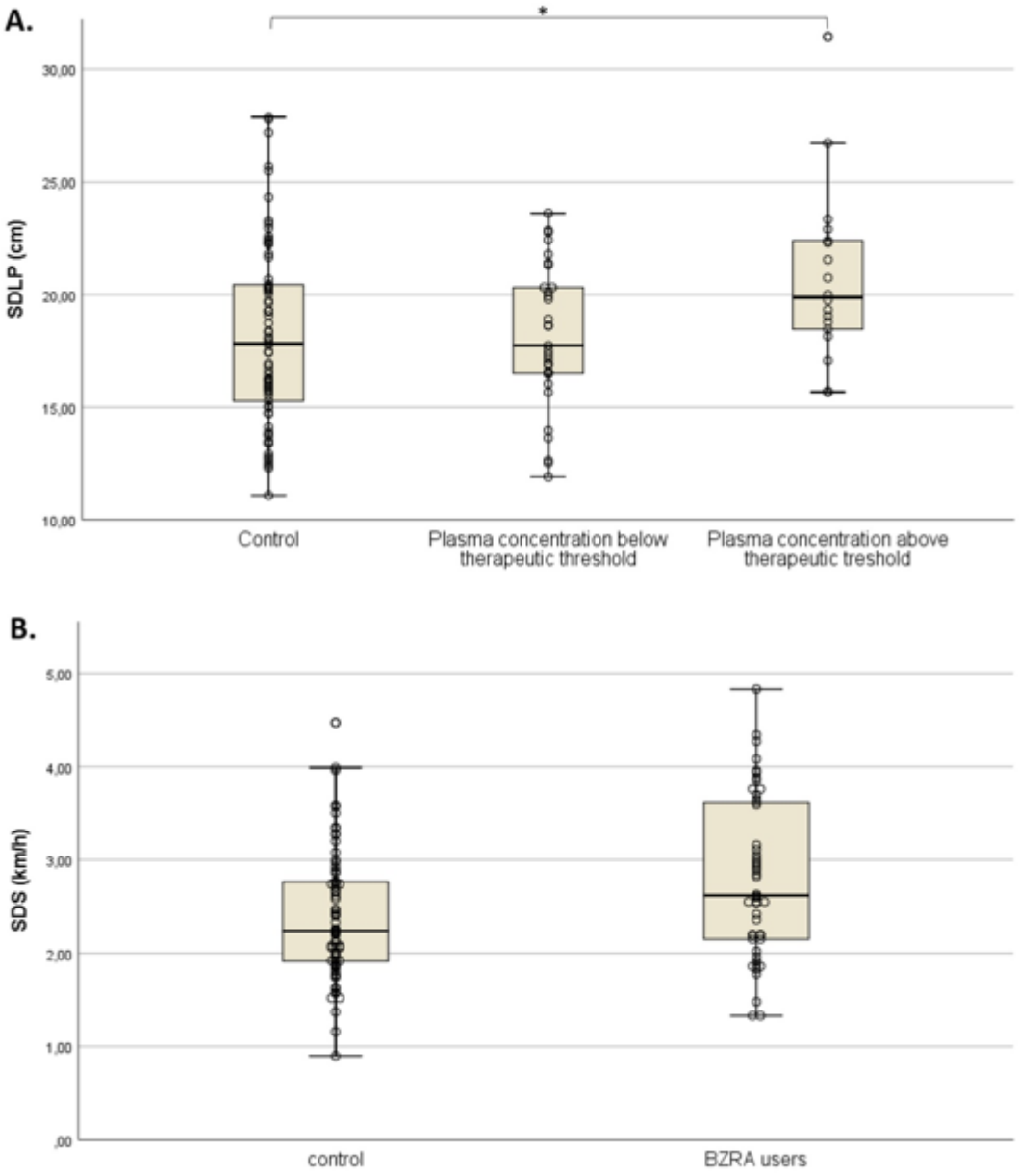

Figure 4. (A) The SDLP values of the on-the-road driving test for controls and benzodiazepine receptor agonist (BZRA) users grouped by estimated BZRA plasma concentrations, i.e. estimated BZRA plasma concentrations below the therapeutic threshold (C_{BZRA}-) and BZRA plasma concentrations exceeding the therapeutic threshold (C_{BZRA}+). (B) Unstandardized predicted values of SDS of controls and patients corrected for the effect of Age. *Estimated marginal mean difference relative to control was found to be statistically significant (p<.05). **Estimated marginal mean difference was found to be statistically significant(p<.01).

A separate ANCOVA with ATTPT accuracy as the dependent variable showed that neither age ($F(1,121)= 2.81$, $p= .096$), clinical complaint composite score ($F(1,121)= 0.21$, $p= .649$), nor C_{BZRA} ($F(1,121)= 0$, $p= .995$) significantly contributed to the model and were consequentially removed. Hereafter, it was found that no performance differences were apparent between the BZRA users and control group ($F(1,126)= .94$, $p= .333$).

A post-hoc comparison of self-reported alcohol use between the BZRA user groups and control participants was performed. Control participants (Mdn= 4.5) indicated to consume significantly more alcoholic beverages per week than BZRA users (Mdn=3 U= 1479.5, p= .015), as demonstrated by a Mann-Whitney U-test. The two BZRA user groups (C_{BZRA}- : Mdn= 2, C_{BZRA}+: Mdn= 4, U= 282.5, p= .804) did not differ with respect to self-reported number of alcoholic beverages per week.

Discussion

The goal of the current analysis was to investigate the neurocognitive and driving performance of long-term, regular BZRA users, controlled for estimated BZRA plasma concentrations and severity of clinical complaints. In addition, considering the wide age range of participants and the known effects of age on cognitive and psychomotor functioning, age was also controlled for. Performance of long-term and regular BZRA users on various psychomotor tasks and a standardized on-the-road driving test was compared to that of healthy controls.

Ordinary least squares factor analysis yielded a three factor solution. The first factor, *Response latency,* correlated strongly and positively with simple reaction time tests consisting of one type of target stimulus and one response option. The second factor, *Processing speed,* correlated strongly with tasks that share a level of complexity relative to simple reaction time tasks in that multiple response options are presented from which the correct one should be selected as fast as possible. This operation requires higher level functioning such as conscious stimulus identification, working memory and response matching. Hence, performance depends on the speed with which a stimulus is identified and the correct response is selected. The third factor, *Sustained attention,* was best explained by the ADT, PVT, and on-the-road driving test performance, albeit modest. Arguably, what these tasks share is the requirement of sustained attention. The ADT places high demands on the participants' attentional resources. The stimulus presentation rate adapts to the participants performance which assures a constant high level of difficulty. This arguably induces mental fatigue.

The PVT and on-the-road driving tests are prolonged, monotonous tasks which require effort to remain vigilant.

Separate analyses of covariance demonstrated that healthy controls performed better than BZAR users on all three factors. However, clinical symptoms and estimated BZRA plasma concentration grouped relative to the therapeutic threshold did not significantly contribute to the model, which suggests that there is no direct influence of BZRA use or clinical symptomatology on neurocognitive performance. Similar results were found for the separate analysis of SDS. A separate analysis of the ATTPT failed to show any difference in accuracy between groups or as a function of any of the covariates.

The SDLP, the main outcome of the on-the-road driving test, did not load strongly on any of the factors and was therefore considered separately. This analysis showed that when the difference between controls and BZRA users was corrected for the effect of estimated BZRA plasma concentrations, the group difference was no longer apparent. A post hoc analysis of variance and subsequent multiple comparisons demonstrated that BZRA users with estimated BZRA plasma concentration exceeding the therapeutic threshold exhibited significantly higher SDLP values. This suggests that residual acute effects of BZRA use can have a negative impact on road tracking ability, which contrasts the results regarding the Sustained attention factor, where residual acute drug effects were not found to significantly explain performance differences, despite the SDLP being one of the greater contributors to that factor. Arguably, what differentiates the SDLP from the Sustained attention factor is the task duration and the stimuli and response modes.

The duration of the on-road driving test is significantly longer than the duration of the other two largest contributors, the ADT and PVT. This characteristic is conceivably of central importance to detect any residual acute drug effects caused by BZRA use. During the driving test, SDLP increases as a function of time on task (Verster & Roth, 2011, 2013b), likely due to driver fatigue. It is plausible that BZRA users might be less able to counteract fatigue during prolonged monotonous tasks. Also, the stimuli and response modes might explain the discrepant findings. Unlike the PVT and ADT, the

on-road driving test does not involve discrete stimuli and response options. The “stimulus” during the on-road driving test is the deviation from the aimed direction. This is not a discrete stimulus since the magnitude of the deviation is of central importance for determining the magnitude of the correction, which also makes the steering correction a graded, non-discrete response. Arguably, this added level of complexity, together with the prolonged test duration, makes the on-road driving test the most sensitive test in the battery to pick up impairments in sustained attention.

The mean SDLP was found to be 2.55cm higher compared to the control group. Previous research demonstrated that a difference of 2.5cm is comparable to the increase in SDLP observed at blood-alcohol concentration (BAC) of 0.5g/L compared to placebo (Jongen et al., 2017). This BAC has been found to be associated with a significant increase in crash risk (Borkenstein, Crowther, & Shumate, 1974). It follows that an increase in SDLP of ≥2.5cm implies a significantly increase in crash risk (Ramaekers, 2017). In conclusion, long-term and regular BZRA users with BZRA plasma concentration exceeding the therapeutic threshold show impaired road-tracking ability which potentially increases their respective crash risk.

Overall, the results confirm the previous findings by Barker et al. (2004a, 2004b) that long-term benzodiazepine users are impaired on tests of psychomotor functioning compared to healthy controls. In addition, the current results suggest that the observed impairments cannot always be explained by residual acute effects of BZRA’s, with the exception of road tracking (SDLP). This later notion suggests that residual acute BZRA effects can still play a causal role in the observed performance differences between BZRA users and controls during prolonged tasks with indiscrete stimuli and graded response requirements.

Despite the previous findings that BZRA blood plasma concentrations appeared to correlate poorly with road tracking performance during the standardized on-road driving test (Verster & Roth, 2013a), the current findings suggests that BZRA blood plasma concentrations can be useful for the estimation of drug effects when interpreted relative to the therapeutic threshold, i.e. in a binary sense rather than as a continuous linear correlate. However, it should be stressed that the BZRA

plasma concentration are estimated based on average pharmacokinetic parameters of the respective drugs as described in the literature. Absorption, distribution, metabolism, and excretion are generally subject to considerable inter-individual variation. Also, age is known to slow metabolic rate of drugs, but was not taken into account for the estimations. Another potential limitation regarding the BZRA estimates is the role of co-medications. Despite the observation that the relative frequency of number and types of CNS co-medications was similar in both BZRA user groups, a more detailed evaluation of potential hepatic drug interactions was not performed. The potential of certain co-medications to inhibit or enhance the actions of the cytochrome P450 enzymes necessary for the breakdown of BZRAs might have resulted in over- or under estimations of the BZRA plasma concentrations. Future research should ideally include actual blood plasma samples for the determination of the BZRA plasma concentrations. However, provided the comparable age distribution and large difference in estimated equivalent BZRA plasma concentrations between the two BZRA user subgroups, the estimates are arguably sufficiently accurate to allow for a group comparison. Another conceivable shortcoming of this investigation is the regular consumption of alcoholic beverages. Although the regular consumption of alcohol is unlikely to have contributed to the observed impairments in BZRA users, since BZRA users reported to have less drinks per week than controls, it might have attenuated the differences in performance. This is because of the notion of cross-tolerance for the effects of ethanol and BZRAs, which has been demonstrated in rodents (Chan, Schanley, Aleo, & Leong, 1985; Le, Khanna, Kalant, & Grossi, 1986). For now it remains unclear at what dose and frequency alcohol consumption can induce cross-tolerance in man. It is therefore plausible that some BZRA users, especially those with the maximally allowed weekly alcohol intake, might experience less functional impairment introduced by the use of their respective BZRA.

Despite the observation that the Clinical Complaint Composite Score was not found to significantly predict performance, the potential role of clinical complaints cannot be excluded (van der Sluiszen et al., 2017; Verster & Roth, 2014; Wingen, Ramaekers, & Schmitt, 2006). It should be noted that the applied depression, anxiety, and sleep quality assessments that make up the

complaint composite score predominantly inquire about the BZRA users' experiences in the recent past (i.e. yesterday, last few days, weeks, and months) and were administered before execution of the test battery. Hence, none of the applied questionnaires were used to quantify task related stress during or immediately after completion of the tasks. Increased anxiety in response to being subjected to a test is generally known as test anxiety and has been repeatedly demonstrated to be associated with poorer test performance (Cook et al., 2007; Eysenck, 1985; Hembree, 1988) and too often co-exist with anxiety and depression related complaints (Akinsola & Nwajei, 2013). Research by Michael Eysenck and others (Eysenck, 1985) has demonstrated that performance on tasks requiring higher order cognitive functions is more affected by task-related anxiety than performance on tasks drawing on low level functions. In line with this it was found that the impairment of Processing speed, the factor summarizing the most complex tasks, due to the effect of Group was of large magnitude (f= .88), while the magnitude of impairments on the lower level Response latency and Sustained attention factors were small to medium (*f*= .27 in both instances). It therefore remains plausible that the performance of BZRA users suffered from test anxiety. Future research should include measures of task related stress such as the Dundee stress state questionnaire(Matthews et al., 1999) in order to more specifically investigate the role of this potential confounder.

Conclusion

Long-term regular BZRA users are impaired on various tasks of psychomotor functioning. Specifically, it was found that response latency, processing speed during more complex psychomotor tasks, and sustained attention are impaired as compared to healthy controls. These impairments were not explained by estimated BZRA plasma concentrations, which makes the role of acute drug effects seem less likely. However, road tracking during a standardized driving test in real traffic did appear to be significantly impaired in BZRA users with therapeutically relevant estimated BZRA levels, but not for the other BZRA user subgroup with low estimated BZRA levels. The magnitude of the effect was comparable to the impairment observed while driving at a BAC of 0.5g/L. These findings

suggest that residual acute BZRA effects are relevant to task performance, depending on the tasks characteristics. Severity of depressive, anxiety, and sleep complaints combined did not explain any of the observed impairments. However, it is argued that based on the applied assessment instruments, the role of acute task related stress cannot be ruled out. Future research should include assessment of experienced task related stress in order to control for this potential confounder. It is also advised to collect blood plasma samples in order to prevent any inaccuracies inherent to the estimations of the BZRA plasma concentrations based on medication profiles.

References

Akinsola, E. F., & Nwajei, A. D. (2013). Test anxiety, depression and academic performance: assessment and management using relaxation and cognitive restructuring techniques. *Psychology, 4*(06), 18.

Ashton, H. (1994). Guidelines for the rational use of benzodiazepines. *Drugs, 48*(1), 25-40.

Barker, M. J., Greenwood, K. M., Jackson, M., & Crowe, S. F. (2004a). Cognitive effects of long-term benzodiazepine use. *CNS drugs, 18*(1), 37-48.

Barker, M. J., Greenwood, K. M., Jackson, M., & Crowe, S. F. (2004b). Persistence of cognitive effects after withdrawal from long-term benzodiazepine use: a meta-analysis. *Archives of Clinical Neuropsychology, 19*(3), 437-454.

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical psychology review, 8*(1), 77-100.

Berlin, A., & Dahlström, H. (1975). Pharmacokinetics of the anticonvulsant drug clonazepam evaluated from single oral and intravenous doses and by repeated oral administration. *European journal of clinical pharmacology, 9*(2-3), 155-159.

Borkenstein, R. F., Crowther, R. F., & Shumate, R. (1974). The role of the drinking driver in traffic accidents (The Grand Rapids Study). *Blutalkohol, 11*(Suppl), 1-131.

Bushnell, G. A., Stürmer, T., Gaynes, B. N., Pate, V., & Miller, M. (2017). Simultaneous antidepressant and benzodiazepine new use and subsequent long-term benzodiazepine use in adults with depression, United States, 2001-2014. *JAMA psychiatry, 74*(7), 747-755.

Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry research, 28*(2), 193-213.

Caille, G., Du Souich, P., Spenard, J., Lacasse, Y., & Vezina, M. (1984). Pharmacokinetic and clinical parameters of zopiclone and trimipramine when administered simultaneously to volunteers. *Biopharmaceutics & drug disposition, 5*(2), 117-125.

Chan, A. W., Schanley, D. L., Aleo, M. D., & Leong, F. W. (1985). Cross-tolerance between ethanol and chlordiazepoxide. *Alcohol, 2*(2), 209-213.

Cook, J. M., Biyanova, T., Masci, C., & Coyne, J. C. (2007). Older patient perspectives on long-term anxiolytic benzodiazepine use and discontinuation: a qualitative study. *Journal of general internal medicine, 22*(8), 1094-1100.

Crevoisier, C., Delisle, M., Joseph, I., & Foletti, G. (2003). Comparative single-dose pharmacokinetics of clonazepam following intravenous, intramuscular and oral administration to healthy volunteers. *European neurology, 49*(3), 173-177.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior research methods, instruments, & computers, 17*(6), 652-655.

Divoll, M., Greenblatt, D. J., Harmatz, J. S., & Shader, R. I. (1981). Effect of age and gender on disposition of temazepam. *Journal of pharmaceutical sciences, 70*(10), 1104-1107.

Divoll, M., Greenblatt, D. J., Ochs, H. R., & Shader, R. I. (1983). Absolute bioavailability of oral and intramuscular diazepam: effects of age and sex. *Anesthesia and analgesia, 62*(1), 1-8.

Eatman, F., Colburn, W., Boxenbaum, H., Posmanter, H., Weinfeld, R., Ronfeld, R., . . . Kaplan, S. (1977). Pharmacokinetics of diazepam following multiple-dose oral administration to healthy human subjects. *Journal of pharmacokinetics and biopharmaceutics, 5*(5), 481-494.

Eysenck, M. W. (1985). Anxiety and cognitive-task performance. *Personality and Individual Differences, 6*(5), 579-586. doi:https://doi.org/10.1016/0191-8869(85)90007-8

Fernandez, C., Martin, C., Gimenez, F., & Farinotti, R. (1995). Clinical pharmacokinetics of zopiclone. *Clinical pharmacokinetics, 29*(6), 431-441.

Fortier-Brochu, É., Beaulieu-Bonneau, S., Ivers, H., & Morin, C. M. (2012). Insomnia and daytime cognitive performance: a meta-analysis. *Sleep medicine reviews, 16*(1), 83-94.

Franklin, S. B., Gibson, D. J., Robertson, P. A., Pohlmann, J. T., & Fralish, J. S. (1995). Parallel analysis: a method for determining significant principal components. *Journal of Vegetation Science, 6*(1), 99-106.

Gaillot, J., Heusse, D., Hougton, G., Aurele, J. M., & Dreyfus, J. (1983). Pharmacokinetics and metabolism of zopiclone. *Pharmacology, 27*(Suppl. 2), 76-91.

Gerlach, L. B., Maust, D. T., Leong, S. H., Mavandadi, S., & Oslin, D. W. (2018). Factors associated with long-term benzodiazepine use among older adults. *JAMA internal medicine, 178*(11), 1560-1562.

Greenblatt, D. J. (1981). Clinical pharmacokinetics of oxazepam and lorazepam. *Clinical pharmacokinetics, 6*(2), 89-105.

Greenblatt, D. J., Abernethy, D. R., Locniskar, A., Harmatz, J. S., Limjuco, R. A., & Shader, R. I. (1984). Effect of age, gender, and obesity on midazolam kinetics. *Anesthesiology: The Journal of the American Society of Anesthesiologists, 61*(1), 27-35.

Greenblatt, D. J., Abernethy, D. R., Locniskar, A., Ochs, H. R., Harmatz, J. S., & Shader, R. I. (1985). Age, sex, and nitrazepam kinetics: relation to antipyrine disposition. *Clinical Pharmacology & Therapeutics, 38*(6), 697-703.

Greenblatt, D. J., Allen, M. D., Harmatz, J. S., & Shader, R. I. (1980). Diazepam disposition determinants. *Clinical Pharmacology & Therapeutics, 27*(3), 301-312.

Greenblatt, D. J., Blaskovich, P. D., Nuwayser, E., Harmatz, J. S., Chen, G., & Zinny, M. A. (2005). Clonazepam Pharmacokinetics: Comparison of Subcutaneous Microsphere Injection With Multiple-Dose Oral Administration. *The Journal of Clinical Pharmacology, 45*(11), 1288-1293.

Greenblatt, D. J., Divoll, M., Harmatz, J. S., & Shader, R. I. (1982). Pharmacokinetic comparison of sublingual lorazepam with intravenous, intramuscular, and oral lorazepam. *Journal of pharmaceutical sciences, 71*(2), 248-252.

Greenblatt, D. J., Harmatz, J. S., Friedman, H., Locniskar, A., & Shader, R. I. (1989). A large-sample study of diazepam pharmacokinetics. *Therapeutic drug monitoring, 11*(6), 652-657.

Greenblatt, D. J., Harmatz, J. S., Roth, T., Singh, N. N., Moline, M. L., Harris, S. C., & Kapil, R. P. (2013). Comparison of pharmacokinetic profiles of zolpidem buffered sublingual tablet and zolpidem oral immediate-release tablet: results from a single-center, single-dose, randomized, open-label crossover study in healthy adults. *Clinical therapeutics, 35*(5), 604-611.

Greenblatt, D. J., & Wright, C. E. (1993). Clinical pharmacokinetics of alprazolam. *Clinical pharmacokinetics, 24*(6), 453-471.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of educational research, 58*(1), 47-77.

Hildebrand, M., Hellstern, A., Hümpel, M., Hellenbrecht, D., & Saller, R. (1990). Plasma levels and urinary excretion of lormetazepam in patients with liver cirrhosis and in healthy volunteers. *European journal of drug metabolism and pharmacokinetics, 15*(1), 19-26.

IBM corporation. (2017). IBM SPSS Statistics for Windows (Version 25.0 Armonk). *NY: IBM Corp*.

Jochemsen, R., Hogendoorn, J., Dingemanse, J., Hermans, J., Boeijinga, J., & Breimer, D. (1982). Pharmacokinetics and bioavailability of intravenous, oral, and rectal nitrazepam in humans. *Journal of pharmacokinetics and biopharmaceutics, 10*(3), 231-245.

Jochemsen, R., Wesselman, J., Hermans, J., Van Boxtel, C., & Breimer, D. (1983). Pharmacokinetics of brotizolam in healthy subjects following intravenous and oral administration. *British journal of clinical pharmacology, 16*(S2), 285S-290S.

Jongen, S., Vermeeren, A., van der Sluiszen, N. N. J. J. M., Schumacher, M. B., Theunissen, E. L., Kuypers, K. P. C., . . . Ramaekers, J. G. (2017). A pooled analysis of on-the-road highway driving studies in actual traffic measuring standard deviation of lateral position (i.e., "weaving") while driving at a blood alcohol concentration of 0.5 g/L. *Psychopharmacology, 234*(5), 837-844. doi:10.1007/s00213-016-4519-z

Jongen, S., Vuurman, E. F. P. M., Ramaekers, J. G., & Vermeeren, A. (2018). Comparing the effects of oxazepam and diazepam in actual highway driving and neurocognitive test performance: a validation study. *Psychopharmacology, 235*(4), 1283-1294. doi:10.1007/s00213-018-4844-5

Kampf, D., Huempel, M., Lerche, U., & Kessel, M. (1981). Effects of uremia and hemodialysis on lormetazepam disposition. *Clinical Pharmacology & Therapeutics, 30*(1), 77-85.
Langley, M. S., & Clissold, S. P. (1988). Brotizolam. *Drugs, 35*(2), 104-122.
Langtry, H. D., & Benfield, P. (1990). Zolpidem. *Drugs, 40*(2), 291-313.
Le, A., Khanna, J., Kalant, H., & Grossi, F. (1986). Tolerance to and cross-tolerance among ethanol, pentobarbital and chlordiazepoxide. *Pharmacology Biochemistry and Behavior, 24*(1), 93-98.
Leufkens, T. R., & Vermeeren, A. (2014). Zopiclone's residual effects on actual driving performance in a standardized test: a pooled analysis of age and sex effects in 4 placebo-controlled studies. *Clinical therapeutics, 36*(1), 141-150.
Lombardo, F., Obach, R. S., Shalaeva, M. Y., & Gao, F. (2002). Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data. *Journal of medicinal chemistry, 45*(13), 2867-2876.
Lucki, I., Rickels, K., & Geller, A. M. (1986). Chronic use of benzodiazepines and psychomotor and cognitive test performance. *Psychopharmacology, 88*(4), 426-433.
Malacrida, R., Fritz, M. E., Suter, P. M., & Crevoisier, C. (1992). Pharmacokinetics of midazolam administered by continuous intravenous infusion to intensive care patients. *Critical care medicine, 20*(8), 1123-1126.
Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state big three. *Personality psychology in Europe, 7*, 335-350.
McDermott, L. M., & Ebmeier, K. P. (2009). A meta-analysis of depression severity and cognitive function. *Journal of affective disorders, 119*(1-3), 1-8.
Mulder-Hajonides van der Meulen, R., & Van den Hoofdakker, R. (1990). The Groningen sleep quality scale. *Scan. J. Psych*.
Neuwirth, W., & Benesch, M. (2003). Determination Test. *Dr G. Schuhfried GmbH, Mödling, Austria*.
Noble, S., Langtry, H. D., & Lamb, H. M. (1998). Zopiclone. *Drugs, 55*(2), 277-302.
O'Hanlon, J. (1984). Driving performance under the influence of drugs: rationale for, and application of, a new test. *British journal of clinical pharmacology, 18*(S1), 121S-129S.
Ochs, H. R., Steinhaus, E., Locniskar, A., Knüchel, M., & Greenblatt, D. J. (1982). Desmethyldiazepam kinetics after intravenous, intramuscular, and oral administration of clorazepate dipotassium. *Klinische Wochenschrift, 60*(8), 411-415.
Olubodun, J. O., Ochs, H. R., Trüten, V., Klein, A., von Moltke, L. L., Harmatz, J. S., . . . Greenblatt, D. J. (2002). Zolpidem pharmacokinetic properties in young females: influence of smoking and oral contraceptive use. *The Journal of Clinical Pharmacology, 42*(10), 1142-1146.
Owen, R., & Tyrer, P. (1983). Benzodiazepine dependence. *Drugs, 25*(4), 385-398.
Paterniti, S., Dufouil, C., & Alpérovitch, A. (2002). Long-term benzodiazepine use and cognitive decline in the elderly: the Epidemiology of Vascular Aging Study. *Journal of clinical psychopharmacology, 22*(3), 285-293.
Peirce, G. L., Smith, M. J., Abate, M. A., & Halverson, J. (2012). Doctor and pharmacy shopping for controlled substances. *Medical care*, 494-500.
Pomara, N., Tun, H., DaSilva, D., & Hernando, R. (1998). The acute and chronic performance effects of alprazolam and lorazepam in the elderly: relationship to duration of treatment and self-rated sedation. *Psychopharmacology bulletin, 34*(2), 139.
Prieler, J. (2008). Reaction test. *RT Test Manual version, 31*.
Ramaekers, J. G. (2017). Drugs and driving research in medicinal drug development. *Trends in pharmacological sciences, 38*(4), 319-321.
Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and motor skills, 8*(3), 271-276.
Roth, T., Eklov, S. D., Drake, C. L., & Verster, J. C. (2014). Meta-analysis of on-the-road experimental studies of hypnotics: effects of time after intake, dose, and half-life. *Traffic injury prevention, 15*(5), 439-445.

Salvà, P., & Costa, J. (1995). Clinical pharmacokinetics and pharmacodynamics of zolpidem. *Clinical pharmacokinetics, 29*(3), 142-153.
Scavone, J., Greenblatt, D., Harmatz, J., & Shader, R. (1986). Kinetic and dynamic interaction of brotizolam and ethanol. *British journal of clinical pharmacology, 21*(2), 197-204.
Schuhfried, G. (2009). Adaptive Tachistoscopic Traffic Perception Test. *ATAVT Test Manual version, 22*.
Schulz, M., Iwersen-Bergmann, S., Andresen, H., & Schmoldt, A. (2012). Therapeutic and toxic blood concentrations of nearly 1,000 drugs and other xenobiotics. *Critical care, 16*(4), R136.
Schwarz, H. (1979). Pharmacokinetics and metabolism of temazepam in man and several animal species. *British journal of clinical pharmacology, 8*(Suppl 1), 23S.
Shader, R. I., Greenblatt, D. J., Ciraulo, D. A., Divoll, M., Harmatz, J. S., & Georgotas, A. (1981). Effect of age and sex on disposition of desmethyldiazepam formed from its precursor clorazepate. *Psychopharmacology, 75*(2), 193-197.
Shinfuku, M., Kishimoto, T., Uchida, H., Suzuki, T., Mimura, M., & Kikuchi, T. (2019). Effectiveness and safety of long-term benzodiazepine use in anxiety disorders: a systematic review and meta-analysis. *International clinical psychopharmacology, 34*(5), 211-221.
Smith, R., Kroboth, P., Vanderlugt, J., Phillips, J., & Juhl, R. (1984). Pharmacokinetics and pharmacodynamics of alprazolam after oral and IV administration. *Psychopharmacology, 84*(4), 452-456.
Sonne, J., Loft, S., Døssing, M., Vollmer-Larsen, A., Olesen, K., Victor, M., . . . Andreasen, P. (1988). Bioavailability and pharmacokinetics of oxazepam. *European journal of clinical pharmacology, 35*(4), 385-389.
Sorock, G. S., & Shimkin, E. E. (1988). Benzodiazepine sedatives and the risk of falling in a community-dwelling elderly cohort. *Archives of Internal Medicine, 148*(11), 2441-2444.
Spielberger, C. D. (2010). State-Trait anxiety inventory. *The Corsini encyclopedia of psychology*, 1-1.
Taylor, D. J., Lichstein, K. L., Durrence, H. H., Reidel, B. W., & Bush, A. J. (2005). Epidemiology of Insomnia, Depression, and Anxiety. *Sleep, 28*(11), 1457-1464. doi:10.1093/sleep/28.11.1457
Uzun, S., Kozumplik, O., Jakovljević, M., & Sedić, B. (2010). Side effects of treatment with benzodiazepines. *Psychiatria Danubina, 22*(1), 90-93.
van der Sluiszen, N. N. J. J. M., Vermeeren, A., Verster, J. C., van de Loo, A. J. A. E., van Dijken, J. H., Veldstra, J. L., . . . Ramaekers, J. G. (2019). Driving performance and neurocognitive skills of long-term users of benzodiazepine anxiolytics and hypnotics. *Human Psychopharmacology: Clinical and Experimental, 34*(6), e2715. doi:10.1002/hup.2715
van der Sluiszen, N. N. J. J. M., Wingen, M., Vermeeren, A., Vinckenbosch, F., Jongen, S., & Ramaekers, J. G. (2017). Driving performance of depressed patients who are untreated or receive long-term antidepressant (SSRI/SNRI) treatment. *Pharmacopsychiatry, 50*(05), 182-188.
Vermeeren, A. (2004). Residual Effects of Hypnotics. *CNS drugs, 18*(5), 297-328. doi:10.2165/00023210-200418050-00003
Verster, & Roth. (2013a). Blood drug concentrations of benzodiazepines correlate poorly with actual driving impairment. *Sleep medicine reviews, 17*(2), 153-159.
Verster, J. C., Peters, L. V., van de Loo, A. J. A. E., Bouwmeester, H., Tiplady, B., Alford, C., & Roth, T. (2016a). Next-morning effects of hypnotic drugs on attention, psychomotor performance, and memory functioning: implications for traffic safety. *Journal of Sleep Research, 25*, 126-127.
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. doi:10.2147/IJGM.S19639
Verster, J. C., & Roth, T. (2013b). Vigilance decrement during the on-the-road driving tests: The importance of time-on-task in psychopharmacological research. *Accident Analysis & Prevention, 58*, 244-248.

Verster, J. C., & Roth, T. (2014). Insomnia and Driving Ability. *Sleep, 37*(9), 1411-1412. doi:10.5665/sleep.3984

Verster, J. C., van de Loo, A. J. A. E., Vermeeren, A., van der Sluiszen, N. N. J. J. M., Brookhuis, K. A., Veldstra, J. L., . . . Ramaekers, J. G. (2016b). *Beïnvloeding van de rijvaardigheid bij langdurig gebruik van ICADTS-categorie III geneesmiddelen.* https://www.rijksoverheid.nl/documenten/rapporten/2016/12/01/beinvloeding-van-de-rijvaardigheid-bij-langdurig-gebruik-van-icadts-categorie-iii-geneesmiddelen.

Verster, J. C., Veldhuijzen, D. S., & Volkerts, R. (2004). Residual effects of sleep medication on driving ability. *Sleep medicine reviews, 8*(4), 309-325.

Verster, J. C., Veldhuijzen, D. S., & Volkerts, R. (2005). Is it safe to drive a car when treated with anxiolytics? Evidence from onthe-road driving studies during normal traffic. *Current Psychiatry Reviews, 1*(2), 215-225.

Vinkers, C. H., & Olivier, B. (2012). Mechanisms underlying tolerance after long-term benzodiazepine use: a future for subtype-selective GABAA receptor modulators? *Advances in pharmacological sciences, 2012*.

Wakamatsu, A., Aoki, K., Sakiyama, Y., Ohnishi, T., & Sugita, M. (2013). Predicting pharmacokinetic stability by multiple oral administration of atypical antipsychotics. *Innovations in clinical neuroscience, 10*(3), 23.

Wechsler, D. (1955). Manual for the Wechsler adult intelligence scale.

Wingen, M., Ramaekers, J. G., & Schmitt, J. A. (2006). Driving impairment in depressed patients receiving long-term antidepressant treatment. *Psychopharmacology, 188*(1), 84-91.

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., . . . Sayeeda, Z. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research, 46*(D1), D1074-D1082.

Wright, C. E. (1995). Clinical pharmacokinetics of alprazolam extended release: a summary. *Current therapeutic research, 56*(9), 947-956.

Yu, Y., Jiang, C., Xu, H., Yang, Q., Li, J., Xu, Y., . . . Lv, F. (2018). Impaired cognitive control of emotional conflict in trait anxiety: a preliminary study based on clinical and non-clinical individuals. *Frontiers in psychiatry, 9*, 120.

Subj	BZRA	Body weight (kg)	Dose (mg)	Last dose (h) [a]	Dosing interval (h)	Therapeutic Treshold (ng/ml)	Css(t) (ng/ml)[b]	C̄ss (ng/ml)[c]	Css(t) DIA (ng/ml)[d]	C̄ss DIA (ng/ml)[e]
1	Temazepam	60	20	14	24	20	181,4	184,8	91,6	91,4
2	Temazepam	78	20	12,5	24	20	149,7	142,2	75,6	70,3
3	Oxazepam	78	10		24	200	27,8	57,1	14,3	28,9
4	Lorazepam	100	1	15	24	80	5,6	6,6	59,3	68,7
5	Lorazepam	78	1	11,75	24	80	8,4	8,4	89,4	88,1
6	Oxazepam	125	30	12,5	24	200	101,2	106,9	52,3	54,0
7	Lorazepam	90	1	2,5	8	80	24,0	21,9	256,5	229,0
8	Zolpidem	70	10	11	24	80	0,2	12,0	0,1	8,0
9	Oxazepam	105	10	1,75	24	200	65,9	42,4	34,0	21,4
10	Temazepam	75	10	10,5	8	20	181,9	221,8	91,9	109,7
11	Zopiclon	65	7,5	12	24	10	10,9	16,9	8,7	13,2
12	Oxazepam	70	5	1	24	200	110,8 **[f]**	95,4 **[f]**	57,2 **[g]**	48,2 **[g]**
	Oxazepam		10	10,5	24	200				
13	Temazepam	125	20	10,5	24	20	102,1	88,7	51,6	43,9
14	Diazepam	84	5	11,5	24	100	112,3	111,0	114,7	111,0
15	Lormetazepam	95	3	11,5	24	2	4,9	4,6	47,1	42,9
16	Nitrazepam	85	2,5	11	24	30	23,1	22,7	22,2	21,3
17	Oxazepam	105	10	13,75	24	200	37,4	42,4	19,3	21,4
18	Oxazepam	74	7,5	2,5	24	200	72,1	45,1	37,7	23,1
19	Zolpidem	60	15	15	24	80	0,0	21,0	0,0	14,1
20	Temazepam	122	20	12,25	24	20	96,8	90,9	48,9	45,0
21	Oxazepam	94	10	2,75	12	200	115,0	94,8	59,4	47,9
22	Oxazepam	81	10	11,5	24	200	55,2	55,0	28,5	27,8
23	Lorazepam	68	0,5	11	24	80	5,0	4,8	53,2	50,5
24	Zopiclon	78	3,75	12,75	24	10	3,3	7,0	2,7	5,5
25	Oxazepam	70	10	0,25	4,8	200	307,8	318,1	158,9	160,8
26	Zopiclon	68	7,5	10	24	10	13,7	16,1	10,9	12,6
27	Zolpidem	81	10	12,5	24	80	0,1	10,3	0,0	6,9
28	Zopiclon	75,5	7,5	8,5	24	10	14,8	14,5	11,9	11,4
29	Brotizolam	76	0,25	13,25	24	1	0,4	0,9	21,9	46,4
30	Zolpidem	76	10	11,5	24	80	0,1	11,0	0,1	7,4
31	Alprazolam	58	0,5	5	12	5	10,4	9,6	222,2	201,0
32	Midazolam	68	3,75	11	24	40	1,5	3,5	3,7	8,7
33	Oxazepam	81	10	1,5	6	200	236,5	220,0	122,1	111,2
34	Temazepam	74	10	11,75	24	20	81,6	74,9	41,3	37,1
35	Midazolam	57	15	11	24	40	7,0	16,6	17,9	41,6
36	Lorazepam	60	1	13	24	80	10,2	11,0	109,2	114,5
37	Zopiclon	94	7,5	10,5	24	10	6,8	11,7	5,4	9,1
38	Temazepam	71	20	10	24	20	141,8	116,9	71,7	57,8
39	Zolpidem	55	10	12	24	80	0,1	15,2	0,1	10,2
40	Lorazepam	71	1,25	11,5	24	80	11,7	11,6	124,3	121,0
41	Lorazepam	140	1	1,5	8	80	15,6	14,1	166,9	147,2
42	Alprazolam	82	1	1,25	12	5	19,0	13,6	405,1	284,3
43	Lormetazepam	78	2	9,5	24	2	4,3	3,7	24,4	20,9
44	Zolpidem	107	10	11,5	24	80	0,1	7,8	21,6 **[g]**	26,3 **[g]**
	Oxazepam	107	10	11,5	24	200	41,8	41,6		
45	Zopiclon	102	7,5	12	24	10	4,0	10,7	2413,6**[g]**	2335,6**[g]**
	Clorazepate	102	30	2	6	200	3427,3	3379,4		
46	Zolpidem	55	10	12,5	24	80	0,1	15,2	690,3 **[g]**	688,4 **[g]**
	Clonazepam	55	2	12,5	24	22	32,4	32,5		
47	Zolpidem	89	10	11,75	24	80	0,1	9,4	195,9 **[g]**	199,3 **[g]**
	Lorazepam	89	2,5	11,75	24	80	18,4	18,5		
48	Oxazepam	110	10	1,75	24	200	62,9	40,5	32,4	20,5
49	Alprazolam	71	0,25	1,5	12	5	5,5	3,9	116,9	82,1
50	Oxazepam	110	10	11,75	24	200	40,1	40,5	36,8 **[g]**	35,3 **[g]**
	Lormetazepam	110	2	11,75	24	2	2,8	2,7		

Appendix 1. Table listing individual medication profiles; **[a]** Elapsed time since last use relative to start of test day; **[b]** estimated drug plasma concentration at steady state at the start of the test day; **[c]** estimated average drug plasma concentration at steady state; **[d]** estimated equivalent diazepam plasma concentration at steady state at the start of the test day; **[e]** estimated equivalent average diazepam plasma concentration at steady state; **[f]** Different doses of the same drug used at different times were calculated separately and the results were then added. Added estimate is listed; **[g]** sum of equivalent diazepam plasma concentrations.

Chapter 6: Validating Lane Drifts as a Predictive Measure of Drug or Sleepiness Induced Driving Impairment

Abstract

Standard Deviation of Lateral Position (SDLP) has been accepted as a reliable parameter for measuring driving impairment due to lowered vigilance caused by sleepiness or the use of sedating drugs. Recently, lane drifts were proposed as an additional outcome measure quantifying momentary lapses of attention. The purpose of this study was to validate lane drifts as outcome measure of driver impairment in a large data pool from 2 independent research centers. Data from 11 placebo controlled studies that assessed the impact of alcohol, hypnotics and sleep deprivation on actual driving performance were pooled. In total, 717 on-road tests performed by 315 drivers were subjected to an automated algorithm to detect occurrences of lane drifts. Lane drifts were defined as deviations > 100 cm from the mean (LD_{mlp}) and from the absolute lateral position (LD_{alp}) for 8 seconds. The number of LD_{mlp} was low and did not differ between treatments and baseline, i.e. 14 vs 3 events respectively. LD_{alp} were frequent and significantly higher during treatment relative to baseline, i.e. 1646 vs 470 events. The correlation between LD_{alp} and SDLP in the treatment conditions was very high(r_s= 0.77). The frequency of the occurrence of treatment induced lane drifts however depended on baseline SDLP of drivers whereas treatment induced changes in SDLP occurred independent of baseline SDLP. LD_{mlp} is not useful as an outcome measure of driver impairment due to its rare occurrence, even when treatment induced increments in SDLP are evident. Treatment effects on LD_{alp} and SDLP are closely related.

Vinckenbosch, F. R. J., Vermeeren, A., Verster, J. C., Ramaekers, J. G., & Vuurman, E. F. (2020). Validating lane drifts as a predictive measure of drug or sleepiness induced driving impairment. *Psychopharmacology*, *237*(3), 877-886. https://doi.org/10.1007/s00213-019-05424-8

Introduction

Sustained attention is a necessary requirement for the safe operation of a motor vehicle in traffic. It has been estimated that up to 15% of the traffic accidents on motor ways are associated with sleepiness (Maycock, 1996). CNS drugs such as hypnotics are known to produce sleepiness and may affect psychomotor and executive function (Jongen, Vuurman, Ramaekers, & Vermeeren, 2018). Consequently, epidemiological studies have repeatedly demonstrated a positive association between the use of sedative medications and accident risk (Barbone et al., 1998; Gustavsen et al., 2008; Movig et al., 2004; Orriols et al., 2011).

For over 30 years, the ”gold standard” for the assessment of the effects of drugs and sleepiness on driver vigilance in experimental, placebo controlled studies has been a standardized on-the-road driving test (Ramaekers, 2017) that was developed in The Netherlands. This naturalistic test requires participants to complete a 100 km test drive on a primary highway while accompanied by a licensed driving instructor who has access to dual controls. The participants are instructed to maintain a steady lateral position in the middle of the right traffic lane at a velocity of 95km/h. During the drive, a camera mounted on top of the vehicle continuously monitors the vehicle’s lateral position relative to the traffic lane demarcation to the left of it.

The standard deviation of lateral position (SDLP), i.e. the weaving of the car, is the main outcome measure of the standardized driving test. It is considered to be a quantification of lane weaving, a measure of vehicle control. It has been repeatedly demonstrated that mean SDLP increases with approximately 2.5cm when driving under a blood alcohol concentration (BAC) of 0.5g/L, relative to driving under placebo. A BAC of 0.5g/L is the legal limit for driving under the influence of alcohol in most European countries, because epidemiological studies have demonstrated that crash risk increases at concentrations exceeding this threshold (Borkenstein, Crowther, & Shumate, 1974). Consequently, an increase in SDLP of 2.5cm or higher represents a clinically relevant change in the on-road driving test when quantifying medicinal drug effects on SDLP. On-road driving studies have demonstrated that the use of sedative medications, such as benzodiazepines and

antidepressants, and sleep deprivation can produce increments in SDLP that are equal to or greater than this clinical threshold (Jongen, Perrier, Vuurman, Ramaekers, & Vermeeren, 2015; Jongen et al., 2018; Ramaekers, Muntjewerff, Van Veggel, Uiterwijk, & O'Hanlon, 1998; Veldhuijzen et al., 2006). The implication is that these challenges can increase crash risk, similar to alcohol. Drug and alcohol induced changes in SDLP observed in the on-the-road driving tests are strongly correlated to drug and alcohol induced crash-risk as assessed in epidemiological studies(Ramaekers, 2017). The implication is that SDLP as measured in the on-the-road driving test is not merely a measure of driver impairment but also predicts crash risk.

However, despite the agreement that the SDLP is a sensitive measure for driver vigilance, it has been argued that the association between drug and sleepiness induced increments in SDLP and crash risk is indirect (Hartman et al., 2015; Lococo & Staplin, 2006; Verster, Bervoets, de Klerk, & Roth, 2014) and that other factors may play a more important role in the occurrence of an accident such as brief moments of inattention, micro-sleeps and distraction (Verster et al., 2014; Verster, Mooren, Bervoets, & Roth, 2018; Verster & Roth, 2014a).

Two additional, potentially relevant measures for predicting crash risk have been proposed, i.e. lane excursions and lane drifts, which are both derived from the same source parameter as SDLP, i.e. lateral position. Lane excursions were quickly discarded as a sensitive measure of crash risk because they occur infrequently and are much less sensitive than SDLP for demonstrating driver impairment (Verster & Roth, 2014a). Lane drifts, however, have been proposed as a valuable measure of lapses of attention that may occur during prolonged driving (Verster et al., 2014). The authors who proposed this measure reanalyzed driving data from two double-blind, placebo-controlled on-the-road driving studies that examined the residual effects of hypnotic drugs on lateral position. The reanalysis showed that both SDLP and lapses of attention were significantly higher following administration of hypnotics as compared to placebo. A lapse of attention in this context was conceptualized by the authors as ''*a short period of inattention during which the driver experiences reduced alertness and does not focus on the task, or actually stops performing the task,*

resulting in driving impairment''. It was argued that a momentary reduction or loss of attention would lead to a relatively large deviation in lateral position. Hence, a lapse of attention during the on-the-road driving test is operationalized as ''*a deviation from mean lateral position of more than 100cm for four or more seconds''(Verster et al., 2014)*. Post hoc, the lapse duration criterion was increased to 8 seconds since it was found that only 4.6% of the observed deviations of >100cm had a duration of 4 to 8 seconds (Verster et al., 2014). In a later publication, an alternative operationalization was applied on the same dataset that defined a lapse "as a continuous change in lateral position of greater than 100 cm, lasting for at least 8 seconds" (Verster et al., 2018). Both operationalizations were reported to be able to detect a significant increase in the number of lane drifts during on-road driving in the morning after nighttime administration of hypnotics such as zolpidem, zopiclone and ramelteon (Verster et al., 2014; Verster et al., 2018). It was therefore concluded that lane drifts are a useful outcome measure of driving impairment during on-road driving that is conceptually distinct from driver impairments assessed with SDLP.

So far, lane drifts as a measure of driving impairment have been piloted in two data sets from on-road driving studies (Mets et al., 2011; Verster et al., 2014; Verster et al., 2018) as described above. In addition, the outcome has been applied in a study on the effects of methylphenidate on the driving performance of ADHD patients where it was found that treatment significantly reduces the number of lane drifts (Verster & Roth, 2014b). Another study into the effects of the orexin antagonist Lemborexant in healthy volunteers did not find any lane drifts at all, which is in contrast to the aforementioned findings which state that lane drifts were also apparent after placebo treatment in healthy volunteers (Vermeeren et al., 2018; Verster et al., 2014). A large scale validation of the measure's sensitivity to drug induced impairment and how it discriminates from SDLP is still lacking. The purpose of this study was to further investigate and validate lane drifts as outcome measure of drug and sleepiness induced driver impairment in a large data pool from 2 independent research centers that included on-road driving data from 11 placebo controlled studies that assessed the impact of alcohol, hypnotics and sleep deprivation on actual driving performance.

Methods

In order to validate proposed operationalizations of lane drifts (Verster et al., 2014; Verster et al., 2018), data was pooled from 11 randomized, placebo-controlled, cross-over studies that employed the on-the-road driving test and were conducted at 2 independent research centers. These studies were originally designed to assess the effects of alcohol(Kuypers, Samyn, & Ramaekers, 2006; van der Sluiszen et al., 2016), hypnotics (Jongen et al., 2018; Leufkens, Lund, & Vermeeren, 2009; Mets et al., 2011; Vermeeren et al., 2018; Vermeeren et al., 2015; Vermeeren et al., 2016; Vermeeren et al., 2014; Verster et al., 2002), and sleep deprivation (Jongen et al., 2015) and were selected for the current investigation because the administered treatments were found to significantly affect the SDLP. The pooled data set included 315 healthy volunteers of both sexes who drove after alcohol (N=49), zopiclone (N=194), zolpidem (N=72), oxazepam (N=43), diazepam (N=21) and sleep deprivation (N=23) and after placebo (N=315). For an overview of the studies included, samples sizes per study, references and research center, see Table 1. For studies where the effect of a substance was assessed at two successive time points, only the first measurement was included. It is noted that for two studies (Mets et al., 2011; Verster et al., 2002) the number of participants differs from the original publications because the data of 11 driving tests could not be recovered.

For every test drive, SDLP was calculated as the standard deviation relative to the mean lateral position over the entire 100 km drive, as opposed to the average of the SDLP across 20 successive segments of 5km, which is traditionally employed (O'Hanlon, 1984; O'Hanlon, Haak, Blaauw, & Riemersma, 1982; Ramaekers, Uiterwijk, & O'hanlon, 1992). The segmented approach corrects for drifts in the mean lane position over time that frequently occur and may lead to inflated SDLP values. Still, it was opted to calculate SDLP from the mean lateral position over the entire test drive to eliminate differences in the calculation of SDLP between the 2 centers. The Maastricht center traditionally uses the segmented approach for calculating SDLP whereas the Utrecht center treats driving data over 100km as a single segment (Verster & Roth, 2011). Consequently, mean SDLP

values from studies conducted in the Maastricht research center will be higher as those reported in their original publications.

Study	Treatment	tstart driving test	research center
Kuypers et al. (2008)	Alcohol 0,5g/L (N=18)	+2h	Maastricht University
	Placebo (N=18)	+2h	
van der Sluiszen et al. (2016)	Alcohol 0,5g/L (N=31)	+1,5h	Maastricht University
	Placebo (N=31)	+1,5h	
Vermeeren et al. (2014)	Zopiclone 7,5mg (N=40)	+9h	Maastricht University
	Placebo (N=40)	+9h	
Vermeeren et al. (2015)	Zopiclone 7,5mg (N=28)	+9h	Maastricht University
	Placebo (N=28)	+9h	
Vermeeren et al. (2016)	Zopiclone 7,5mg (N=24)	+9h	Maastricht University
	Placebo (N=24)	+9h	
Vermeeren et al. (2018)	Zopiclone 7,5mg (N=48)	+9h	Maastricht University
	Placebo (N=48)	+9h	
Mets et al. (2011)	Zopiclone 7,5mg (N=29)	+8,5h/10h	Utrecht University
	Placebo (N=29)	+8,5h/10h	
Leufkens et al. (2009)	Zopiclone 7,5mg (N=25)	+10h	Maastricht University
	Zolpidem 10mg (N=24)	+5h	
	Placebo (N=25)	+5h/+10h	
Verster et al. (2002)	Zolpidem 10mg (N=23)	+4h	Utrecht University
	Zolpidem 20mg (N=25)	+4h	
	Placebo (N=27)	+4h	
Jongen et al. (2018)	Diazepam 10 mg (N=21)	+4h	Maastricht University
	Oxazepam 30mg (N=22)	+4h	
	Oxazepam 10mg (N=21)	+4h	
	Placebo (N=22)	+4h	
Jongen et al. (2015)	Sleep deprivation (N=23)	-	Maastricht University
	Normal sleep (N=23)		

Table 1. Summary of included data. The time of the start of the driving test (tstart driving test) is relative to the time of drug administration.

In previous publications (Verster et al., 2014; Verster et al., 2018) lane drifts were evaluated and classified by visual inspection of the individual driving data. In order to control for interrater differences and to objectify the scoring procedure an automated algorithm was created to detect occurrences of lane drifts in all individual drives. Lane drifts were defined as deviations > 100 cm from the mean (LD_{mlp}) and from the absolute lateral position (LD_{alp}) for 8 seconds as previously suggested (Verster et al., 2014; Verster et al., 2018). The former essentially defines a lane drift as the event where a driver crosses the lane border for more than 8 secs, whereas the latter qualifies a lane drift as any lateral displacement >100cm that occurs within a time window of 8 secs. The automated algorithm was programmed to calculate mean SDLP and to count the number of lane drift according to the 2 definitions given above.

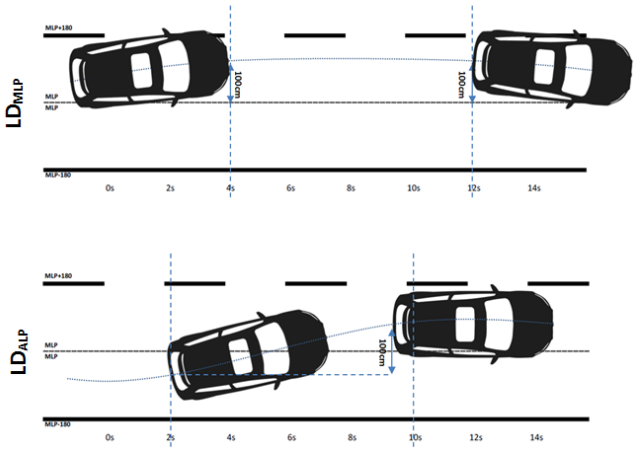

Figure 1. Illustration depicting a lane drift relative to the mean lateral position (LDmlp) and lane drifts relative to the absolute lateral position (LDalp).

Statistical Analysis

For the statistical analysis, participant data was grc o assure that the method of calculating the SDLP, i.e. as the standard deviation of the entire test as opposed to the average of the standard deviations per 5km segment, did not affect the relative drug effects reported in the original investigations, the Pearson correlation was determined between the two calculation methods for the largest group in the sample, i.e. participants treated with zopiclone 7.5mg and placebo (N=194), but not for all available test drives since the aggregation and parallel analysis of all datasets is time costly and was not considered necessary to make this minor point.

Next, in order to investigate whether lane drifts are indeed sensitive measures for observing driving impairment, the statistical significance of the difference in SDLP values between each treatment and the respective baseline condition was tested with a paired samples t-test. For the statistical testing of the difference in number of LD_{mlp} and LD_{alp} between the treatment and baseline conditions, the non-parametric Wilcoxon signed rank test was selected because of the severe skewness of the data which could not be resolved by logarithmic transformations. In addition, the correlation between the SDLP and number of LD_{alp} in the treatment conditions, and the correlation between the increase in SDLP (ΔSDLP) and the number of lane drifts was determined. It was hypothesized that if lane drifts are a true measure of driving impairment rather than a mere transformation of the SDLP, the relationship between the number of LD and the ΔSDLP should be stronger and more consistent than the relationship between the LD and the absolute SDLP values during treatment, considering that it is the increase in SDLP which conveys information about driving impairment rather than the absolute SDLP value. If the correlation between the lane drifts and absolute SDLP values are stronger, this suggests that lane drifts are a mere transformation of SDLP rather than a measure of driving impairment. Because the scatterplots suggested a quadratic pattern for both relationships, Spearman's $\rho(r_s)$ was chosen over Pearson's *r*. Lastly, in order to further investigate the independence of lane drifts from the SDLP, the Spearman correlation was calculated between the baseline values of the SDLP of individuals and their ΔSDLP during treatments, as well as the correlation between the baseline SDLP values of individuals and their ΔLD_{alp} during treatment. It is to be expected that neither the ΔSDLP nor the ΔLD_{alp} correlate significantly with the baseline SDLP since the variation in the baseline SDLP is not assumed to reflect driving performance or driver fitness. For all statistical tests, the significance level was set at α=.05. All statistical analyses were conducted using the statistical package for the social sciences (*SPSS) edition 24* offered by IBM.

Results

The Pearson correlation between SDLP values calculated from the mean lateral position over the entire ride and across segments was highly significant for the 194 zopiclone 7.5mg drives ($r[163]= .84$, $p < .001$), as well as for the 194 corresponding baseline drives ($r[163]= .89$, $p < .001$). For all treatment conditions, a significant increase in SDLP was observed compared to baseline. Overall, only 14 LD_{mlp} events were detected during treatment conditions and only 3 LD_{mlp} events occurred during baseline. LD_{mlp} did not significantly differ between treatments and baseline. LD_{alp} events however occurred very frequently. A total of 1646 LD_{alp} events were established in the treatment conditions vs. 470 events during the baseline conditions. As can be seen in figure 2, the distribution was positively skewed, with the majority of test drives in the treatment conditions (54.5%) including no or a single LD_{alp} event. About 7.5% of the sample accounted for approximately 50% of all detected LD_{alp}. Overall, the number of LD_{alp} was significantly higher during most treatments as compared to baseline, with the exception of the oxazepam 10mg and diazepam 10mg conditions.

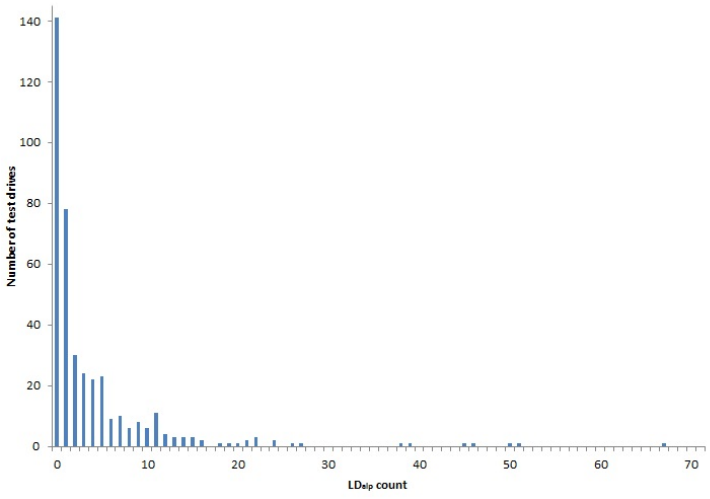

Figure 2. Histogram of the number of lane drifts relative to the absolute lateral position (LDalp) during the treatment conditions (N=402).

The correlation analyses demonstrated a consistent and highly significant correlation between SDLP and the number of LD_{alp} after treatment ($rs(400)= .77$, $p< .001$). Also, the correlation between the ΔSDLP and the number of LD_{alp} during the treatment conditions was significant ($rs(400)=$

.50, p< .001). Significant correlations between SDLP and number of $LD_{alp,}$ and between ΔSDLP and LD_{alp} across treatment conditions are shown in Figure 3.

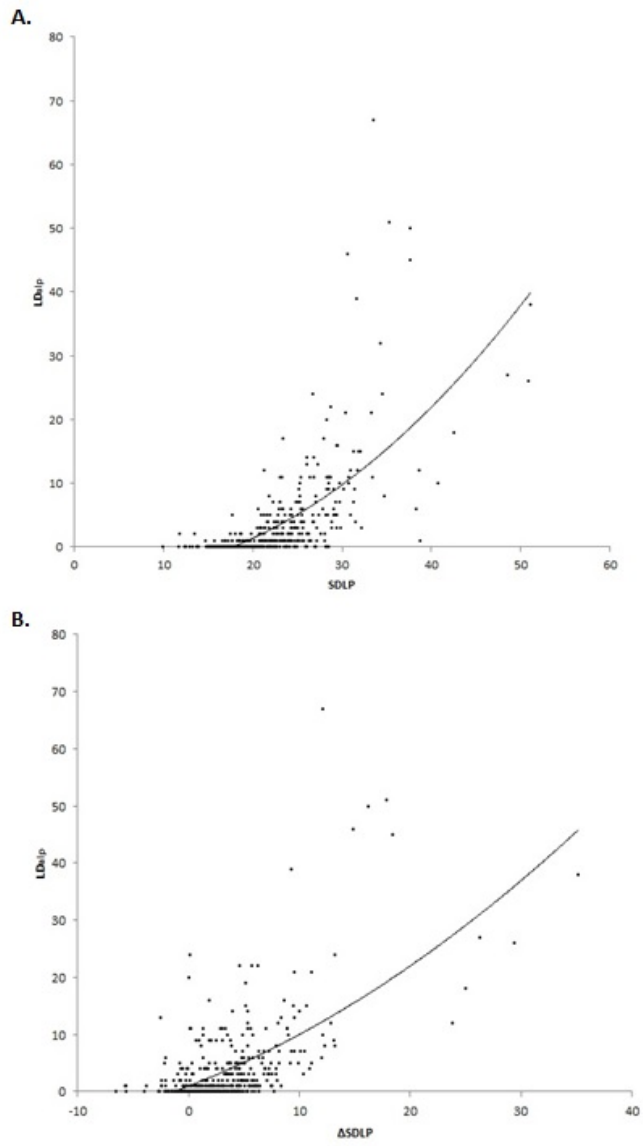

Figure 3. Scatterplots depicting the relationships between [A]the SDLP values and the number of lane drifts relative to the absolute lateral position(LDalp), and [B] the change in SDLP between placebo and treatment and the number of LDalp.

For each treatment separately, correlations between SDLP and number of LD_{alp} in the treatment conditions was found to be significant as well as the correlation between the ΔSDLP and LD_{alp} with the exception of the alcohol condition. The correlation analysis of the relationship between the baseline SDLP values on the one hand and ΔSDLP and ΔLD_{alp} on the other only yielded a significant Spearman correlation between the baseline SDLP and the ΔLD_{alp} in the zopiclone 7.5mg group ($r_s(192)= .20$, $p= .004$), but no significant correlation between the baseline SDLP values and the ΔSDLP in this same treatment group ($r_s(192)= -.11$, $p= .14$). All other correlations were found to be insignificant. Table 2 provides a summary of the results.

Treatment	N	Mean SDLP (SD)	Mean ΔSDLP (SE) [a]	Lane drifts relative to MLP	paired difference: + : - [$_1$]	W+ : W- [$_2$] [b]	Lane drifts relative to ALP	paired difference: + : - [$_1$]	W+ : W- [$_2$][b]	rs(SDLP-LD_{ALP}) [c]	rs(ΔSDLP-LD_{ALP}) [c]	rs($SDLP_{pla}$-ΔSDLP) [c]	rs($SDLP_{pla}$-ΔLD_{ALP}) [c]
Alcohol	49	22.97 (5.19)	**+2.65 (0.42)*** **	5	1 : 0	5 : 0	197	25 : 10	**455 : 175***	**.826*** **	**.271**	-.019	-.009
Alcohol Placebo		20.32 (4.33)		0			137						
Zopiclone	194	21.92 (3.82)	**+2.60 (0.19)*** **	1	1 : 0	1 : 0	413	94 : 20	**5424.5 : 1130.5*** **	**.707*** **	**.422*** **	-.106	**.204****
Zopiclone Placebo		19.32 (3.82)		0			178						
Zolpidem 10mg	47	23.10 (5.52)	**+3.78 (0.68)*** **	5	3 : 2	10 : 5	305	25 : 9	**504.4 : 90.5*** **	**.835*** **	**.585*** **	-.120	.145
Zolpidem Placebo		19.37 (3.59)		3			85						
Zolpidem 20mg	25	27.98 (11.23)	**+9.86 (2.12)*** **	3	2 : 1	3.5 : 2.5	380	17 : 5	**227.5 : 25.5****	**.849*** **	**.788*** **	.157	.089
Zolpidem Placebo		18.12 (3.99)		2			47						
Oxazepam 10mg	21	21.02 (4.57)	**+1.62 (0.71)***	0	-	-	47	10 : 5	82.5 : 37.5	**.769*** **	.488*	-.238	-.084
Oxazepam Placebo		19.4 (4.17)		0			30						
Oxazepam 30mg	22	27.01 (5.94)	**+7.74 (1.17)*** **	0	-	-	138	20 : 2	**222.5 : 30.5****	**.682*** **	**.611****	-.281	-.137
Oxazepam Placebo		19.27 (4.11)		0			30						
Diazepam 10mg	21	22.4 (1.01)	**+3.28 (0.56)*** **	0	-	-	75	11 : 3	82.5 : 22.5	**.729*** **	**..512***	-.117	.213
Diazepam placebo		19.12 (0.91)		0			30						
Sleep deprivation	23	20.62 (3.87)	**+3.36 (0.63)*** **	0	-	-	91	12 : 3	**114 : 6****	**.863*** **	**.543****	-.154	.303

Table2. Summary of the results. [$_1$] Ratio of the number of positive and negative paired differences (treatment - placebo); [$_2$] ratio of sum of positive (W+) and sum of negative (W-) differences;[a] Paired samples t-test; [b] Wilcoxon signed rank test; [c] Spearman correlation; * p < .05; ** p < .01; *** p < .001

Discussion

The primary goal of the current investigation was to validate lane drifts as an outcome measure of drug and sleepiness induced driver impairment as proposed by Verster et al.(Verster et al., 2014; Verster et al., 2018). An automated algorithm determined the number of lane drifts relative to the mean lateral position with a duration of at least 8 seconds (LD_{mlp}) and the number of lane drifts relative to the absolute lateral position within a time window of 8 seconds (LD_{alp}) in a large data pool. This data pool contained data of 315 test drives after placebo administration or no treatment, and 402 test drives after administration of a sedative substance or sleep deprivation. Over all 717 test drives, only 19 LD_{mlp} were detected. The number of LD_{mlp} was not found to be significantly more prevalent in the treatment conditions. A total of 2116 LD_{alp} were identified in all test drives. It was also found that the number of LD_{alp} was significantly higher in the treatment conditions with the exception of the oxazepam 10mg and diazepam 10mg conditions.

It appears that LD_{mlp} are rare events. As a result, the outcome measure is unable to significantly discriminate between any treatment and respective baseline condition, despite the evident increase of the SDLP after all treatments. It is therefore clear that ΔLD_{mlp} as an outcome measure has an inferior sensitivity, if any, to driving impairment compared to the ΔSDLP. In contrast to LD_{mlp}, multiple LD_{alp} were detected in every condition. The events appeared relatively frequently. However, the majority of drivers exhibited no or only one LD_{alp} during treatment. The distribution of LD_{alp} was highly skewed with less than ten percent of the participants accounting for approximately half of all detected events. The number of LD_{alp} was nearly always significantly higher in the treatment than in the baseline conditions, with the exception of the oxazepam 10mg and diazepam 10mg conditions. The absence of an effect could be attributed to the fact that in these conditions two drives were prematurely terminated, resulting in less opportunity for lane drifts to occur. In contrast the increase in SDLP was found to be significant in all treatment conditions which suggests that the ΔSDLP is more sensitive for the detection of driving impairment, even if the duration of the test is shortened. An inspection of the correlation coefficients in table 2 demonstrates that there is a

close positive relationship between the SDLP and the number of LD_{alp} in the treatment conditions, as well as between the ΔSDLP and the LD_{alp}. The former relationship was overall found to be stronger and more consistent across different treatments. The close relationship between the outcome measures is not surprising given that they are both derived from the lateral position of the vehicle.

As mentioned, the relationship between the absolute SDLP and LD_{alp} was overall stronger than the relationship between the ΔSDLP and LD_{alp}. However, if LD_{alp} is a true measure of driving impairment, it would be expected that a closer relationship should exist between the ΔSDLP and LD_{alp}. The absolute SDLP value is known to differ considerably between healthy individuals treated with placebo, while the test-retest reliability of the SDLP has been found to be high(Verster & Roth, 2011). Hence, the SDLP can be considered as a driving characteristic that varies considerably between healthy individuals, but not within, and therefore conveys little information about driving impairment per se. For this reason, it is the increase in SDLP (ΔSDLP) relative to placebo or baseline performance that is used in on-road driving studies to assess driving impairment. The close relationship between the absolute SDLP and $LD_{alp,}$ while the relationship between the ΔSDLP and the LD_{alp} was found to be less strong and less consistent, suggests that LD_{alp} is simply a transformation of the SDLP rather than an independent measure of driving impairment. The absence of a correlation between baseline SDLP and treatment induced ΔSDLP, and the presence of a significant correlation between baseline SDLP and a treatment (zopiclone) induced ΔLD_{alp} supports this notion. It demonstrates that the ΔSDLP is independent from the baseline SDLP, while detection of LD_{alp} is biased towards participants with higher baseline SDLP values. However, for LD_{alp} to be a true measure of driving impairment, the increase in LD_{alp} should occur independent from the baseline SDLP value. The observation that this is not the case is problematic.

Consideration of the statistical nature of the SDLP can explain why the relationship between the absolute SDLP and the LD_{alp} count is stronger than the relation between the ΔSDLP and ΔLD_{alp}, as well as provide an explanation for the modest correlation between the baseline SDLP and the increase in LD_{alp}. The standard deviation of the lateral position is a parameter of the width of the

confidence interval of the lateral position. It can be calculated that the 99% confidence interval of the lateral position is about 100cm wide when a participant drives with a mean SDLP of 19.45 cm. As a result, participants with a SDLP under 19.45 cm would likely exhibit no LD_{alp}, i.e. never produce a lane drift that spans more than 100 cm, despite the possibility of experiencing significant road tracking impairment as indicated by the SDLP. Therefore, the relationship between the absolute SDLP and LD_{alp} count is stronger than the relation between ΔSDLP and LD_{alp}. This also means that participants with a low baseline SDLP of e.g. 16cm should show greater impairment in road tracking ability as quantified by the ΔSDLP during the treatment condition than a participant with a high baseline SDLP of e.g. 19cm in order to exceed the threshold of 19.45cm. Taken together, the high correlation between the absolute SDLP and absolute LD_{alp} count, and the statistical bias towards participants with higher inherent SDLP values suggest that LD_{alp} is a linear transformation of the SDLP that conceivably adds no new information and is less sensitive to impairment than the ΔSDLP.

Besides the abovementioned problems with the LD_{mlp} and LD_{alp} as parameters of driving impairment, a more fundamental issue remains. Verster et al.(Verster et al., 2014; Verster et al., 2018; Verster & Roth, 2014a) proposed lane drifts as a measure of momentary lapses of attention which were defined as "*short periods of inattention during which the driver experiences reduced alertness and does not focus on the task, or actually stops performing the task, resulting in driving impairment*". However, it is uncertain whether a momentary lapse of attention leads to a significant change in lateral position. Of course, it appears likely that a significant lateral displacement would occur if the participant stops performing the task altogether. However, an event like this can arguably not be conceived as a lapse of attention, rather than falling asleep or losing consciousness.

A lapse of attention is usually conceptualized as moment during which internal task irrelevant information is being processed at the cost of the processing of incoming external information, and are often referred to as task unrelated thought (TUT) or mind wandering (Smallwood et al., 2004). A frequently employed lab task for the assessment of attentional lapses in this sense is the sustained attention to response task (SART). During this task, participants are

instructed to respond to stimuli which are presented at a rapid pace, and to withhold their response whenever a relatively infrequently presented target stimulus appears. A lapse is defined as a commission error, i.e. responding to the target stimulus which requires inhibition of a response. It has been argued that commission errors during the SART are the result of poor top-down motor control (motor decoupling), rather than inattention to external stimuli (perceptual decoupling)(Head & Helton, 2013), or both(Seli, 2016). Whatever the case, it is clear that higher level functioning can be impaired while lower level functioning, i.e. indiscriminant and automated responding to any stimulus, remains intact. Arguably, for experienced drivers, road tracking is also a highly automated skill which requires little focused attention. It is therefore reasonable to assume that a short period of inattention does not necessarily result in a measurable change in lateral position. Future research should adopt established physiological and behavioral measures of attention in order to assess if and when drivers experience attentional lapses and whether this is reflected in the lateral position of the vehicle.

Conclusion

Lane drifts relative to the mean lateral position (≥ 100cm) which last for at least 8 seconds are not a useful outcome measure of drug and sleepiness induced driving impairment. Due to the rare occurrence it is unable to demonstrate driving impairment in various treatment conditions despite SDLP showing significant treatment induced increases. Lane drifts relative to the absolute lateral position (≥ 100cm) which occur within a window of 8 seconds did occur frequently and were able to demonstrate treatment induced driving impairment in most conditions. However, the measure seems to be a simple transformation of the SDLP with inferior sensitivity to treatment induced driver impairment. Also, the detection rate is biased in the direction of drivers with higher inherent SDLP values. It is therefore concluded that this outcome has little or no additional value to the SDLP as an outcome measure of the on-the-road driving test.

References

Barbone, F., McMahon, A., Davey, P., Morris, A., Reid, I., McDevitt, D., & MacDonald, T. (1998). Association of road-traffic accidents with benzodiazepine use. *The Lancet, 352*(9137), 1331-1336.

Borkenstein, R. F., Crowther, R. F., & Shumate, R. (1974). The role of the drinking driver in traffic accidents (The Grand Rapids Study). *Blutalkohol, 11*(Suppl.), 1-131.

Gustavsen, I., Bramness, J. G., Skurtveit, S., Engeland, A., Neutel, I., & Mørland, J. (2008). Road traffic accident risk related to prescriptions of the hypnotics zopiclone, zolpidem, flunitrazepam and nitrazepam. *Sleep medicine, 9*(8), 818-822.

Hartman, R. L., Brown, T. L., Milavetz, G., Spurgin, A., Pierce, R. S., Gorelick, D. A., . . . Huestis, M. A. (2015). Cannabis effects on driving lateral control with and without alcohol. *Drug and alcohol dependence, 154*, 25-37.

Head, J., & Helton, W. S. (2013). Perceptual decoupling or motor decoupling? *Consciousness and Cognition, 22*(3), 913-919.

Jongen, S., Perrier, J., Vuurman, E. F., Ramaekers, J. G., & Vermeeren, A. (2015). Sensitivity and validity of psychometric tests for assessing driving impairment: effects of sleep deprivation. *PLoS one, 10*(2), e0117045.

Jongen, S., Vuurman, E., Ramaekers, J., & Vermeeren, A. (2018). Comparing the effects of oxazepam and diazepam in actual highway driving and neurocognitive test performance: a validation study. *Psychopharmacology, 235*(4), 1283-1294.

Kuypers, K. P., Samyn, N., & Ramaekers, J. G. (2006). MDMA and alcohol effects, combined and alone, on objective and subjective measures of actual driving performance and psychomotor function. *Psychopharmacology, 187*(4), 467-475.

Leufkens, T. R., Lund, J. S., & Vermeeren, A. (2009). Highway driving performance and cognitive functioning the morning after bedtime and middle-of-the-night use of gaboxadol, zopiclone and zolpidem. *Journal of sleep research, 18*(4), 387-396.

Lococo, K. H., & Staplin, L. (2006). *Literature review of polypharmacy and older drivers: identifying strategies to study drug usage and driving functioning among older drivers*. Retrieved from

Maycock, G. (1996). Sleepiness and driving: the experience of UK car drivers. *Journal of sleep research, 5*(4), 229-231.

Mets, M. A., De Vries, J. M., de Senerpont Domis, L. M., Volkerts, E. R., Olivier, B., & Verster, J. C. (2011). Next-day effects of ramelteon (8 mg), zopiclone (7.5 mg), and placebo on highway driving performance, memory functioning, psychomotor performance, and mood in healthy adult subjects. *Sleep, 34*(10), 1327-1334.

Movig, K. L., Mathijssen, M., Nagel, P., Van Egmond, T., De Gier, J. J., Leufkens, H., & Egberts, A. C. (2004). Psychoactive substance use and the risk of motor vehicle accidents. *Accident Analysis & Prevention, 36*(4), 631-636.

O'Hanlon, J. (1984). Driving performance under the influence of drugs: rationale for, and application of, a new test. *British journal of clinical pharmacology, 18*(S1), 121S-129S.

O'Hanlon, J., Haak, T., Blaauw, G., & Riemersma, J. (1982). Diazepam impairs lateral position control in highway driving. *Science, 217*(4554), 79-81.

Orriols, L., Philip, P., Moore, N., Castot, A., Gadegbeku, B., Delorme, B., . . . Group, C. R. (2011). Benzodiazepine-like hypnotics and the associated risk of road traffic accidents. *Clinical Pharmacology & Therapeutics, 89*(4), 595-601.

Ramaekers, J. (2017). Drugs and driving research in medicinal drug development. *Trends in pharmacological sciences, 38*(4), 319-321.

Ramaekers, J., Muntjewerff, N., Van Veggel, L., Uiterwijk, M., & O'Hanlon, J. (1998). Effects of nocturnal doses of mirtazapine and mianserin on sleep and on daytime psychomotor and driving performance in young, healthy volunteers. *Human Psychopharmacology: Clinical and Experimental, 13*(S2), S87-S97.

Ramaekers, J. G., Uiterwijk, M., & O'hanlon, J. (1992). Effects of loratadine and cetirizine on actual driving and psychometric test performance, and EEG during driving. *European journal of clinical pharmacology, 42*(4), 363-369.

Seli, P. (2016). The attention-lapse and motor decoupling accounts of SART performance are not mutually exclusive. *Consciousness and Cognition, 41*, 189-198.

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition, 13*(4), 657-690.

van der Sluiszen, N., Vermeeren, A., Jongen, S., Theunissen, E., van Oers, A., Van Leeuwen, C., . . . Ramaekers, J. (2016). On-the-road driving performance after use of the antihistamines mequitazine and l-mequitazine, alone and with alcohol. *Psychopharmacology, 233*(18), 3461-3469.

Veldhuijzen, D. S., van Wijck, A. J., Verster, J. C., Kenemans, J. L., Kalkman, C. J., Olivier, B., & Volkerts, E. R. (2006). Acute and subchronic effects of amitriptyline 25mg on actual driving in chronic neuropathic pain patients. *Journal of psychopharmacology, 20*(6), 782-788.

Vermeeren, A., Jongen, S., Murphy, P., Moline, M., Filippov, G., Pinner, K., . . . Van Oers, A. (2018). On-the-road driving performance the morning after bedtime administration of lemborexant in healthy adult and elderly volunteers. *Sleep*.

Vermeeren, A., Sun, H., Vuurman, E. F., Jongen, S., Van Leeuwen, C. J., Van Oers, A. C., . . . Heirman, I. (2015). On-the-road driving performance the morning after bedtime use of suvorexant 20 and 40 mg: a study in non-elderly healthy volunteers. *Sleep, 38*(11), 1803-1813.

Vermeeren, A., Vets, E., Vuurman, E. F., Van Oers, A. C., Jongen, S., Laethem, T., . . . Li, X. (2016). On-the-road driving performance the morning after bedtime use of suvorexant 15 and 30 mg in healthy elderly. *Psychopharmacology, 233*(18), 3341-3351.

Vermeeren, A., Vuurman, E. F., Leufkens, T. R., Van Leeuwen, C. J., Van Oers, A. C., Laska, E., . . . Roth, T. (2014). Residual effects of low-dose sublingual zolpidem on highway driving performance the morning after middle-of-the-night use. *Sleep, 37*(3), 489-496.

Verster, J. C., Bervoets, A. C., de Klerk, S., & Roth, T. (2014). Lapses of attention as outcome measure of the on-the-road driving test. *Psychopharmacology, 231*(1), 283-292. doi:10.1007/s00213-013-3236-0

Verster, J. C., Mooren, L., Bervoets, A. C., & Roth, T. (2018). Highway driving safety the day after using sleep medication: the direction of lapses and excursions out-of-lane in drowsy drivers. *Journal of sleep research, 27*(3), e12622.

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. doi:10.2147/IJGM.S19639

Verster, J. C., & Roth, T. (2014a). Excursions out-of-lane versus standard deviation of lateral position as outcome measure of the on-the-road driving test. *Human Psychopharmacology: Clinical and Experimental, 29*(4), 322-329.

Verster, J. C., & Roth, T. (2014b). Methylphenidate significantly reduces lapses of attention during on-road highway driving in patients with ADHD. *Journal of clinical psychopharmacology, 34*(5), 633-636.

Verster, J. C., Volkerts, E. R., Schreuder, A. H., Eijken, E. J., van Heuckelum, J. H., Veldhuijzen, D. S., . . . Danjou, P. (2002). Residual effects of middle-of-the-night administration of zaleplon and zolpidem on driving ability, memory functions, and psychomotor performance. *Journal of clinical psychopharmacology, 22*(6), 576-583.

Chapter 7: Determining the statistical significance of change in road tracking performance of individual drivers in repeated measures designs

Abstract

Currently, no cutoffs indicating statistically significant change in road tracking performance during standardized on-the-road driving testing have been determined to qualify driving impairment in individual drivers. The current article applies classical test theory (CTT) and least squares linear regression (LSLR) as two methods for determining the statistical significance of change in road tracking performance at the individual level over repeated tests. Normative parameters for both methods were estimated based on a pooled dataset including 98 pairs of standardized on-the-road driving tests completed 7 days apart during a placebo period of the respective study. These estimates were used to determine confidence intervals of normative change (change under null hypothesis). The CTT method yielded the following limits for 68%, 80%, 90%, 95%, and 99% confidence intervals (i.e. cutoffs) respectively: reference SDLP ± 1.64, 2.1, 2.7, 3.21, and 4.21cm. Confidence interval limits yielded by the LSLR method varied as a function of reference SDLP, with wider intervals as the reference SDLP deviated further from its mean. The predicted value corresponding to the mean reference SDLP was associated with the following 68%, 80%, 90%, 95%, and 99% confidence intervals respectively: 1.92 + 0.88*reference SDLP ± 1.37, 1.78, 2.3, 2.74, and 3.64cm. Subsequently, both methods were applied to a dataset containing the SDLP from 182 healthy drivers who completed a standardized on-the-road driving test at a blood alcohol concentration (BAC) of 0.05% and 0.00% (placebo). Proportions of drivers classified as (statistically) impaired, unchanged, or improved were compared between CTT and LSLR classification approaches. Compared to the CTT approach, the LSLR approach classified a greater proportion of drivers with a BAC of 0.05% as impaired and a smaller proportion as improved, which suggests greater sensitivity and specificity. The clinical relevance of the proposed cutoffs remains to be determined. An advantage of the CTT method is its simplicity, while the LSLR method has presumably greater sensitivity and specificity and the added benefit of being able to include multiple predictors which could further increase classification sensitivity and specificity.

Introduction

Assessing how licit and illicit drugs affect driving performance is an important topic in drug development, policy making, and patient care. Experimental research has demonstrated the impairing effects of various drugs, especially central nervous system depressants and dissociatives, on real-world and simulated driving performance, as well as on laboratory tests of driving related skills (Arkell et al., 2020; Hayley et al., 2018; Rapoport et al., 2009; Strand, Vindenes, Gjerde, Mørland, & Ramaekers, 2019; Vermeeren, 2004; Verster & Volkerts, 2004). In addition to demonstrating impairment, it is often of interest to investigate potential beneficial effects of medicinal drugs on driving performance, for example for the treatment of attentional disorders or hypersomnia (Verster et al., 2008; Vinckenbosch et al., 2020).

For over 35 years, the Dutch standardized on-the-road driving test has been applied for demonstrating the impairing or beneficial effects of numerous licit and illicit substances on road tracking performance (Ramaekers, 2017). The test consists of an approximately 100km drive on a public motorway in a specially equipped test vehicle and under supervision of a licensed driving instructor who has access to dual controls. Participants are instructed to maintain a steady lateral position within the borders of the right traffic lane(s) and to maintain a steady velocity of 95km/h. During the test, the distance of the vehicle relative to the lane demarcations on its left, i.e. the lateral position, is continuously recorded. After removal of overtake maneuvers and other artifacts, the standard deviation of the lateral position (SDLP) is calculated from the lateral position data. The SDLP is the main outcome measure of the driving test and is considered to be a parameter of road tracking control, with higher values reflecting poorer road tracking, and vice versa. The parameter has been shown to be very reliable over repeated tests of the same driver (Verster & Roth, 2011). However, considerable interindividual variation exists between drivers (Verster & Roth, 2011). This makes the SDLP suitable as an outcome measure for experimental crossover trials, i.e. repeated measures designs, rather than between-subject randomized clinical trials, which would require considerably larger sample sizes than those required in repeated measure designs. Applied as such, the mean

SDLP increment between experimental conditions has been proven to be very sensitive for detecting drug induced effects on driving performance and is unparalleled in this respect compared to other commonly used tests of psychomotor performance (Jongen, Perrier, Vuurman, Ramaekers, & Vermeeren, 2015; Jongen, Vuurman, Ramaekers, & Vermeeren, 2016).

The clinical relevance of the drug-induced mean increment in SDLP is usually interpreted by comparing its magnitude to that observed at a blood-alcohol concentration (BAC) of 0.5g/L (0.05%), i.e. the legal limit in most of Western Europe and many other countries worldwide. Early epidemiological studies into the effects of BAC on crash risk demonstrated that a BAC of 0.05% marks the threshold for significantly increased crash risk (Borkenstein, Crowther, & Shumate, 1974; Hurst, Harte, & Frith, 1994). Hence, it is reasoned that if an investigational substance induces a mean SDLP increment compared to placebo equal to or greater than that observed at a BAC of 0,05%, relative crash risk is likely to be significantly increased for that compound as well. It has been repeatedly demonstrated that a BAC of 0.05% induces a mean SDLP increase of approximately 2.4 - 2.5cm (Jongen et al., 2017). Hence, a mean increase of 2.5cm in SDLP has been adopted as the threshold for impairment of clinical relevance (Ramaekers, 2017).

The majority of studies on the effects of drugs on driving performance employing the standardized on-the-road driving test are randomized, placebo-controlled, double-blind crossover trials (Ramaekers, 2017). Sample sizes usually range from 12-40 participants, depending on the complexity of the applied statistical models (usually linear or generalized linear models). The central aim of these statistical models is to determine whether group-level mean performance differs between the experimental conditions and in which direction. Hence, past experimental studies have focused on investigating the effects of drugs on driving performance at the aggregate level, with little regard to driving performance at the individual level.

Some consideration of the significance of change at the individual level is required to determine the margins of the McNemar test (Lachenbruch, 2014), which is a frequently applied post-hoc test in addition to the conventional parametric tests for demonstrating group level mean

differences (e.g. t-tests, ANOVA, mixed effects modelling) in drugs and driving studies applying the standardized on-the-road driving test (Jongen et al., 2017; Vermeeren et al., 2019; Vermeeren et al., 2015; Vermeeren et al., 2016; Vermeeren et al., 2014). The McNemar test is a test of marginal homogeneity, meaning that it assesses whether significant asymmetry exists in the margins of the score range, i.e. the number of observed values exceeding an upper threshold and the number below a lower threshold, while disregarding the midrange values. In the case of the standardized on-road driving test, the McNemar test has been applied for the detection of asymmetry of SDLP change scores between placebo and drug conditions in several studies (Jongen et al., 2017; Vermeeren et al., 2019; Vermeeren et al., 2015; Vermeeren et al., 2016; Vermeeren et al., 2014). The idea behind disregarding the midrange values is that small differences between conditions are arguably not representative for any effects of clinical relevance, or simply caused by measurement error. By disregarding presumed insignificant differences, the McNemar test is reasoned to be able to determine whether differences in mean SDLP are due to a meaningless shift of SDLP for the whole population or a clinically meaningful effect in a subset of the population (Laska, Meisner, & Wanderling, 2012). However, for SDLP change scores at the level of the individual driver, it is currently unclear what magnitude of change should constitute the threshold for impaired or improved road tracking performance. Past studies adopted the 2.4-2.5cm threshold used for group-level mean differences as the threshold for identifying individual impairment (Jongen et al., 2017; Vermeeren et al., 2019; Vermeeren et al., 2015; Vermeeren et al., 2016; Vermeeren et al., 2014), although this approach has not been validated. For marked improvement, no such threshold has been proposed. In fact, it has not been determined what magnitude of SDLP change at the individual level should be considered as statistically significant.

Identifying subsets of participants who demonstrate a statistically significant deviation from their placebo or baseline performance (henceforth referred to as “reference” performance) is not only useful in the context of the McNemar test, but might be of considerable value for the evaluation of individual driver performance after administration of an investigational substance. In turn this

could help to formulate more fine-tuned medical advice and prescription guidelines, as well as aiding in devising better targeted safety policies and medical treatments. The purpose of the current analysis is to present methods for finding cutoffs that signify statistically significant change in SDLP for individual drivers in experimental crossover trials employing the standardized on-the-road driving test. Two methods for determining statistical significance of individual drivers' SDLP change between conditions are proposed, one based on classical test theory and the other on least squares linear regression. Benefits and drawbacks of both methods will be discussed as well as possible applications in future research.

Methods

In this section, two methods, i.e. classical test theory (CTT) and least squares linear regression (LSLR), for determining statistical significance of individuals' SDLP change across repeated tests are described. Next, a summary of a pooled normative dataset from previous experimental crossover studies applying the standardized on-the-road driving test is provided. This normative dataset will be used for estimating the necessary (normative) parameter values needed for implementation in the methods described. Hereafter, the estimated parameter values will be implemented and applied to an explorative dataset containing pooled data from experimental crossover trials on the effects of a blood alcohol concentration (BAC) of 0.05% on road tracking performance during the standardized on-the-road driving test.

Methods for determining statistical significance of individuals' SDLP change across experimental conditions: Classical test theory method (CTT method)

In classical test theory, any obtained measurement is assumed to be the sum of a true score and measurement error. The term measurement error can be somewhat misleading because it does not only refer to actual measurement error, i.e. inaccuracies of the measurement instruments and act of applying them, but also to actual variations within the construct of interest which are introduced by

external random factors that are not the focus of investigation. In CTT, measurement error is assumed to be random, independent of individual, group, or true score, and normally distributed around zero. The standard deviation of the measurement error, known as the standard error of the measurement (SEM) is calculated from the observed standard deviation (SD) and the estimated reliability ($r_{xx'}$) of the test scores as follows:

$$SEM = SD * \sqrt{1 - r_{xx'}}$$

(Crocker & Algina, 1986). Similar to the standard deviation of the mean, it can be stated that, if the measurement of a construct for any individual *i* is repeated an infinite number of times, the true score will lie between the observed score ± 1.96 * SEM for approximately 95% of the obtained measurements (±1.96 is an approximation of the critical Z-values marking the limits of a 95% confidence interval). This is the 95% confidence interval for the true score.

However, the current analysis is not concerned with describing the (un)certainty of the true score estimate, but with the (un)certainty regarding true change in SDLP over repeated tests. Similar to a single observation, the observed difference in test scores over repeated tests is conceived as the sum of the true change and measurement error. If a condition is repeated under similar circumstances, the true change for any individual is expected to be zero. As with a single observation, the measurement error of the true change is assumed to be normally distributed with a mean of zero. However, the variance of the measurement error for the change scores is twice as large as for individual scores since two independent measurement errors are involved in the computation of a change score. Therefore, the SEM of the true change, denoted here as SEM_c is calculated as follows:

$$SEM_c = SD * \sqrt{2(1 - r_{xx'})} = \sqrt{2} * SD * \sqrt{1 - r_{xx'}}$$

(Crocker & Algina, 1986). From this, a 95% confidence interval for the true change can be determined as the observed difference between any given pair of repetitions *j* and *k* ± 1.96 * SEM_c.

The now determined confidence interval can be applied to perform a simple hypothesis test. Under the null hypothesis, it is assumed that no difference exists between the experimental conditions. Hence, the measurements can be regarded as test repetitions, and we expect the

difference between the conditions to equal zero. Because of measurement error, the observed difference for any individual will likely deviate from zero. The question then is what range of observed deviations from zero are plausible if there is no true change. If the confidence interval for true change of an individual does not include zero, it can be concluded that the observed difference is reflecting an actual true change, and the null hypothesis can be rejected in favor of the alternative hypothesis. This test has a Type I error rate of 5%. Depending on the purpose of the test, the Type I error rate can be adjusted accordingly to adopt a more liberal or conservative approach.

Methods for determining statistical significance of individuals' SDLP change across experimental conditions: Least squares linear regression method (OLLSR method)

As an alternative to the CTT method, a method based on least squares linear regression (LSLR) is proposed. Unlike the CTT method, which aims to assess whether the observed change across conditions is indicative of a true change, we apply LSLR to determine how unusual or atypical an observed score is in the light of a reference score.

In LSLR, the linear relationship between two observed variables *X* and *Y* is summarized as $Y =$

$$a + \beta X_i + \varepsilon$$

with *a* being the regression constant or intercept (i.e. the estimated mean value of *Y* if *X* = 0), β being the regression coefficient or slope (the estimated mean change in *Y* for every unit of increment of *X)*, and ε representing a residual or error term. The intercept and the slope are often referred to as *linear predictors*. These parameters can be generalized beyond the linear model to predict the value of *Y* for any new observation *X,* as long as the model satisfies the assumptions of ordinary least squares linear regression, the error is normally distributed, and the new observation belongs to the same population as the sample that was used to fit the model. As the presence of the error term implies, this prediction is subject to uncertainty. This uncertainty can be quantified in the form of a *standard error of the prediction* (SE_{pred}) (Faraway, 2002) and is calculated as

$$SE_{pred} = \sqrt{SE_{est}{}^2 + SE_{fit}{}^2}$$

with SE_{est} representing the standard error of the estimate, which is the standard deviation of the error in the regression model, i.e. the standard deviation of the residual. It is calculated as

$$SE_{est} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}[Y_i - (a + bX_i)]^2}$$

Most statistical software automatically calculates the SE_{est} when performing an LSLR. The SE_{est} refers to the mean predicted value of Y for every corresponding value of X, or put more simply: the regression line. Hence, every individual prediction will have the same $SE_{est.}$ The other term, SE_{fit}, is the standard error of the fitted value. This statistic quantifies the standard error of the predicted, or fitted, mean values of *Y*. The SE_{fit} is calculated as

$$SE_{fit} = SE_{est}\sqrt{\frac{1}{n} + \frac{(X-\bar{X})^2}{\sum(X_i-\bar{X})^2}}$$

where *n* is the number of observations, or the sample size. Unlike the SE_{est}, the SE_{fit} will be greater for the fitted values corresponding to the extremes of regressor variable *X*. In other words, the further away the value of *X* is from its mean, the greater the uncertainty associated with the predicted value $\hat{Y}$. This results in the characteristic double funnel shape of the confidence interval of the predicted values, a.k.a. the prediction interval.

Once the SE_{pred} is determined, it can be applied to calculate a prediction interval for any predicted value $\hat{Y}$ based on a newly observed score on variable *X* as

$$\hat{Y} \pm t_{\alpha/2}^{n-2} * SE_{pred}$$

where $t_{\alpha/2}^{n-2}$ is the t-value corresponding to the desired Type I error rate α , with *n-2* degrees of freedom and for a two-tailed test. Of note is that the LSLR approach assumes normality of the error distribution, which is not an explicit assumption in ordinary least squares linear regression. This assumption allows for the construction of confidence intervals around the predicted value base on the t-distribution.

In experimental crossover trials applying the on-the-road driving test, under the null hypothesis it is assumed that the SDLP values in the active condition is a linear function of the SDLP values in the reference condition, apart from measurement error. Hence, we would expect the predicted value to fall inside the prediction interval, i.e. the confidence interval around the predicted SDLP value based on the observed SDLP in the reference condition. It follows that if an observed SDLP in the active condition falls outside of the prediction interval, this can be considered as an "atypical observed value", and the null hypothesis is rejected with a type I error rate that was preselected. Hence, it is then concluded that the observation in the active condition differs significantly from the reference score, and that the driver exhibited significant impairment or improvement, depending on the direction of the prediction error.

Normative and explorative datasets

A normative dataset was compiled containing 196 (i.e. 98 pairs of) reference SDLP measurements from 3 studies including the standardized on-the-road driving test (Vermeeren et al., 2019; Vermeeren et al., 2015; Vermeeren et al., 2016). All participants in these studies were volunteers with (corrected to) normal vision, and free from any physical or psychological conditions that might affect driving performance, as established by a medical screening by a licensed clinician. Demographics of the study samples are summarized in table 1. Each of the selected studies included two standardized on-road driving tests for each experimental condition, scheduled 7 days apart, meaning that each of the 98 participants completed two reference (placebo) driving tests 7 days apart from each other.

An explorative dataset was retrieved from a study pooling data from 9 experimental crossover trials investigating the effects of alcohol (BAC 0.05%) on road tracking performance during the standardized on-road driving test (Jongen et al., 2017). This pooled sample contains data from 182 healthy participants who completed both an alcohol (BAC 0.05%) and alcohol placebo condition. Demographics of this sample are summarized in table 1.

	Study	N	%male	Age range (years)	Mean age (SD)
Normative dataset (N=98)	Vermeeren, A., Sun, H., Vuurman, E. F., Jongen, S., Van Leeuwen, C. J., Van Oers, A. C., ... & McCrea, J. (2015). On-the-road driving performance the morning after bedtime use of suvorexant 20 and 40 mg: a study in non-elderly healthy volunteers. *Sleep*, *38*(11), 1803-1813.	27	46	23-46	45.6 (13.2)
	Vermeeren, A., Vets, E., Vuurman, E. F., Van Oers, A. C., Jongen, S., Laethem, T., ... & Sun, H. (2016). On-the-road driving performance the morning after bedtime use of suvorexant 15 and 30 mg in healthy elderly. *Psychopharmacology*, *233*(18), 3341-3351.	24	58	65-76	68.8 (2.7)
	Vermeeren, A., Jongen, S., Murphy, P., Moline, M., Filippov, G., Pinner, K., ... & Vuurman, E. F. (2019). On-the-road driving performance the morning after bedtime administration of lemborexant in healthy adult and elderly volunteers. *Sleep*, *42*(4), zsy260.	47	54	23-78	58.5 (13.3)
Explorative dataset (N=182)	Jongen, S., Vermeeren, A., van der Sluiszen, N. N. J. J. M., Schumacher, M. B., Theunissen, E. L., Kuypers, K. P. C., ... & Ramaekers, J. G. (2017). A pooled analysis of on-the-road highway driving studies in actual traffic measuring standard deviation of lateral position (ie,"weaving") while driving at a blood alcohol concentration of 0.5 g/L. *Psychopharmacology*, *234*(5), 837-844.	182	51	21-59	-

Table 1. Overview of studies and respective samples that were included in the normative (test - retest under placebo) and explorative (placebo – BAC 0.05%) datasets.

Statistical analysis

CTT method

The standard deviation (SD) and reliability (i.e. Pearson correlation of test and retest, denoted as $r_{xx'}$) of the SDLP were determined from the normative dataset. The obtained estimates were then used to calculate the SEM_c. Next, the data of the explorative dataset was considered. For each pair of observations the limits of various two-sided confidence intervals (i.e. 68%, 80%, 90%, 95%, and 99%) were calculated around the observed difference, based on the SEM_c determined from the normative dataset and the critical Z-score corresponding to the limits of the desired confidence interval. Finally, it was inspected whether the respective confidence interval contained zero, in which case the null hypothesis was maintained. If not, the null hypothesis was rejected with a Type I error rate corresponding to the width of the respective confidence interval, and it was concluded that the

observed difference across conditions was reflecting true change greater or smaller than zero for the respective individual.

LSLR method

The obtained SDLP values for the first and second driving test repetition in the normative dataset were randomly assigned to the predictor ($SDLP_{test}$) and dependent ($SDLP_{retest}$) variable in order to mimic the randomized treatment order inherent to randomized experimental crossover designs. Next, a least squares linear regression was performed with $SDLP_{test}$ as the predictor and $SDLP_{retest}$ as the dependent variable. The Shapiro-Wilk test for normality of the standardized residuals (Shapiro & Wilk, 1965) and the Breusch-Pagan-Koenker(BPK) test for heteroscedasticity (Breusch & Pagan, 1979) were performed to check the respective model assumptions. Next, the linear equation yielded by fitting the normative data was applied to the explorative dataset, implementing the observed SDLP values in the placebo condition as the regressor values in order to predict the SDLP values in the alcohol condition that would be expected under the null hypothesis (i.e. absence of effect of BAC 0.05% on SDLP for the respective individual). In addition, the limits of various confidence (or prediction) intervals (i.e. 68%, 80%, 90%, 95%, and 99%CIs) were calculated around the predicted values as based on the calculated SE_{pred} and the critical t-values marking the limits of the desired two-sided prediction interval with 92 degrees of freedom (i.e. degrees of freedom for estimation of regression parameters in normative sample after removal of influential cases). Finally, it was determined whether the observed SDLP value in the active condition fell in- or outside of the respective prediction interval. Values that fell outside were considered “atypical observed values” which was interpreted as reflecting a true difference in SDLP across the experimental conditions.

Results

CTT method

The mean difference in SDLP between test and retest was observed to be 0.26cm (SE= 0.17cm). Zero falls well within the 95% confidence interval of the mean change, supporting the assumption that the expected difference between test repetitions under reference conditions is zero. The SD of the SDLP, estimated as the average of the SD at test and retest, was found to be 2.93cm. The test-retest correlation for the SDLP was strong and highly significant (r(96)= .844, p<.001). With these estimates, the SEM_c was calculated as: $SEM_c = \sqrt{2} * 2.93 * \sqrt{1 - .844} = 1.64cm$

Finally, based on the SEM_c and critical Z-values, the limits of the two-sided confidence intervals (i.e. 68%, 80%, 90%, 95%, and 99%) around the observed differences between the experimental conditions of the explorative dataset were calculated for each individual. The CTT method yielded the following limits for 68%, 80%, 90%, 95%, and 99% confidence intervals (i.e. cutoffs) respectively: reference SDLP ± 1.64, 2.1, 2.7, 3.21, and 4.21cm.

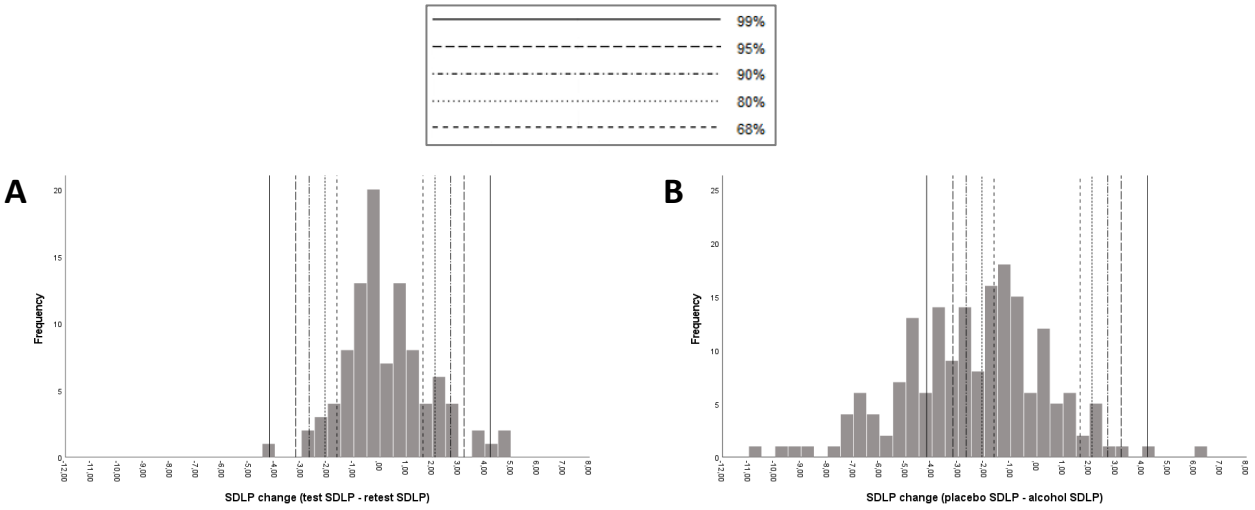

Figure 1. Illustration of the outcomes of the CTT method. **A.** Distribution of SDLP change (cm) across test and retest in the normative dataset. The horizontal lines mark the limits of the confidence intervals, as indicated in the legend in the top left quadrant, for normative change at the individual level. Negative change values indicate higher SDLP and therefore poorer road tracking at retest. **B.** Distribution of SDLP change in the explorative dataset (placebo – BAC 0.05%). Horizontal lines mark the limits of the confidence intervals for individual change as derived from the normative dataset. Negative change scores indicate higher SDLP and therefore poorer road tracking during the BAC 0.05% condition.

LSLR method

The LSLR of the normative data demonstrated that SDLP explained a large proportion of the retest SDLP scores (R^2= .712, F(1,96)= 236.82, p< .001). The regression constant (intercept) was estimated to be 2.17 (t(96)= 2.27, p< .05) and the unstandardized coefficient (slope) was estimated as 0.86 (β=.84, t(96)= 15.39, p< .001). The corresponding SE_{est} was 1.59cm.

The Shapiro-Wilk test did not provide any evidence for a deviation from normality (W(98)= .983, p= .23). However, the BPK test for heteroscedasticity suggested the presence of heteroscedasticity of the residuals (χ^2(1, N= 98)=8.04, p< .01). An inspection of the scatterplot of the standardized predicted values and residuals seemed to suggest that the heteroscedasticity in the data was caused by six cases with large standardized residuals (> |2|), rather than a systematic bias in the model. Hence, it was investigated if these cases exerted an unduly large influence on the model. To this end, it was inspected whether the Cook's Distance values (D_i) associated with these cases were larger than 4/(n-k-1), with n being sample size and k being the number of predictors. Hence, values with an associated D_i > 0.042 were considered as distorting influential cases. It was found that 4 out of the 6 cases with a standardized residual > |2| had a D_i > 0.042. These cases were excluded from the model and the LSLR was performed again on the remaining (N=94) cases.

In the new model, SDLP explained a slightly larger proportion of the retest SDLP scores (R^2= .769, F(1,92)= 305.56, p< .001). The SE_{est} was also found to have decreased to 1.37cm after removal of the influential cases, indicating increased accuracy of the predicted values. The regression constant (intercept) was found to be 1.92 (t(92)= 2.25, p< .05) and the unstandardized coefficient (slope) was 0.88 (β=.88, t(92)= 17.48, p< .001). The Shapiro-Wilk test did not suggest any deviation from normality (W(94)= .991, p= .8). The BPK test was not indicative of significant heteroscedasticity in the data (χ^2(1, N= 94)=1.08, p= .3). The SE_{fit} in the normative dataset, calculated from a mean of 16.78 cm (SD= 2.83) of the baseline SDLP and sample size of N = 94, ranged from 0.14 to 0.46 cm with a mean of 0.19 cm (SD= 0.06). The mean SE_{pred} in the normative data was calculated to be 1.38 cm (SD= 0.01), ranging from 1.38 to 1.44 cm.

The mean SE_{pred} in the explorative data was calculated to be 1.4 cm (SD= 0.05), ranging from 1.38 to 1.71 cm. For the mean value of predictor *X* (placebo SDLP, $\bar{X} = 19.58cm$), the limits of the 68%, 80%, 90%, 95%, and 99% confidence intervals were calculated to be at 19.15cm ± 1.37, 1.78, 2.3, 2.74, and 3.64cm, respectively. The SE_{pred} for predicted values $\hat{Y}$ corresponding to predictor values *X* that lie ~2 standard deviations from the mean was observed to be 1.48cm, with the limits of the 68%, 80%, 90%, 95%, and 99% confidence intervals located at 1.92 + 0.88*X* ± 1.46, 1.91, 2.45, 2.92, and 3.88cm, respectively.

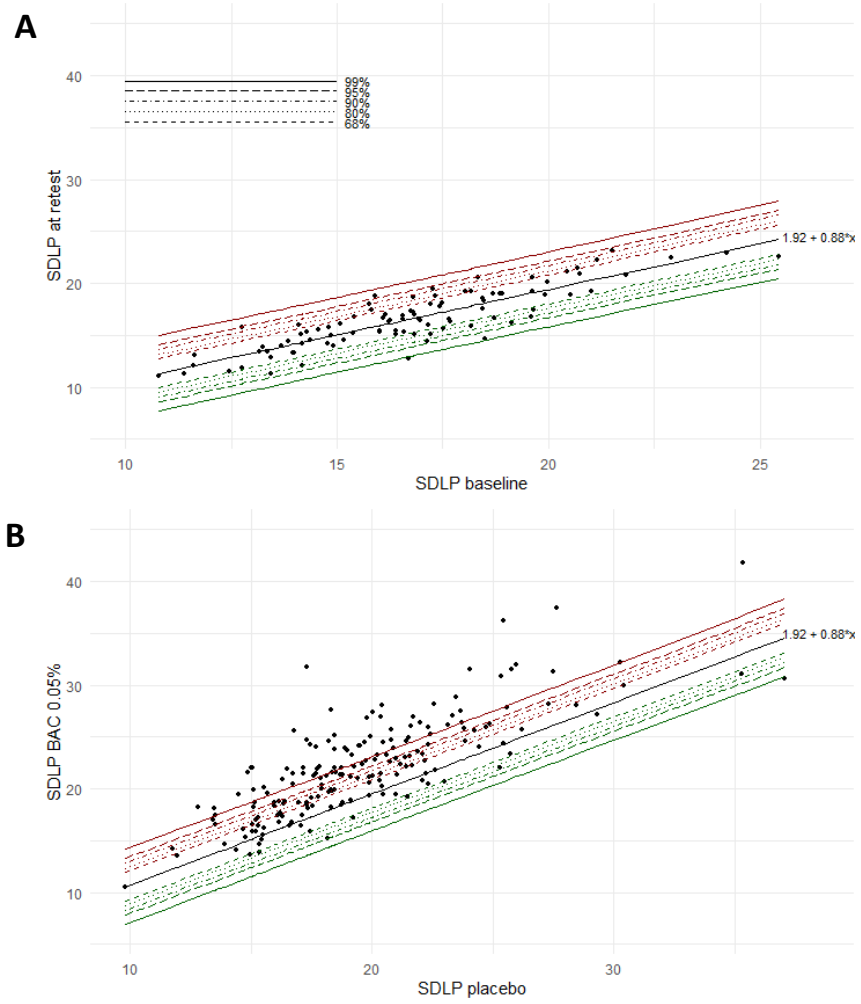

Figure 2. Illustration of the outcomes of the LSLR method. **A.** Distribution of test vs. retest SDLP (cm) for the normative dataset and the corresponding estimated regression line and prediction intervals. Values above the regression line indicate higher SDLP (poorer road tracking) at retest than what was predicted by the regression equation (indicated in top right), and vice versa. **B.** Distribution of placebo SDLP vs. BAC 0.05% SDLP for the explorative dataset. The regression line and prediction intervals, derived from the normative dataset, indicate the expected SDLP in the hypothetical absence of an effect of the BAC 0.05% based on the SDLP observed during the placebo condition. Values above the regression line indicate poorer road tracking than predicted under the null hypothesis (no drug effect).

Figure 3. Overview of the proportion of drivers in the explorative dataset demonstrating statistically significant improvement or impairment, or no statistically significant change in the BAC 0.05% condition relative to placebo performance for the various indicated confidence levels for both the CTT and LSLR methods.

Discussion

The purpose of this manuscript is to present methods for determining the statistical significance of SDLP change across repeated tests at the individual level. To this end, two methods were discussed, one based on classical test theory (CTT) and another on least squares linear regression (LSLR). The necessary parameter estimates for application of these methods were determined from a normative dataset of 98 pairs of test drives completed 7 days apart by the same driver during a placebo treatment period and then applied to an explorative dataset containing the pooled data from 9 randomized, double blind, placebo-controlled crossover trials on the effects of BAC of 0.05% on road tracking performance.

Through application of the CTT method, it was determined that cutoffs for statistically significant change in SDLP across repeated tests lie at the reference SDLP value ± 1.64, 2.1, 2.7, 3.21, and 4.21cm for the 68%, 80%, 90%, 95%, and 99% two-sided confidence intervals of SDLP change across conditions, respectively. The limits of the confidence intervals around the predicted values

using the LSLR method vary depending on the reference SDLP (predictor variable *X*), with reference values that lie further from the mean giving rise to greater uncertainty of the corresponding predicted value and consequently wider confidence intervals. The limits of the confidence interval of the predicted value corresponding to the mean reference value (reference SDLP, $\bar{X} = 19.58cm$) were calculated to be 19.15cm ± 1.37, 1.78, 2.3, 2.74, and 3.64cm for the 68%, 80%, 90%, 95%, and 99% two-sided confidence intervals, respectively. These intervals widen with ~0.18cm at each end when the predictor value *X* lies ~2 standard deviations from the mean $\bar{X}$ (i.e. covering ~95% of the observed predictor values).

The lack of a ground truth comparator prevents to determine with certainty which method is the most sensitive and specific alternative. Nevertheless, the comparison of the CTT and LSLR methods regarding the number of drivers in the alcohol dataset classified as impaired, unchanged, or improved (figure 3) suggests that the LSLR method is the most sensitive alternative for detecting drug-induced impairment of road tracking performance. In addition, it also appears to be the most specific approach, as is implied by the number of drivers classified as improved in the BAC 0.05% condition. Moskowitz, Burns, Fiorentino, Smiley, and Zador (2000) have reported that nearly all drivers demonstrate alcohol induced impairment of at least one driving related function at a BAC of 0.05% (Moskowitz et al., 2000). Of note is that in the respective report, any performance difference smaller than zero was interpreted as indicative of impairment, regardless of measurement error, although some corrections were made for time-of-day and day-to-day influences. Regardless, it seems doubtful that any healthy driver would improve driving performance in response to a BAC of 0.05%. Hence, the observed improvements in road tracking for some individuals with a BAC of 0.05% is arguably due to external factors that negatively impacted road tracking performance in the placebo condition. The results presented in figure 3 suggest that the LSLR method leads to less false positive classification as improved road tracking at a BAC of 0.05%, and therefore has greater specificity. This is conceivably because, unlike the CTT method, the LSLR method takes into account the notion of regression towards the mean. This notion holds that drivers with an initially high SDLP

are more likely to exhibit a lower SDLP at retest, and vice versa. The LSLR accommodates this principle as increased uncertainty regarding the predicted value as the predictor value moves farther away from its mean. This increased uncertainty is reflected in the wider confidence intervals surrounding predicted values at the extremes. The CTT method does not account for this. It does imply that larger changes between baseline and retest scores are less likely than small ones. However, it does not specify which observations are more likely to exhibit larger changes at retest. It therefore seems plausible that the LSLR method has superior specificity when dealing with high or low SDLP values in the reference condition.

A practical advantage of the LSLR method is that it can be further improved through the inclusion of additional predictors. Factors that potentially influence SDLP, e.g. traffic density (Teh, Jamson, Carsten, & Jamson, 2014), could be included in an LSLR model to determine if they account for a significant portion of the observed variation in the model. If so, the estimates of the regression parameter values are expected to be more precise, resulting in a lower SE_{est} which is the chief determinant of the SE_{pred}. Hence, the SE_{pred} would become smaller, and with it the width of the prediction intervals, subsequently increasing the sensitivity of the method, and potentially outperforming the CTT approach which does not lend itself to the consideration of multiple predictors.

The major practical advantage of the CTT method over the LSLR method is its simplicity. The CTT method provides clear cut-off values, depending on the selected type I error rate, which can be readily applied to a new sample. In contrast, the LSLR method requires the calculation of a unique SE_{pred} for every value of the regressor. For example, for the purpose of a post-hoc McNemar test, the cut offs provided by the CTT method appear preferable as they can be readily applied to every case in the sample.

A general note on the cutoffs yielded by both methods is that statistical significance does not necessarily correspond to clinical or real-world significance. Clinical significance of change in SDLP is generally conceived as change in SDLP that is associated with a statistically significant increase (or

decrease) in relative crash risk. However, it remains to be determined if a statistically significant increase in SDLP at the individual level at any of the proposed cutoffs is associated with increased crash risk. Arguably, any functional impairment will lead to a statistically significant increase in crash risk. However, an increase might be so small that it is extremely unlikely to contribute in any meaningful way to absolute crash risk over a lifetime of driving. Hence, a first step in determining cutoffs for clinically relevant impairment at the individual level should be to establish a threshold for meaningful increases in crash risk before investigating the level of impairment that is required to achieve such an increase in crash risk. Despite the uncertainty of the clinical relevance of the proposed cutoffs, the methods and outcomes presented in this paper arguably provide a more accurate approach to establishing cutoffs determining impairment of improvement of relative SDLP compared to the prior approach of adopting group-level thresholds to the individual level. Until further research identifies appropriate cutoffs for clinically meaningful change, it is proposed that the presented methods and corresponding cutoffs can be used as best approximations of clinically meaningful change in SDLP over repeated testing at the individual level.

References

Arkell, T. R., Vinckenbosch, F., Kevin, R. C., Theunissen, E. L., McGregor, I. S., & Ramaekers, J. G. (2020). Effect of cannabidiol and Δ9-tetrahydrocannabinol on driving performance: a randomized clinical trial. *JAMA, 324*(21), 2177-2186.

Borkenstein, R. F., Crowther, R. F., & Shumate, R. (1974). The role of the drinking driver in traffic accidents (The Grand Rapids Study). *Blutalkohol, 11*(Suppl), 1-131.

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica, 47*(5), 1287-1294. doi:10.2307/1911963

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*: ERIC.

Faraway, J. J. (2002). *Practical regression and ANOVA using R* (Vol. 168): University of Bath Bath.

Hayley, A. C., Green, M., Downey, L. A., Stough, C. K., Keane, M., Shiferaw, B., . . . Shehabi, Y. (2018). The acute and residual effects of escalating, analgesic-range doses of ketamine on driving performance: A simulator study. *Progress in neuro-psychopharmacology and biological psychiatry, 86*, 83-88.

Hurst, P. M., Harte, D., & Frith, W. J. (1994). The grand rapids dip revisited. *Accident Analysis & Prevention, 26*(5), 647-654.

Jongen, S., Perrier, J., Vuurman, E. F., Ramaekers, J. G., & Vermeeren, A. (2015). Sensitivity and validity of psychometric tests for assessing driving impairment: effects of sleep deprivation. *PLoS one, 10*(2), e0117045.

Jongen, S., Vermeeren, A., van der Sluiszen, N., Schumacher, M. B., Theunissen, E., Kuypers, K., . . . Ramaekers, J. (2017). A pooled analysis of on-the-road highway driving studies in actual traffic measuring standard deviation of lateral position (ie,“weaving”) while driving at a blood alcohol concentration of 0.5 g/L. *Psychopharmacology, 234*(5), 837-844.

Jongen, S., Vuurman, E., Ramaekers, J., & Vermeeren, A. (2016). The sensitivity of laboratory tests assessing driving related skills to dose-related impairment of alcohol: A literature review. *Accident Analysis & Prevention, 89*, 31-48.

Lachenbruch, P. A. (2014). McNemar test. *Wiley StatsRef: Statistics Reference Online*.

Laska, E., Meisner, M., & Wanderling, J. (2012). A maximally selected test of symmetry about zero. *Statistics in medicine, 31*(26), 3178-3191.

Moskowitz, H., Burns, M., Fiorentino, D., Smiley, A., & Zador, P. (2000). *Driver characteristics and impairment at various BACs*. Retrieved from

Ramaekers, J. (2017). Drugs and driving research in medicinal drug development. *Trends in pharmacological sciences, 38*(4), 319-321.

Rapoport, M. J., Lanctôt, K. L., Streiner, D. L., Bedard, M., Vingilis, E., Murray, B., . . . Herrmann, N. (2009). Benzodiazepine use and driving: a meta-analysis. *Journal of Clinical Psychiatry, 70*(5), 663.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591-611.

Strand, M. C., Vindenes, V., Gjerde, H., Mørland, J. G., & Ramaekers, J. G. (2019). A clinical trial on the acute effects of methadone and buprenorphine on actual driving and cognitive function of healthy volunteers. *British journal of clinical pharmacology, 85*(2), 442-453.

Teh, E., Jamson, S., Carsten, O., & Jamson, H. (2014). Temporal fluctuations in driving demand: The effect of traffic complexity on subjective measures of workload and driving performance. *Transportation research part F: traffic psychology and behaviour, 22*, 207-217.

Vermeeren, A. (2004). Residual effects of hypnotics. *CNS drugs, 18*(5), 297-328.

Vermeeren, A., Jongen, S., Murphy, P., Moline, M., Filippov, G., Pinner, K., . . . Van Oers, A. C. (2019). On-the-road driving performance the morning after bedtime administration of lemborexant in healthy adult and elderly volunteers. *Sleep, 42*(4), zsy260.

Vermeeren, A., Sun, H., Vuurman, E. F., Jongen, S., Van Leeuwen, C. J., Van Oers, A. C., . . . Heirman, I. (2015). On-the-road driving performance the morning after bedtime use of suvorexant 20 and 40 mg: a study in non-elderly healthy volunteers. *Sleep, 38*(11), 1803-1813.

Vermeeren, A., Vets, E., Vuurman, E. F., Van Oers, A. C., Jongen, S., Laethem, T., . . . Li, X. (2016). On-the-road driving performance the morning after bedtime use of suvorexant 15 and 30 mg in healthy elderly. *Psychopharmacology, 233*(18), 3341-3351.

Vermeeren, A., Vuurman, E. F., Leufkens, T. R., Van Leeuwen, C. J., Van Oers, A. C., Laska, E., . . . Roth, T. (2014). Residual effects of low-dose sublingual zolpidem on highway driving performance the morning after middle-of-the-night use. *Sleep, 37*(3), 489-496.

Verster, J. C., Bekker, E. M., de Roos, M., Minova, A., Eijken, E. J., Kooij, J. S., . . . Olivier, B. (2008). Methylphenidate significantly improves driving performance of adults with attention-deficit hyperactivity disorder: a randomized crossover trial. *Journal of Psychopharmacology, 22*(3), 230-237.

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.

Verster, J. C., & Volkerts, E. R. (2004). Antihistamines and driving ability: evidence from on-the-road driving studies during normal traffic. *Annals of allergy, asthma & immunology, 92*(3), 294-304.

Vinckenbosch, F., Lammers, G., Overeem, S., Chen, D., Wang, G., Carter, L., . . . Vermeeren, A. (2020). 0763 Effects of solriamfetol on driving performance in participants with narcolepsy. *Sleep, 43*, A290.

CHAPTER 8: general discussion

In the introduction of this dissertation, it was argued that there is a need for methods for the assessment of individual driving performance. It was asserted that researchers, policymakers, and clinicians bear the ethical responsibility of generating insights, advice, and regulations that prevent impaired drivers from taking to the roads, while also preventing unjustified denial of driving rights, as losing the right to drive can have a significant negative impact on independence and quality of life (Gilliam et al., 1997; Musselwhite & Haddad, 2010; Owsley, 2002; Timmermans et al., 2018). Currently, policies and clinical advice are often based on empirical findings from studies that consider outcomes at the aggregate (group) level. However, the heterogeneity of clinical populations, exemplified by the clinical sample discussed in chapter 5 of this dissertation, impedes the direct translation of findings at the group level to the individual. In other words, if a given population is found to exhibit overall impaired driving performance, this impairment is unlikely to be present (in a clinically meaningful way) for every member of that respective population. Hence, in order to formulate fair and adequate policies and clinical advice, an individual appraisal of driving capabilities is warranted for each patient for whom concerns regarding driving performance exist. An additional benefit of an individualized approach would be an increased yield of information from clinical trials. Determining which participants demonstrate (a(n)) (un)favorable response(s) to the experimental treatment(s) could provide valuable insights into patient characteristics that predict treatment response. Although several test batteries exist that serve the purpose of individual driving assessment, these are generally aimed at assessing driving performance of members of single narrowly defined clinical population, e.g. Alzheimer's or Parkinson's patients (Brown et al., 2005; Devos et al., 2007; Hunt et al., 1997), and can therefore not be readily applied to other patient groups. Hence, the need for methods of individual driving assessment for many more clinical populations remains.

In chapter 2, an overview of various tools for assessing driving performance and skills related to driving are introduced. Each of these tools is capable of assessing specific, although often overlapping, aspects for driving behavior or related functions. At Maastricht University, the most prominent tool is the standardized on-the-road driving test (Ramaekers, 2017; Verster & Roth, 2011), introduced in chapters 1 and 2. Although this test is by no means intended to assess the whole range of behaviors and functions that mediate driving behavior under every thinkable circumstance, the monotonous and prolonged nature of the test does make it ideal for assessing sustained attention or vigilance during prolonged motorway driving. This is illustrated by the successful application of the standardized on-the-road driving test and the SDLP for the detection of the effects of numerous sedative drugs (Jongen, Vuurman, Ramaekers, & Vermeeren, 2018; van der Sluiszen et al., 2016; Vermeeren et al., 2002; Verster, Veldhuijzen, Patat, Olivier, & Volkerts, 2006), sleepiness (Jongen, Perrier, Vuurman, Ramaekers, & Vermeeren, 2015), and mental load (Jongen, van der Sluiszen, Brown, & Vuurman, 2018), as well as by the convergent validity with laboratory tests designed for assessing vigilance and attention (Jongen et al., 2015; Jongen, Vuurman, Ramaekers, & Vermeeren, 2016; Ramaekers, Uiterwijk, & O'hanlon, 1992). Hence, the standardized on-the-road driving test is well-suited for investigating the effects of CNS depressants and/or the effects of medical conditions that are often characterized by impairments of sustained attention and vigilance on driving performance, for example the clinical trials on the effects of the novel stimulant drug solriamfetol on the driving performance of narcolepsy and obstructive sleep apnea patients presented in chapters 3 and 4 of this dissertation.

Maastricht University is currently one of the few research institutes that routinely performs the standardized on-the-road driving tests. Although several research groups conduct driving assessments on public roads, albeit with different procedures and outcomes (Brown et al., 2005; Devos et al., 2007; Hunt et al., 1997; Patomella, Caneman, Kottorp, & Tham, 2004), Maastricht University is able to do so for the investigation of the effects of (novel) medicinal and illicit substances on driving performance, which is rarely permitted elsewhere because of legal and safety

concerns. Of note is the excellent safety record of the on-the-road driving test at Maastricht University, with zero accidents after about 1.5 million completed kilometers and counting, thanks to the excellent safety supervision by the driving instructors. The impeccable safety history of the standardized on-the-road driving tests at Maastricht University is without a doubt the key condition for being able to continue this empirically valuable test.

The rather unique position of Maastricht University with regards to the standardized on-the-road driving test, especially for psychopharmacological research, has the advantage of attracting many external partners, such as pharmaceutical industry, government entities, and other research institutes which bring with them ample research opportunities as they seek collaboration in order to exploit the exceptional ecological validity of the test for answering their research questions. However, from a scientific point of view, it's unfortunate that Maastricht University is one of the few research institutes that routinely applies the standardized on-the-road driving test. The lack of a sizable research community employing the standardized on-the-road driving test implies limited capacity for replication of on-road study findings across independent research groups and limited means and capacity to further develop or expand the golden 'standard' of drugs and driving research.

Indeed, throughout the years, the application of the standardized on-the-road driving test and its outcomes have remained unaltered. The test has almost exclusively been employed in randomized, placebo-controlled crossover trials for the investigation of drug effects on SDLP at the group level. Although outcome measures other than the SDLP, such as mean and standard deviation of the velocity, time to lane crossing, time driven out of lane, steering angle and standard deviation of steering angle, and eye gaze have been recorded incidentally or regularly since the early days of the standardized on-the-road driving test, these were often derivative measures of SDLP or had lower sensitivity to drug effects as compared to SDLP (Jongen et al., 2016; Verster & Roth, 2011). More recently, an attempt was made to develop lane drifts as a novel driving measure, independent of SDLP Verster, Bervoets, de Klerk, & Roth, 2014; Verster & Roth, 2014). However, our pooled

analysis of data collected in on-road driving studies that was presented in chapter 6 demonstrated that lane drifts are not a sensitive outcome measure of driver impairment due to its rare occurrence, even when treatment-induced increments in SDLP are evident. So far, no efforts to explore ways of applying the standardized on-the-road driving test for extracting information about individual driving performance have been made. As will be elaborated upon in the next section, the SDLP is not an ideal outcome for assessing driving performance at the individual level. It will be argued that in order to apply the standardized on-the-road driving test for the assessment of individual driving performance, additional outcome measures need to be implemented.

Limitations of the SDLP as a parameter of fitness to drive in individual drivers

As posed in the general introduction of this dissertation, the wide range of normative SDLP values impedes its use as a direct measure of driving performance of individual drivers. What is considered a normal SDLP for one healthy and capable driver, might be indicative of severe impairment for another as compared to the individual's own baseline score. In contrast, the SDLP is a very reliable measure within the individual, meaning that subsequent assessments of the same driver under similar circumstances will yield SDLP values that are very close to each other. To illustrate the imbalance of the magnitude of variability of SDLP values between subjects and within subjects, consider the following: Verster and Roth (2011) reported a normative mean SDLP of 18.79cm with a standard deviation of 4.13cm. Hence, it is expected that 95% of healthy drivers will exhibit an SDLP between 10.7cm and 26.88cm (normative range). In contrast, in chapter 7 of this dissertation it was determined that a change in SDLP larger than ~3.2cm can be considered as statistically significant for an individual driver, with the conventional type I error rate of 5%. Hence, what is considered to be a nominal difference in SDLP between two drivers can, in extremis, be up to ~5 times as large as what is considered a statistically significant difference in SDLP for a single driver tested at successive times.

A practical consequence of this is that in order to employ the SDLP for assessing individual driving performance, repeated testing is needed, which excludes its application as an outcome for

individual assessment in study designs that do not include repeated measures, e.g. the cross-sectional study on long-term use of benzodiazepines receptor agonists presented in chapter 5. For these and many other patient groups, a pre-treatment (or pre-medical condition) SDLP as a reference is simply not available or readily obtainable, as this would require resource draining longitudinal approaches or unethical experimental interventions (e.g. administering a potentially addictive substance such as BZRAs for prolonged periods). Hence it can often not be determined whether driving performance deteriorated after BZRA treatment initiation for an individual driver in the sample discussed in chapter 5. From a research point of view, this means that those patients who are actually impaired cannot be considered in isolation from those who are not, which limits the potential to investigate which characteristics might predict driving impairment in response to a medicinal treatment or as a consequence of a medical condition from an individualized analysis approach. In clinical practice and policy making, this means that no clear conclusion can be reached about driving fitness for most patients based on performance during the standardized on-the-road driving test, as there is no reference to compare performance to. Arguably, this is a waste of the potential of this ecologically valid test. A possible solution could be to explore ways for advancing standardized on-the-road driving test in order to enable the formulation of valid statements about driving fitness based on the observed performance during a single test.

Within repeated measures paradigms, the interpretation of individual SDLP change across conditions is limited by the lack of a threshold for clinically significant change across repeated tests at the individual level. Although statistical significance of change can be demonstrated using the methods discussed in chapter 7, it is important to note that statistical significance does not automatically imply a clinically meaningful effect. The conventional conceptualization of clinically relevant impairment is impairment to a sufficient magnitude to be statistically associated with an increase in crash risk (or decrease for improvements) (Ramaekers, 2017). At the group level, the mean increase in SDLP (~2.5cm) observed at a blood alcohol content (BAC) of 0.05% is used as a threshold for clinical significance (Jongen et al., 2017; Ramaekers, 2017). The reasoning for this is

based on early epidemiological research which demonstrated a statistically significant increase in crash risk from a BAC of ~0.05% onwards (Borkenstein, Crowther, & Shumate, 1974; Jongen et al., 2017; Ramaekers, 2017). Hence the reasoning is that a mean increase of 2.5cm in SDLP is indicative of a clinically relevant impairing effect at the group level. For the lack of better alternatives, several past studies have adopted the 2.5cm increase as the threshold for clinical significance at the individual level as well (within the context of post-hoc McNemar test which requires a threshold for individual performance change; see chapter 7) (Jongen et al., 2017; Vermeeren et al., 2019; Vermeeren et al., 2015; Vermeeren et al., 2016; Vermeeren et al., 2014). However, whether the 2.5cm threshold for clinical significance can be readily applied to individual drivers is unclear. The 2.5cm threshold represents the average increase in SDLP in response to a BAC of 0.05%. This increase has a standard deviation of ~3cm (Jongen et al., 2017), resulting in a wide range of normative change values (-3.4cm to 8.4cm). From this it can be derived that approximately 40% of the drivers do not reach the point of statistically significant increase, as in chapter 7 it was determined that the point of statistical significance for change at the individual level lies at ~+3.2cm (for a 5% type I error rate). Note that the lower border of the confidence interval (-3.4cm) suggests that some individuals even demonstrated statistically significant improvement in road-tracking performance. Although changes at the individual level are conceivably more susceptible to random external influences which might obscure the actual change in road tracking skill, as will be argued further on in this section, it seems plausible that a significant portion of drivers will not experience any significant impairment at a BAC of 0.05%. Whether it is a valid approach to estimate a threshold of clinical relevance for individual drivers based on a sample that possibly contains a considerable portion of unimpaired drivers remains to be determined.

If a threshold for clinically relevant change in SDLP at the individual level were established, it would be possible to infer whether an individual is impaired when SDLP is inflated beyond that threshold. However, it would still remain unclear whether a clinically meaningful decrease in SDLP in individual patients with suspected driving impairment is indicative of restoration of normative

performance or just reflects an attenuation of impairment. Similarly, it is not possible to conclude whether baseline performance was already impaired for any given individual, because of the limited informational value of absolute SDLP as assessed in a single drive. Hence, the interpretation of observed SDLP would often remain elusive in patient samples where assumptions about baseline driver state (impaired vs. unimpaired) suffer from a considerable degree of uncertainty, as opposed to samples consisting of healthy control participants for which it is reasonable to assume that baseline driving performance is unimpaired. Yet, it is clear that absolute rather than relative driving performance is the central focus for the clinical and legal management of driving while using certain medications or when suffering from certain medical conditions. Knowing that a persons' driving performance has worsened is of limited informational value without knowing if it worsened enough to be considered as impaired in an absolute sense (i.e. statistically significant increase of crash risk) and vice versa for improvement.

A final conceivable problem with using SDLP for individual driving assessment, as opposed to aggregate level assessments, is the increased uncertainty whether changes in SDLP are attributable to changes road tracking *aptitude* (i.e. latent skill level that is semi-constant over time and circumstances; see chapter 2 for definition). For example, it has been observed that the SDLP increases in response to increasing traffic density and mobility (Teh, Jamson, Carsten, & Jamson, 2014). Arguably, this increase is caused by increased allocation of attentional resources to monitoring traffic at the cost of road tracking. If traffic density by chance increases across experimental conditions, the SDLP could also be observed to increase. However, this observed change in SDLP could be the consequence of a strategic choice by the driver to reallocate attention to maximize safety, and not because of decreased road tracking aptitude. Hence the observed increase in SDLP would not necessarily be indicative of impaired driving performance. Many more speculative yet plausible examples could be proposed that would nudge the SDLP in either direction. At the aggregate level, the influence of such random influences are of limited concern since it is often reasonable to assume that these factors are indeed random and therefore not systematically

associated with the experimental condition(s) or independent variable(s). However, on the individual level, the signal (contribution of drivers' road tracking skills) to noise (other factors that might affect SDLP) ratio is much smaller, which decreases the confidence to ascribe changes in SDLP to changes in road tracking skill.

In conclusion, the standardized on-the-road driving test is a valuable tool for driving assessment given its ecologically valid nature. Past research has demonstrated the potential of the test and its main outcome measure, the SDLP, for demonstrating the effects of impaired attention and vigilance during real-world on-road driving at the aggregate level. However, the wide range of inter-individual differences in SDLP limits its use for the evaluation of fitness to drive of individual drivers based on a single driving tests. Change scores across multiple conditions are more informative for individual driving assessment. However, interpretation of change scores is limited to establishing the statistical significance and direction of an effect, without an unequivocal interpretation of real-world significance as long as a threshold for clinically relevant change at the individual level is not established. Furthermore, in the hypothetical instance that a threshold for clinically relevant change in SDLP were established, it would remain difficult to determine whether presumed impaired drivers (e.g. increased risk individuals such as sleep apnea patients) are still impaired in an absolute sense after a clinically meaningful decrease in SDLP, rather than demonstrating an attenuation of impairment. Also, it would not be possible to make a confident statement about the respective participant's baseline driving performance, which would be warranted when investigating whether an investigational treatment causes remission of presumed driving impairment (which implies the presence of driving impairment at baseline). Finally, for many populations it is often logistically too complex to obtain a suitable reference SDLP. Taken together, the presented discussion on the limitations of the SDLP as a measure of individual driving performance suggests a need for additional outcomes that would allow to make statements about individual driver fitness, ideally based on a single test drive.

Recommendations for measures of individual fitness to drive

In order to enable the standardized on-the-road driving test to be employed for individual driving assessment during a single test repetition, existing driver monitoring systems (DSMs) can be considered as a starting point. These systems are aimed at continuously monitoring and assessing individual driver behavior (e.g. eye opening and blink rate) or driving performance (e.g. steering wheel movement) and triggering timely warning messages when impairment is detected (Hayley et al., 2021; Kang, 2013). Impairment is usually operationalized as the detection of parameter values that are outside of a predetermined normative range (Hayley et al., 2021; Kang, 2013). Arguably, such normative operationalizations risk achieving suboptimal sensitivity if the normative range of the respective measure is too wide, as is the case with SDLP. Furthermore, DSMs tend to focus on either driver behavior (e.g. gaze, head movement, etc.) or driving performance (i.e. vehicle based measures such as steering wheel movements and braking power, etc.). It is arguably preferable to collect both driver based and vehicle based measures in parallel in order to ascertain that presumed impairments observed in vehicle based measures are associated with indications of impairments of driver attention in driver based measures, and vice versa. Proposedly, indications of sleepiness, fatigue or inattention might not be accompanied by observable driving impairment if the attentional impairment experienced by the driver is mild. This cannot be confidently elucidated without time synced information of driving performance (i.e. vehicle based measures). Conversely, presumed indications of driving impairment might as well be caused by random external events (e.g. decreased visibility) instead of by an impaired driver. This cannot be ascertained without time synced information on driver state. Indeed, it has been posed that DSMs would benefit from an hybrid approach that includes both behavioral parameters of driver attention and functional outcomes of driving performance (Grüner & Ansorge, 2017; Hayley et al., 2021; Kang, 2013). However, a practical hurdle is the development of reliable fusion methods that allow for the integration of driver and vehicle based measures to enable the classification of driver state (i.e. impaired vs. unimpaired) (Hayley et al., 2021; Kang, 2013).

The use of artificial intelligence might be a promising approach to solving the complex issue of integrating driver and vehicle based measures. Training artificial intelligence as classifiers through machine learning algorithms has gained traction over the last years and will likely play an increasingly important role in future research and engineering (Bonaccorso, 2017; El Naqa & Murphy, 2015; Zhang, 2020). Simply put, machine learning is the process of learning a computer to recognize complex patterns in data through the presentation of many already classified examples (i.e. ground truth). It is conceptually possible to train a computer to distinguish between impaired and unimpaired drivers through the presentation of examples of a collection of vehicle and driver based measures coming from drivers known to be impaired or unimpaired. However, this is potentially also the Achilles heel of a machine learning approach. In order to properly train a computer to distinguish between impaired and unimpaired driver, a large dataset of drivers whose absolute performance status is already known is required. This might seem somewhat circular, since elucidating whether an individual driver is impaired or not is exactly the question we want to know the answer to.

Nevertheless, gathering data from known impaired and unimpaired drivers is arguably possible with a reasonable degree of accuracy within the context of an experimental crossover trial if three requirements are met. Firstly, participants should be healthy volunteers with a valid drivers' license and who drive regularly. This requirement should provide a reasonable degree of certainty that performance of each driver in the reference (baseline or placebo) condition is unimpaired and can therefore be used as an example of an unimpaired driver for the machine learning algorithm. Secondly, in the active condition, the experimental challenge should be selected based on its established association with increased accident risk and impairing effects on measures of driving performance and related functioning at the population level, e.g. alcohol at a BAC of 0.05% (Borkenstein et al., 1974; Jongen et al., 2017; Kuypers, Legrand, Ramaekers, & Verstraete, 2012; Moskowitz, Burns, Fiorentino, Smiley, & Zador, 2000) or the hypnotic drug zopiclone 7.5mg (Gustavsen et al., 2008; Vermeeren et al., 2002; Verster, Warren Spence, Shahid, R Pandi-Perumal, & Roth, 2011). This requirement should increase the chance and plausibility that any observed

impairments are reflective of increased crash risk, and therefore have clinical relevance. Thirdly, individual drivers are only labeled as impaired in the active experimental condition whenever both vehicle based and driver based measures show statistically significant deteriorations. Statistical significance of change across conditions can be determined by adopting the statistical methods described in chapter 7 of this dissertation, which can arguably be applied to all continuous measures with sufficient reliability. Ascertaining that both vehicle and driver based measures show significant deterioration prevents the classification of drivers as being impaired when in fact the perceived deterioration of vehicle based measures is reflective of the influence of random factors rather than impaired attention to the driving task.

However, it should first be decided which measures will be collected during the driving task for presentation to the algorithm. As argued above, the presented data should consist of both vehicle based parameters of driving performance and driver based measures of attention and vigilance. Among the candidates for parameters of arousal and attention are the outcomes of various laboratory neuropsychological tests that are specifically designed for assessing these cognitive constructs. Laboratory tests of arousal and attention often rely on the visual modality or require multiple response buttons and are therefore not easily integrated into the standardized on-the-road driving test. Nevertheless, several auditory reaction time tasks could conceivably be adapted for implementation in the standardized on-the-road driving test (Doverspike, Cellar, & Barrett, 1986; Jung, Ronda, Czeisler, & Wright Jr, 2011; Seli, Cheyne, Barton, & Smilek, 2012). Proposedly, it would be possible to adapt visual attention tasks for execution during the standardized on-the-road driving test as long as the stimuli are presented in the periphery of the visual field of the driver or are presented very briefly centrally (e.g. through the use of a head up display). However, a possible concern of introducing additional tasks to be executed during the standardized on-the-road driving task is that it might interfere with driving performance (Broeker et al., 2020; Jongen, van der Sluiszen, et al., 2018; Levy & Pashler, 2008; Wester, Verster, Volkerts, Böcker, & Kenemans, 2010). In general, task performance decreases as a function of increasing complexity of the concurrent task

(McDowd, Vercruyssen, & Birren, 2020; Vaportzis, Georgiou-Karistianis, & Stout, 2013). Hence, in order to minimize the influence of concurrent attention tasks on driving performance during the standardized on-the-road driving test, simple tasks such as simple reaction time tasks (one stimulus and one response option) tests appear preferable to more complex tasks such as choice reaction time tasks or tasks requiring higher order cognitive processes. Nevertheless, it remains important to investigate the impact of any introduced concurrent task on measures of driving performance during the standardized on-the-road driving test in order ensure that potential signatures of inattention in the lateral position are generalizable to driving without the parallel execution of the respective concurrent task. This consideration also goes the other way, i.e. performance on a secondary attention task might fluctuate as a function of fluctuations in demands of the primary driving task (Becic et al., 2010). Therefore, discrete events in the secondary task that seem to suggest inattention (e.g. missed stimuli or prolonged reaction times) might be caused by pressing traffic circumstances which require prioritization of the primary driving task at the cost of the secondary attention task, and not because of an overall decrease in vigilance or attention.

In addition to (adapted) laboratory tests of arousal and attention, physiological measures of attention and arousal, such as power of neural oscillations in specific frequency bands associated with wakefulness and attention as measured trough electroencephalography (EEG), muscle tensions as measured through electromyography (EMG) (Lohani, Payne, & Strayer, 2019; Mahmoodi & Nahvi, 2019), breathing rate, heart rate (HR) and heart rate variability (HRV)(Fujiwara et al., 2018; Lohani et al., 2019; Wolkow et al., 2020), pupil size(Lohani et al., 2019; Schwalm, Keinath, & Zimmer, 2008), electrodermal activity (EDA) (Collet, Petit, Priez, & Dittmar, 2005; Larue, Rakotonirainy, & Pettitt, 2011), and peripheral body temperature (Lohani et al., 2019) can be considered for integration in the standardized om-the-road driving test. Collecting physiological parameters often requires attaching sensors to the participant. This might cause discomfort during the drive, especially with EEG, as it requires multiple electrodes to be attached to the head and face of the participant which might restrict head and torso movements. Also, movements and vibrations during the test could distort the

signal. The practical limitations of EEG measurements during the standardized on-the-road driving tests are unfortunate as EEG has excellent sensitivity for detecting fluctuations in activity of corticothalamic networks that underlie sleep-wake behavior (Oken, Salinsky, & Elsas, 2006). Despite its practical limitations, EEG has been successfully applied in the context of the standardized on-the-road driving test in the past (Perrier et al., 2016; Perrier et al., 2015; Ramaekers et al., 1992). Of note is that these studies did not report (Ramaekers et al., 1992) or establish a statistically significant correlation between changes in SDLP and fluctuations EEG alpha, beta, and theta power spectra (Perrier et al., 2016; Perrier et al., 2015). However, the primary statistical omnibus tests did suggest changes in the same direction for SDLP and EEG power spectra and the sample sizes (n=16 and n=20, respectively) suggest that the correlation analysis was underpowered for demonstrating a statistically significant correlation of medium strength (Pearson correlation coefficient of .03 - .05). Changes in EEG activity have been observed to be significantly associated with drowsiness and attention in numerous other studies (Lal & Craig, 2001). Hence, it is argued that EEG is worth being considered as an add-on measure during the standardized on-the-road driving test. Several wireless or simplified EEG systems (Gangadharan & Vinod, 2021; Ogino & Mitsukura, 2018; Shameen, Yusoff, Saad, Malik, & Muzammel, 2018) are available or are currently being developed, which makes EEG a promising technology for the purpose of quantifying driver attentional states. Other physiological parameters such as EMG, HR, HRV, BP, EDA, and peripheral body temperature can generally be collected in more convenient and swift way. Although these physiological parameters are generally considered to be less sensitive than EEG or have not progressed far enough to achieve similar accuracy in the driving context, their combined input together with machine learning techniques might allow for the utilization of what limited the information they may provide such that their combined input allows for a sufficiently accurate classification of driver attentional states (Lohani et al., 2019).

Apart from physiological measures, observable driver behavior can be monitored to derive measures that are indicative of driver attentional state. The relevant locus for monitoring driver

behavior is the head, and more specifically the eyes (Ebrahim, Stolzmann, & Yang, 2013; Hayley et al., 2021; Shiferaw, Downey, & Crewther, 2019; Shiferaw, Crewther, & Downey, 2019; Stawarczyk, François, Wertz, & D'Argembeau, 2020). Head and eye position and movements are strong behavioral correlates of attention and vigilance and are conveniently monitored through the collection of video footage (Ebrahim et al., 2013; Shekari Soleimanloo et al., 2019; Shiferaw et al., 2019; Shiferaw et al., 2019; Stawarczyk et al., 2020). This excludes the need for time consuming, movement restricting, and uncomfortable sensor attachment. The difficulty in using behavioral correlates of attention and vigilance lies in objective quantification of the parameters. In order to objectively quantify relevant aspects of driver behavior, researchers have turned to machine learning techniques. Many successful algorithms have been developed that can detect faces, eye movements, measure the speed and frequency of eye movements, determine direction of gaze, measure blink rates and speed, and so on (Ebrahim et al., 2013; Shekari Soleimanloo et al., 2019; Shiferaw et al., 2019; Shiferaw et al., 2019). Ocular parameters have been observed to change in response to sleepiness and alcohol intoxication during simulated driving and to correlate with laboratory performance measures of attention (Ebrahim et al., 2013; Grüner & Ansorge, 2017; Shekari Soleimanloo et al., 2019; Shiferaw et al., 2019; Stawarczyk et al., 2020; Zandi, Quddus, Prest, & Comeau, 2019).

It is clear that there are plenty of candidate driver based physiological and behavioral parameters to be included in the standardized on-the-road driving test. This leaves the consideration of which vehicle based measures to include. Ever since the dawn of the standardized on-the-road driving test, the SDLP has been the main outcome measure and has proven its value in numerous experimental and observational studies, as discussed throughout this thesis. Although not fit for individual driving assessments by itself, the SDLP is arguably deserving of continued use as a parameter of driving performance within the context of an advanced DSM integrating various parameters. Proposedly, the evolution of the SDLP over time within a single drive might prove to be a useful parameter of individual driving performance. However, the SDLP is not necessarily the only

useful parameter of driving performance based on the lateral position signal. A prior explorative investigation has attempted to extend the outcomes of the standardized on-the-road driving test with *lane drifts* which were suggested to capture momentary lapses of attention (Verster et al., 2014, discussed in chapter 6 of this dissertation). The proposed operationalization was concluded to provide little information in addition to the SDLP, as it was found to be a correlate of SDLP with less sensitivity due to its infrequent occurrence (Vinckenbosch, Vermeeren, Verster, Ramaekers, & Vuurman, 2020). Furthermore, it remains to be established whether such events would actually reflect attentional fluctuations. Nevertheless, the idea of identifying discrete signatures in the lateral position which reflect meaningful fluctuations in driver attention is worth further exploration. Alternatively, additional meaningful parameters might be derived from steering wheel movement, velocity, braking behavior, or gap acceptance (Hayley et al., 2021; Kang, 2013; Mahajan & Velaga, 2020; Papantoniou, Papadimitriou, & Yannis, 2017; Yadav & Velaga, 2019).

The selection of driver and vehicle based measures should arguably be based on their sensitivity, specificity, and practicality. However, it is suggested that it is the balance between the sensitivity and specificity on the one hand, and practicality on the other, that should be decisive when deciding on the application of a given measure. Measures with low sensitivity and/or specificity can still be considered as additions when their collection can be achieved in a convenient and affordable way. One of the strengths of machine learning is that it can embed Bayesian inferential statistics (Tipping, 2003), which allows otherwise low information yielding measures to provide a meaningful contribution the a DSM's overall acuity when it is considered against a background of other measures.

A final note on machine learning techniques for the development of advanced DSMs to be incorporated in the standardized on-the-road driving test is that a machine learning algorithm will only learn to recognize a specific type of driving impairment. Conceivably, driving impairment due to alcohol intoxication might differ qualitatively from that induced by sleep deprivation. Hence, a

trained algorithm can only be applied to populations or experimental manipulations that are qualitatively comparable with respect to the dataset that was used to train it. This means that different algorithms need to be trained for different purposes. Nevertheless, the efforts required to make this approach a reality are arguably worthwhile in the light of the potentially considerable gain of valuable information for empirical studies and applicability for the assessment of individual driving performance which is needed for the formulation of adequate and fair clinical advice and legal policies.

Conclusion

In this final chapter, it is argued that the standardized on-the-road driving test in its current form, i.e. with SDLP as its main and de facto only outcome measure, is not fit for individualized assessments of driving performance. It is proposed that in order to exploit the ecological validity of the standardized on-the-road driving test to meet the need for individualized assessments, the set of outcome measures should be appended. It is reasoned that a comprehensive set of outcomes comprises both vehicle and driver based measures. The data originating from these multiple sources could conceivably be integrated through machine learning techniques. Ground truth examples which are required for machine learning algorithms can arguably be provided through carefully designed experimental crossover trials including healthy participants and an experimental challenge that is known to be associated with increased accident risk and impaired driving performance. Proposedly, participants who demonstrate statistically significant deterioration (established through application of the methods outlined in chapter 7) on both vehicle based measures of driving performance and driver based measures of driver attentional state, can be reasonably assumed to demonstrate clinically meaningful driving impairment. Hence, data from these drivers can be used as ground truth examples of impaired driver data.

References

Becic, E., Dell, G. S., Bock, K., Garnsey, S. M., Kubose, T., & Kramer, A. F. (2010). Driving impairs talking. *Psychonomic bulletin & review, 17*(1), 15-21.

Bonaccorso, G. (2017). *Machine learning algorithms*: Packt Publishing Ltd.

Borkenstein, R. F., Crowther, R. F., & Shumate, R. (1974). The role of the drinking driver in traffic accidents (The Grand Rapids Study). *Blutalkohol, 11*(Suppl), 1-131.

Broeker, L., Haeger, M., Bock, O., Kretschmann, B., Ewolds, H., Künzell, S., & Raab, M. (2020). How visual information influences dual-task driving and tracking. *Experimental brain research*, 1-13.

Brown, L. B., Ott, B. R., Papandonatos, G. D., Sui, Y., Ready, R. E., & Morris, J. C. (2005). Prediction of on-road driving performance in patients with early Alzheimer's disease. *Journal of the American Geriatrics Society, 53*(1), 94-98.

Collet, C., Petit, C., Priez, A., & Dittmar, A. (2005). Stroop color–word test, arousal, electrodermal activity and performance in a critical driving situation. *Biological psychology, 69*(2), 195-203.

Devos, H., Vandenberghe, W., Nieuwboer, A., Tant, M., Baten, G., & De Weerdt, W. (2007). Predictors of fitness to drive in people with Parkinson disease. *Neurology, 69*(14), 1434-1441.

Doverspike, D., Cellar, D., & Barrett, G. V. (1986). The Auditory Selective Attention Test: A review of field and laboratory studies. *Educational and psychological measurement, 46*(4), 1095-1103.

Ebrahim, P., Stolzmann, W., & Yang, B. (2013, 13-16 Oct. 2013). *Eye Movement Detection for Assessing Driver Drowsiness by Electrooculography.* Paper presented at the 2013 IEEE International Conference on Systems, Man, and Cybernetics.

El Naqa, I., & Murphy, M. J. (2015). What is machine learning? *machine learning in radiation oncology* (pp. 3-11): Springer.

Fujiwara, K., Abe, E., Kamata, K., Nakayama, C., Suzuki, Y., Yamakawa, T., . . . Masuda, F. (2018). Heart rate variability-based driver drowsiness detection and its validation with EEG. *IEEE transactions on biomedical engineering, 66*(6), 1769-1778.

Gangadharan, S., & Vinod, A. (2021). Drowsiness detection using portable wireless EEG. *Computer Methods and Programs in Biomedicine*, 106535.

Gilliam, F., Kuzniecky, R., Faught, E., Black, L., Carpenter, G., & Schrodt, R. (1997). Patient-validated content of epilepsy-specific quality-of-life measurement. *Epilepsia, 38*(2), 233-236.

Grüner, M., & Ansorge, U. (2017). Mobile eye tracking during real-world night driving: A selective review of findings and recommendations for future research. *Journal of Eye Movement Research, 10*(2).

Gustavsen, I., Bramness, J. G., Skurtveit, S., Engeland, A., Neutel, I., & Mørland, J. (2008). Road traffic accident risk related to prescriptions of the hypnotics zopiclone, zolpidem, flunitrazepam and nitrazepam. *Sleep medicine, 9*(8), 818-822.

Hayley, A. C., Shiferaw, B., Aitken, B., Vinckenbosch, F., Brown, T. L., & Downey, L. A. (2021). Driver monitoring systems (DMS): The future of impaired driving management? *Traffic injury prevention, 22*(4), 313-317.

Hunt, L. A., Murphy, C. F., Carr, D., Duchek, J. M., Buckles, V., & Morris, J. C. (1997). Reliability of the Washington University Road Test: A performance-based assessment for drivers with dementia of the Alzheimer type. *Archives of neurology, 54*(6), 707-712.

Jongen, S., Perrier, J., Vuurman, E. F., Ramaekers, J. G., & Vermeeren, A. (2015). Sensitivity and validity of psychometric tests for assessing driving impairment: effects of sleep deprivation. *PLoS one, 10*(2), e0117045.

Jongen, S., van der Sluiszen, N. N., Brown, D., & Vuurman, E. F. (2018). Single-and dual-task performance during on-the-road driving at a low and moderate dose of alcohol: A

comparison between young novice and more experienced drivers. *Human Psychopharmacology: Clinical and Experimental, 33*(3), e2661.
Jongen, S., Vermeeren, A., van der Sluiszen, N., Schumacher, M. B., Theunissen, E., Kuypers, K., . . . Ramaekers, J. (2017). A pooled analysis of on-the-road highway driving studies in actual traffic measuring standard deviation of lateral position (ie,"weaving") while driving at a blood alcohol concentration of 0.5 g/L. *Psychopharmacology, 234*(5), 837-844.
Jongen, S., Vuurman, E., Ramaekers, J., & Vermeeren, A. (2016). The sensitivity of laboratory tests assessing driving related skills to dose-related impairment of alcohol: A literature review. *Accident Analysis & Prevention, 89*, 31-48.
Jongen, S., Vuurman, E., Ramaekers, J., & Vermeeren, A. (2018). Comparing the effects of oxazepam and diazepam in actual highway driving and neurocognitive test performance: a validation study. *Psychopharmacology, 235*(4), 1283-1294.
Jung, C. M., Ronda, J. M., Czeisler, C. A., & Wright Jr, K. P. (2011). Comparison of sustained attention assessed by auditory and visual psychomotor vigilance tasks prior to and during sleep deprivation. *Journal of sleep research, 20*(2), 348-355.
Kang, H.-B. (2013). *Various approaches for driver and driving behavior monitoring: A review.* Paper presented at the Proceedings of the IEEE International Conference on Computer Vision Workshops.
Kuypers, K. P. C., Legrand, S.-A., Ramaekers, J. G., & Verstraete, A. G. (2012). A case-control study estimating accident risk for alcohol, medicines and illegal drugs.
Lal, S. K., & Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological psychology, 55*(3), 173-194.
Larue, G. S., Rakotonirainy, A., & Pettitt, A. N. (2011). Driving performance impairments due to hypovigilance on monotonous roads. *Accident Analysis & Prevention, 43*(6), 2037-2046.
Levy, J., & Pashler, H. (2008). Task prioritisation in multitasking during driving: Opportunity to abort a concurrent task does not insulate braking responses from dual-task slowing. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 22*(4), 507-525.
Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving. *Frontiers in Human Neuroscience, 13*(57). doi:10.3389/fnhum.2019.00057
Mahajan, K., & Velaga, N. R. (2020). Effects of partial sleep deprivation on braking response of drivers in hazard scenarios. *Accident Analysis & Prevention, 142*, 105545.
Mahmoodi, M., & Nahvi, A. (2019). Driver drowsiness detection based on classification of surface electromyography features in a driving simulator. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 233*(4), 395-406.
McDowd, J., Vercruyssen, M., & Birren, J. E. (2020). Aging, divided attention, and dual-task performance *Multiple-task performance* (pp. 387-414): CRC Press.
Moskowitz, H., Burns, M., Fiorentino, D., Smiley, A., & Zador, P. (2000). *Driver characteristics and impairment at various BACs*. Retrieved from
Musselwhite, C., & Haddad, H. (2010). Mobility, accessibility and quality of later life. *Quality in Ageing and Older Adults*.
Ogino, M., & Mitsukura, Y. (2018). Portable drowsiness detection through use of a prefrontal single-channel electroencephalogram. *Sensors, 18*(12), 4477.
Oken, B. S., Salinsky, M. C., & Elsas, S. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clinical neurophysiology, 117*(9), 1885-1901.
Owsley, C. (2002). Driving mobility, older adults, and quality of life. *Gerontechnology, 2*(1).
Papantoniou, P., Papadimitriou, E., & Yannis, G. (2017). Review of driving performance parameters critical for distracted driving research. *Transportation research procedia, 25*, 1796-1805.
Patomella, A.-H., Caneman, G., Kottorp, A., & Tham, K. (2004). Identifying scale and person response validity of a new assessment of driving ability. *Scandinavian journal of occupational therapy, 11*(2), 70-77.

Perrier, J., Jongen, S., Vuurman, E., Bocca, M., Ramaekers, J., & Vermeeren, A. (2016). Driving performance and EEG fluctuations during on-the-road driving following sleep deprivation. *Biological psychology, 121*, 1-11.

Perrier, J., Leufkens, T., Bulla, J., Jongen, S., Bocca, M., Ramaekers, J., & Vermeeren, A. (2015). Electroencephalography during on-the-road driving in older untreated insomnia patients and normal sleepers. *Biological psychology, 109*, 20-28.

Ramaekers, J. G. (2017). Drugs and driving research in medicinal drug development. *Trends in pharmacological sciences, 38*(4), 319-321.

Ramaekers, J. G., Uiterwijk, M. M., & O'hanlon, J. F. (1992). Effects of loratadine and cetirizine on actual driving and psychometric test performance, and EEG during driving. *European journal of clinical pharmacology, 42*(4), 363-369.

Schwalm, M., Keinath, A., & Zimmer, H. D. (2008). Pupillometry as a method for measuring mental workload within a simulated driving task. *Human Factors for assistance and automation, 1986*, 1-13.

Seli, P., Cheyne, J. A., Barton, K. R., & Smilek, D. (2012). Consistency of sustained attention across modalities: Comparing visual and auditory versions of the SART. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 66*(1), 44.

Shameen, Z., Yusoff, M., Saad, M., Malik, A., & Muzammel, M. (2018). Electroencephalography (EEG) based drowsiness detection for drivers: A review. *ARPN J. Eng. Appl. Sci, 13*(4), 1458-1464.

Shekari Soleimanloo, S., Wilkinson, V. E., Cori, J. M., Westlake, J., Stevens, B., Downey, L. A., . . . Howard, M. E. (2019). Eye-blink parameters detect on-road track-driving impairment following severe sleep deprivation. *Journal of clinical sleep medicine, 15*(9), 1271-1284.

Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews, 96*, 353-366.

Shiferaw, B. A., Crewther, D. P., & Downey, L. A. (2019). Gaze entropy measures detect alcohol-induced driver impairment. *Drug and alcohol dependence, 204*, 107519.

Stawarczyk, D., François, C., Wertz, J., & D'Argembeau, A. (2020). Drowsiness or mind-wandering? Fluctuations in ocular parameters during attentional lapses. *Biological psychology, 156*, 107950.

Teh, E., Jamson, S., Carsten, O., & Jamson, H. (2014). Temporal fluctuations in driving demand: The effect of traffic complexity on subjective measures of workload and driving performance. *Transportation research part F: traffic psychology and behaviour, 22*, 207-217.

Timmermans, I., Jongejan, N., Meine, M., Doevendans, P., Tuinenburg, A., & Versteeg, H. (2018). Decreased quality of life due to driving restrictions after cardioverter defibrillator implantation. *Journal of Cardiovascular Nursing, 33*(5), 474-480.

Tipping, M. E. (2003). *Bayesian inference: An introduction to principles and practice in machine learning.* Paper presented at the Summer School on Machine Learning.

van der Sluiszen, N., Vermeeren, A., Jongen, S., Theunissen, E., van Oers, A., Van Leeuwen, C., . . . Ramaekers, J. (2016). On-the-road driving performance after use of the antihistamines mequitazine and l-mequitazine, alone and with alcohol. *Psychopharmacology, 233*(18), 3461-3469.

Vaportzis, E., Georgiou-Karistianis, N., & Stout, J. C. (2013). Dual task performance in normal aging: a comparison of choice reaction time tasks. *PLoS one, 8*(3), e60265.

Vermeeren, A., Jongen, S., Murphy, P., Moline, M., Filippov, G., Pinner, K., . . . Van Oers, A. C. (2019). On-the-road driving performance the morning after bedtime administration of lemborexant in healthy adult and elderly volunteers. *Sleep, 42*(4), zsy260.

Vermeeren, A., Riedel, W. J., van Boxtel, M. P., Darwish, M., Paty, I., & Patat, A. (2002). Differential residual effects of zaleplon and zopiclone on actual driving: a comparison with a low dose of alcohol. *Sleep, 25*(2), 224-231.

Vermeeren, A., Sun, H., Vuurman, E. F., Jongen, S., Van Leeuwen, C. J., Van Oers, A. C., . . . Heirman, I. (2015). On-the-road driving performance the morning after bedtime use of suvorexant 20 and 40 mg: a study in non-elderly healthy volunteers. *Sleep, 38*(11), 1803-1813.

Vermeeren, A., Vets, E., Vuurman, E. F., Van Oers, A. C., Jongen, S., Laethem, T., . . . Li, X. (2016). On-the-road driving performance the morning after bedtime use of suvorexant 15 and 30 mg in healthy elderly. *Psychopharmacology, 233*(18), 3341-3351.

Vermeeren, A., Vuurman, E. F., Leufkens, T. R., Van Leeuwen, C. J., Van Oers, A. C., Laska, E., . . . Roth, T. (2014). Residual effects of low-dose sublingual zolpidem on highway driving performance the morning after middle-of-the-night use. *Sleep, 37*(3), 489-496.

Verster, Warren Spence, D., Shahid, A., R Pandi-Perumal, S., & Roth, T. (2011). Zopiclone as positive control in studies examining the residual effects of hypnotic drugs on driving ability. *Current Drug Safety, 6*(4), 209-218.

Verster, J. C., Bervoets, A. C., de Klerk, S., & Roth, T. (2014). Lapses of attention as outcome measure of the on-the-road driving test. *Psychopharmacology, 231*(1), 283-292.

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.

Verster, J. C., & Roth, T. (2014). Excursions out-of-lane versus standard deviation of lateral position as outcome measure of the on-the-road driving test. *Human Psychopharmacology: Clinical and Experimental, 29*(4), 322-329.

Verster, J. C., Veldhuijzen, D. S., Patat, A., Olivier, B., & Volkerts, E. R. (2006). Hypnotics and driving safety: meta-analyses of randomized controlled trials applying the on-the-road driving test. *Current Drug Safety, 1*(1), 63-71.

Vinckenbosch, F., Vermeeren, A., Verster, J., Ramaekers, J., & Vuurman, E. (2020). Validating lane drifts as a predictive measure of drug or sleepiness induced driving impairment. *Psychopharmacology, 237*(3), 877-886.

Wester, A. E., Verster, J. C., Volkerts, E. R., Böcker, K. B., & Kenemans, J. L. (2010). Effects of alcohol on attention orienting and dual-task performance during simulated driving: An event-related potential study. *Journal of psychopharmacology, 24*(9), 1333-1348.

Wolkow, A. P., Rajaratnam, S. M., Wilkinson, V., Shee, D., Baker, A., Lillington, T., . . . Tucker, A. (2020). The impact of heart rate-based drowsiness monitoring on adverse driving events in heavy vehicle drivers under naturalistic conditions. *Sleep health, 6*(3), 366-373.

Yadav, A. K., & Velaga, N. R. (2019). Effect of alcohol use on accelerating and braking behaviors of drivers. *Traffic injury prevention, 20*(4), 353-358.

Zandi, A. S., Quddus, A., Prest, L., & Comeau, F. J. (2019). Non-intrusive detection of drowsy driving based on eye tracking data. *Transportation research record, 2673*(6), 247-257.

Zhang, X.-D. (2020). Machine learning *A Matrix Algebra Approach to Artificial Intelligence* (pp. 223-440): Springer.

Summary

Many medical conditions and medicinal treatments are known to potentially impair driving performance and/or are associated with increased traffic accident risk. Researchers, clinicians, and policy makers are burdened with the task of deciding which patients are capable of operating a vehicle in traffic safely and which aren't. Keeping impaired drivers of the road is crucial for managing traffic safety. However, preventing unnecessary denial of the right to drive is important for protecting individual patients' mobility and independence, which is important for quality of life.

Clinical advice and legal policies are generally based on empirical findings at the group level. However, clinical populations are often considerably heterogeneous regarding types and severity of complaints as well as with respect to effectiveness and tolerability of (medicinal) treatments and their potential side effects. Hence, applying group level findings to individual patients will arguably lead to a considerable number of cases who are falsely identified as unimpaired, which is undesirable for traffic safety, or as impaired, and subsequently unfairly denied their right to drive. To prevent this from happening, it is argued that efforts should be invested in constructing adequate assessment tools for determining whether an individual driver is impaired or not in order to complement group-level findings.

In **chapter 2**, common research methods in the field of traffic medicine are introduced. Various epidemiological and experimental study designs are discussed, as well as their respective strengths and weaknesses. In a nutshell, it can be stated that a tradeoff exists between epidemiological and experimental studies. The advantage of (observational) epidemiological studies is that they are directly concerned with the outcome of interest, i.e. traffic accidents. Yet, this direct consideration of actual crash risk comes at the cost of not being able to make a strong case for causality of the factors under investigation due to the observational (associative) nature of the study designs. In contrast, experimental studies are well suited for demonstrating causality, yet are unable to directly assess real-world crash risk. Instead, experimental studies are limited to assessing driving performance as a presumed predictor of crash risk. Various laboratory and on-road methods for

assessing driving performance are discussed. Laboratory measures have the advantage of being highly controlled, convenient, and fit for a detailed assessment of different aspects of driving related functions. The main disadvantage is that it is often uncertain how performance on laboratory tests translates to real-world driving performance. In contrast, on-road driving test have a high degree of ecological validity, but are subject to the unpredictability of real-world traffic and are less practically convenient to execute.

One of the introduced on-road methods for assessing driving performance in **chapter 2** is the standardized on-the-road driving test, which is at the center of this dissertation because it is the cornerstone of driving performance assessment at Maastricht University. The test comprises a 100km drive on a public motorway while maintaining a steady velocity of approximately 95km/h. Participants are instructed to stay in the right traffic lane(s), and to only deviate from these instructions for overtaking slower traffic or because of other pressing traffic circumstances. A licensed driving instructor who has access to dual controls accompanies the participant to ensure safety. During the test, a camera on top of the test vehicle continuously monitors the distance to the lane demarcation on the left of the vehicle. The standard deviation of the lateral position (SDLP) is the main outcome measure of the test. It is conceived as a measure of road-tracking control or *lane weaving*. Higher SDLP is thought to reflect poorer road-tracking performance, and therefore also poorer driving performance. A central notion in this dissertation is that the SDLP has a wide normative range which limits the potential to make meaningful inferences about driving performance based on a single SDLP value. In contrast, repeated standardized-on-the road driving tests performed by the same drivers generally yield very similar SDLP values, indicating high reliability of the measure. Therefore, the standardized on-the-road driving test and the SDLP are commonly applied in experimental crossover designs, where the central focus is on the change of the SDLP across the experimental conditions with each driver functioning as their own reference.

One example of such a study is the randomized placebo-controlled, double-blind crossover trial on the effects of the noradrenaline and dopamine reuptake inhibitor solriamfetol on driving

performance of sleep apnea patients suffering from excessive daytime sleepiness, presented in **chapter 3** of this dissertation. In this two-period crossover study, 34 participants received a daily morning dose of solriamfetol (150mg for 3 days, then 300mg for 3 days) or placebo for 7-days before switching over to a 7-day treatment with placebo or solriamfetol respectively. In each 7-day treatment period, two standardized on-the-road driving tests were performed by the patient at 2h and 6h post dose (administration at 8:45am). The SDLP at both time points was compared across the experimental conditions. It was found that SDLP was significantly lower in the solriamfetol condition at both 2h and 6h post dose. The results suggest a statistically significant improvement of road-tracking performance of sleep apnea patients suffering from excessive daytime sleepiness after treatment with solriamfetol that lasts until at least 6h after morning administration. A comparison of the mean SDLP in both conditions with that of a normative sample suggests that the mean SDLP was significantly higher than that of a normative sample in the placebo condition. In the solriamfetol condition, the mean SDLP was comparable to that of the normative sample at 2h post dose, but not at 6h post dose.

A near-identical study on the effects of solriamfetol on driving performance in 24 narcolepsy patients is presented in **chapter 4** of this dissertation. A statistically significant decrease in median SDLP was observed after solriamfetol compared to placebo at 2h post dose. At 6 hours post dose, no significant difference was observed between the experimental conditions regarding SDLP, although the higher number of prematurely terminated driving tests in the placebo condition, primarily at the request of the participants, does suggest that solriamfetol decreased subjective drowsiness in the placebo condition. Comparison to normative data revealed that the mean and median SDLP in the placebo condition fell outside the normative range. At 2h post dose, mean and median SDLP fell within the normative range, but not anymore at 6h post dose. This suggests that solriamfetol has a beneficial effect on road-tracking performance during the standardized on-the-road driving test. However, this effect appears to have dissipated at 6h post dosing.

In **chapter 5,** a cross-sectional study is discussed that focusses on the neurocognitive functioning and driving performance during the standardized on-the-road driving test of long-term users of benzodiazepine receptor agonists (BZRAs), a class of central nervous system depressants. The investigation highlights and attempts to address the problem of heterogeneity regarding clinical complaints and medicinal treatments in real-world clinical populations. The sample of 50 long-term BZRA users collectively used 13 different BZRAs, sometimes combined, at varying doses, frequencies and time of day. Patient participants also differed with respect to severity of clinical complaints and had a wide age range (21–75 years). An attempt was made to determine which participants were likely to have clinically relevant BZRA plasma concentrations, both on average as well as at the start of the administered tests, based on their reported and prescribed medication use. It was determined that 31 BZRA users were considered unlikely to have clinically relevant BZRA plasma at the start of testing and on average (C_{BZRA}-) , while 19 BZRA users were (C_{BZRA}+) . Next, it was investigated whether the BZRA users exhibited poorer neurocognitive and driving performance compared to 76 healthy controls, corrected for BZRA plasma status (C_{BZRA}- or C_{BZRA}+), age, and/or clinical complaint severity. It was found that BZRA users performed significantly worse than healthy controls on tasks of response speed, processing speed, and sustained attention. Clinical complaint severity and BZRA plasma status did not appear to contribute significantly to this difference. For driving performance during the standardized on-the-road driving test, operationalized as the SDLP, it was found that BZRA users with clinically relevant estimated BZRA concentrations exhibited significantly poorer road-tacking performance compared to healthy controls. The magnitude of the differences was comparable to that seen at blood alcohol concentrations of 0.05g/L, which suggests that this patient subgroup might exhibit clinically meaningful driving impairment. The subgroup of BZRA users with estimated plasma concentrations that are not considered clinically relevant did not show any difference to healthy controls on the SDLP.

The sample in **chapter 5** illustrates the discrepancy of a real-world clinical samples with the highly selected samples of clinical trials, such as the ones discussed in **chapters 3 and 4**. Real-world

clinical samples are arguably much more heterogeneous regarding patient characteristics, clinical complaints and comorbidities, and medication use. This notion prompted the question to what extent findings from highly controlled clinical trials can be generalized to the respective clinical population, but even more so, to what extent these findings relate to the individual patients that make up that clinical population. The boxplots in **chapter 5** demonstrate that the clinical and control samples overlap a great deal with regard to the various performance measures. This raises the suspicion that within the heterogeneous real-world clinical populations, a significant portion of patients will often exhibit functioning within the normative range. Identifying which individual patients demonstrate impaired functioning and which don't is important for increasing the understanding of the relevant predictors and mechanisms that underlie impairment in the respective population, it can also allow for fair decision making regarding the ability to perform certain tasks such as driving a vehicle in traffic. Therefore, this dissertation shifted focus to individualized assessments of driving performance using the standardized on-the-road driving test. It has already been posed that the main outcome measure of the standardized on-the-road driving test, the SDLP, is not ideal for individual level assessments due to its wide normative range and other conceptual limitations outlined in **chapter 8**. Hence, it is asserted that additional outcome measures are needed to equip the standardized on-the-road driving test for individual assessments of driving performance.

In **chapter 6**, one such candidate for additional outcome measure was considered. Previously, it was proposed that large and sudden lateral displacements of the test vehicle, i.e. lane drifts, during the standardized on-the-road driving test were indicative of attentional lapses and therefore of impaired driving performance. The statistical characteristics of this measure were explored. The deviations were, as previously proposed, operationalized as deviation in the lateral position of 100cm within an 8 second window, or alternatively, for a duration of at least 8 seconds. It was found that the latter operationalization occurred very rarely. The former operationalization was found to occur more frequently. However, it was concluded that it was a mere transformation of the

SDLP with inferior sensitivity. Lane drifts were therefore concluded to provide little added value as an outcome measure of the standardized on-the-road driving test.

Chapter 7 again focused on the SDLP as an outcome for individual assessments. Although the wide normative range of the SDLP limits the interpretability of a single absolute SDLP value obtained during a single test repetition, SDLP change across repeated tests demonstrates high test-retest reliability and is therefore more informative. Hence, the presented analysis set out to determine cutoffs for statistically significant change in SDLP within an individual driver across repeated tests. Two statistical approaches were considered and applied to a historical dataset containing data 98 pairs of standardized on-the-road driving tests completed by healthy drivers in the placebo condition of the respective study. From this, it was determined that the cutoff for statistically significant change in SDLP across drives completed by the same driver lies as 1.8cm for a type I error rate of 5%. Various confidence interval widths are reported, as well as the findings of one of the two statistical techniques which did not yield straightforward cutoffs and is therefore not described in this summary.

Finally, in **chapter 8**, a more in depth consideration of the limitations of the SDLP as an outcome measure for individualized assessments of driving performance during the standardized on-the-road driving test is presented. The need for additional outcome measures is revisited. It is proposed that additional outcome measures for the on-the-road driving test should include both driver and vehicle based measures. This assures that presumed driver impairment, as observed in driver based measures, is accompanied by functional driving impairment, as observed in vehicle based measures, and vice versa. The potential role for machine learning techniques to integrate multiple driver and vehicle based parameters is discussed. Also, the potential for the statistical methods discussed in chapter 7 in establishing ground truth for machine learning approaches is proposed.

Samenvatting

Van veel medische aandoeningen en medicamenteuze behandelingen is bekend dat ze de rijvaardigheid kunnen aantasten en/of in verband kunnen worden gebracht met een verhoogd risico op verkeersongevallen. Onderzoekers, clinici en beleidsmakers worden belast met de taak om te beslissen welke patiënten in staat zijn om veilig een voertuig te besturen in het verkeer en welke niet. Het van de weg houden van bestuurders met verminderde rijvaardigheid is van cruciaal belang om de verkeersveiligheid in goede banen te leiden. Voorkomen dat het recht om een voertuig te besturen onnodig wordt ontzegd, is echter belangrijk voor de bescherming van de mobiliteit en onafhankelijkheid van individuele patiënten, wat belangrijk is voor de levenskwaliteit.

Klinische adviezen en juridisch beleid zijn over het algemeen gebaseerd op empirische bevindingen op groepsniveau. Klinische populaties zijn echter vaak zeer heterogeen wat betreft de aard en ernst van de klachten en wat betreft de effectiviteit en verdraagbaarheid van (medicamenteuze) behandelingen en de mogelijke bijwerkingen daarvan. De toepassing van bevindingen op groepsniveau op individuele patiënten zal er dan ook waarschijnlijk toe leiden dat een aanzienlijk aantal gevallen ten onrechte als geschikt, wat onwenselijk is voor de verkeersveiligheid, of als ongeschikt wordt geïdentificeerd, en hun vervolgens ten onrechte de rijbevoegdheid wordt ontzegd. Om dit te voorkomen wordt er in dit proefschrift betoogd dat er moet worden geïnvesteerd in de ontwikkeling van adequate beoordelingsinstrumenten om vast te stellen of een individuele bestuurder al dan niet een rijbeperking heeft, als aanvulling op de bevindingen op groepsniveau.

In **hoofdstuk 2** worden gangbare onderzoeksmethoden op het gebied van de verkeersgeneeskunde geïntroduceerd. Verschillende epidemiologische en experimentele onderzoeksopzetten worden besproken, evenals hun respectievelijke sterke en zwakke punten. Kort samengevat kan gesteld worden dat er een trade-off bestaat tussen epidemiologische en experimentele studies. Het voordeel van (observationele) epidemiologische studies is dat ze direct

betrekking hebben op de uitkomst van belang, namelijk verkeersongevallen. Deze directe beschouwing van het feitelijke ongevalsrisico gaat echter ten koste van het aantonen van de causaliteit van de onderzochte factoren vanwege het observationele (associatieve) karakter van de studieopzetten. Experimentele studies daarentegen zijn zeer geschikt om causaliteit aan te tonen, maar zijn niet in staat om het werkelijke risico op ongevallen rechtstreeks te beoordelen. In plaats daarvan beperken experimentele studies zich tot het beoordelen van de rijprestaties als een veronderstelde voorspeller van het ongevalsrisico. Verschillende laboratorium- en verkeersmethoden om de rijvaardigheid te beoordelen worden besproken. Laboratoriummetingen hebben het voordeel dat ze zeer goed gecontroleerd kunnen worden, dat ze praktisch zijn en dat ze geschikt zijn voor een gedetailleerde beoordeling van verschillende aspecten van rijgerelateerde functies. Het belangrijkste nadeel is dat het vaak niet zeker is hoe de prestaties op laboratoriumtests zich vertalen in rijprestaties in de praktijk. Rijtests op de weg daarentegen hebben een hoge mate van ecologische validiteit, maar zijn onderhevig aan de onvoorspelbaarheid van het verkeer in de echte wereld en zijn in de praktijk minder gemakkelijk uit te voeren.

Één van de in **hoofdstuk 2** geïntroduceerde methoden voor het beoordelen van rijprestaties op de weg is de gestandaardiseerde rijtest op de weg, die centraal staat in dit proefschrift omdat het de hoeksteen is van de beoordeling van rijprestaties aan de Universiteit Maastricht. De test bestaat uit een rit van 100 km op een openbare snelweg met een constante snelheid van ongeveer 95 km/u. Deelnemers worden geïnstrueerd om op de rechterrijstrook(ken) te blijven, en alleen van deze instructies af te wijken voor het inhalen van langzamer verkeer of vanwege andere dringende verkeersomstandigheden. Een gediplomeerde rijinstructeur die toegang heeft tot de dubbele besturing begeleidt de deelnemer om de veiligheid te garanderen. Tijdens de test controleert een camera bovenop het testvoertuig voortdurend de afstand tot de rijstrookafbakening aan de linkerkant van het voertuig. De standaardafwijking van de laterale positie (SDLP) is de belangrijkste uitkomstmaat van de test. Zij wordt beschouwd als een maatstaf voor de controle over het volgen van de rijstrook of voor slingergedrag. Een hogere SDLP zou wijzen op een slechtere wegligging en

dus ook op slechtere rijprestaties. Een centraal idee in dit proefschrift is dat de SDLP een breed normatief bereik heeft, wat de mogelijkheid beperkt om zinvolle conclusies te trekken over rijprestaties op basis van een enkele SDLP waarde. Daarentegen leveren herhaalde gestandaardiseerde rijtests op de weg, uitgevoerd door dezelfde bestuurders, over het algemeen zeer vergelijkbare SDLP waarden op, wat duidt op een hoge betrouwbaarheid van de maatstaf. Daarom worden de gestandaardiseerde rijtest op de weg en de SDLP vaak toegepast in experimentele cross-over ontwerpen, waarbij de nadruk ligt op de verandering van de SDLP tussen verschillende experimentele omstandigheden, waarbij elke bestuurder als zijn eigen referentie fungeert.

Een voorbeeld van zo'n studie is de gerandomiseerde placebo-gecontroleerde, dubbelblinde cross-over studie naar de effecten van de noradrenaline en dopamine heropname remmer solriamfetol op de rijvaardigheid van slaapapneu patiënten die lijden aan overmatige slaperigheid overdag, gepresenteerd in **hoofdstuk 3** van dit proefschrift. In deze twee-periodes cross-over studie, kregen 34 deelnemers een dagelijkse ochtenddosis van solriamfetol (150mg gedurende 3 dagen, daarna 300mg gedurende 3 dagen) of placebo gedurende 7 dagen alvorens over te schakelen naar een 7-daagse behandeling met respectievelijk placebo of solriamfetol. In elke 7-daagse behandelingsperiode werden twee gestandaardiseerde rijtesten op de weg uitgevoerd door de patiënt om 2 uur en 6 uur na de dosis (toediening om 8.45 uur). De SDLP op beide tijdstippen werd vergeleken tussen de experimentele condities. Er werd vastgesteld dat de SDLP significant lager was in de solriamfetol conditie, zowel 2u als 6u na de dosis. De resultaten suggereren een statistisch significante verbetering van de wegligging van slaapapneupatiënten die lijden aan overmatige slaperigheid overdag na behandeling met solriamfetol, die aanhoudt tot minstens 6 uur na de ochtendtoediening. Een vergelijking van de gemiddelde SDLP in beide condities met die van een normgroep suggereert dat de gemiddelde SDLP in de placebo conditie significant hoger was dan die van een normgroep. In de solriamfetol conditie was de gemiddelde SDLP vergelijkbaar met die van het normatieve monster op 2 uur na de dosis, maar niet op 6 uur na de dosis.

Een bijna identieke studie naar de effecten van solriamfetol op de rijvaardigheid bij 24 narcolepsiepatiënten wordt gepresenteerd in **hoofdstuk 4** van dit proefschrift. Een statistisch significante afname in mediane SDLP werd waargenomen na solriamfetol vergeleken met placebo, 2 uur na de dosis. Na 6 uur werd er geen significant verschil waargenomen tussen de experimentele condities met betrekking tot SDLP, hoewel het hogere aantal voortijdig afgebroken rijtesten in de placebo-conditie, voornamelijk op verzoek van de deelnemers, suggereert dat solriamfetol de subjectieve slaperigheid in de placebo conditie verminderde. Vergelijking met normgroep toonde aan dat de gemiddelde en mediane SDLP in de placebo conditie buiten het normatieve bereik vielen. Na 2 uur na de dosis vielen de gemiddelde en mediane SDLP binnen het normatieve bereik, maar niet meer na 6 uur na de dosis. Dit suggereert dat solriamfetol een gunstig effect heeft op de wegligging tijdens de gestandaardiseerde on-the-road rijtest. Dit effect lijkt echter te zijn verdwenen 6 uur na de dosering.

In **hoofdstuk 5** wordt een cross-sectionele studie besproken die zich richt op het neurocognitief functioneren en de rijprestaties tijdens de gestandaardiseerde rijtest op de weg van langdurige gebruikers van benzodiazepine receptor agonisten (BZRAs), een groep van verdovende medicijnen. Het onderzoek belicht en tracht het probleem van de heterogeniteit met betrekking tot klinische klachten en medicamenteuze behandelingen in klinische populaties in de echte wereld aan te pakken. De steekproef van 50 langdurige BZRA-gebruikers gebruikte gezamenlijk 13 verschillende BZRAs, soms gecombineerd, in verschillende doses, frequenties en op verschillende tijdstippen van de dag. De deelnemers verschilden ook wat betreft de ernst van de klinische klachten en hadden een brede leeftijdsgroep (21-75 jaar). Er werd getracht te bepalen welke deelnemers waarschijnlijk klinisch relevante BZRA plasmaconcentraties hadden, zowel gemiddeld als bij aanvang van de toegediende tests, op basis van hun gerapporteerde en voorgeschreven medicatiegebruik. Er werd besloten dat het onwaarschijnlijk werd geacht dat 31 BZRA-gebruikers bij het begin van de tests klinisch relevante BZRA-plasma waarden hadden (CBZRA-) , terwijl 19 BZRA-gebruikers dat wel verondersteld werden te hebben (CBZRA+) . Vervolgens werd onderzocht of de BZRA gebruikers

slechtere neurocognitieve en rijprestaties vertoonden in vergelijking met 76 gezonde controles, gecorrigeerd voor BZRA plasma status (CBZRA- of CBZRA+), leeftijd, en/of ernst van de klinische klachten. Het bleek dat BZRA-gebruikers significant slechter presteerden dan gezonde controles op taken van reactiesnelheid, verwerkingssnelheid, en volgehouden aandacht. De ernst van de klinische klachten en de BZRA plasma status bleken niet significant bij te dragen aan dit verschil. Voor de rijvaardigheid tijdens de gestandaardiseerde rijtest op de weg, geoperationaliseerd als de SDLP, werd gevonden dat BZRA-gebruikers met klinisch relevante geschatte BZRA-concentraties significant slechtere rijvaardigheid vertoonden in vergelijking met gezonde controles. De omvang van de verschillen was vergelijkbaar met die bij bloedalcoholconcentraties van 0.05g/L (wettelijke limiet in o.a. Nederland en België), wat suggereert dat deze patiënten subgroep een klinisch betekenisvolle rijbeperking zou kunnen vertonen. De subgroep van BZRA-gebruikers met geschatte plasmaconcentraties die niet als klinisch relevant worden beschouwd, vertoonde geen significant verschil met gezonde controles op de SDLP.

De steekproef in **hoofdstuk 5** illustreert de discrepantie van klinische steekproeven in de echte wereld met de sterk geselecteerde steekproeven van klinische studies, zoals die besproken in de **hoofdstukken 3 en 4**. Klinische steekproeven uit de echte wereld zijn opmerkelijk heterogener wat betreft patiëntkarakteristieken, klinische klachten en comorbiditeit, en medicatiegebruik. Deze notie leidde tot de vraag in hoeverre bevindingen uit zeer gecontroleerde klinische trials kunnen worden gegeneraliseerd naar de betreffende klinische populatie, maar meer nog, in hoeverre deze bevindingen betrekking hebben op de individuele patiënten die deel uitmaken van die klinische populatie. De boxplots in **hoofdstuk 5** laten zien dat de klinische en controlegroepen elkaar sterk overlappen met betrekking tot de verschillende prestatie-indicatoren. Dit wekt het vermoeden dat binnen de heterogene klinische populaties in de echte wereld, een aanzienlijk deel van de patiënten vaak binnen het normatieve bereik zal functioneren. Daarom werd in dit proefschrift de aandacht verlegd naar geïndividualiseerde beoordelingen van rijprestaties m.b.v. de gestandaardiseerde rijtest op de weg. Er is reeds gesteld dat de belangrijkste uitkomstmaat van de gestandaardiseerde rijtest

op de weg, de SDLP, niet ideaal is voor beoordelingen op individueel niveau, vanwege het brede normatieve bereik en andere conceptuele beperkingen die in **hoofdstuk 8** worden beschreven. Daarom wordt gesteld dat aanvullende uitkomstmaten nodig zijn om de gestandaardiseerde rijtest op de weg uit te rusten voor individuele beoordelingen van rijprestaties.

In **hoofdstuk 6** werd een dergelijke kandidaat voor een aanvullende uitkomstmaat in overweging genomen. In eerdere publicaties werd voorgesteld dat grote en plotselinge zijdelingse verplaatsingen van het testvoertuig, d.w.z. *lane drifts*, tijdens de gestandaardiseerde rijtest op de weg indicatief waren voor aandachtsverslappingen en dus voor verminderde rijvaardigheid. De statistische kenmerken van deze maat werden onderzocht. De afwijkingen werden, zoals eerder voorgesteld, geoperationaliseerd als een afwijking in de laterale positie van 100 cm binnen een tijdspanne van 8 seconden, of als alternatief, voor een duur van ten minste 8 seconden. Het bleek dat de laatste manier van operationaliseren zeer zelden voorkwam. De eerste manier van operationaliseren bleek vaker voor te komen. Er werd echter geconcludeerd dat de maat slechts een transformatie van de SDLP is met een inferieure gevoeligheid. Er werd daarom geconcludeerd dat *lane drifts* weinig toegevoegde waarde hebben als uitkomstmaat van de gestandaardiseerde rijtest op de weg.

Hoofdstuk 7 richt zich opnieuw op de SDLP als uitkomstmaat voor individuele beoordelingen. Hoewel het brede normatieve bereik van de SDLP de interpreteerbaarheid van een enkele absolute SDLP waarde verkregen tijdens een enkele testherhaling beperkt, bezit SDLP een hoge test-hertest betrouwbaarheid en is daarom meer informatief. De gepresenteerde analyse is dan ook bedoeld om grenswaarden vast te stellen voor statistisch significante verandering in SDLP binnen een individuele bestuurder bij herhaalde tests. Twee statistische benaderingen werden overwogen en toegepast op een historische dataset met gegevens van 98 paren gestandaardiseerde rijtests op de weg die werden afgelegd door gezonde bestuurders in de placeboconditie van de respectieve studie. Op basis hiervan werd bepaald dat de cut-off voor statistisch significante verandering in SDLP tussen door dezelfde bestuurder afgelegde ritten ligt bij 1,8 cm voor een type I-foutenpercentage van 5%.

Verschillende betrouwbaarheidsintervalbreedtes worden gerapporteerd, evenals de bevindingen van een van de twee statistische technieken die geen eenduidige cutoffs opleverde en daarom niet in deze samenvatting wordt beschreven.

Tenslotte wordt in **hoofdstuk 8** een meer diepgaande beschouwing gegeven van de beperkingen van de SDLP als een uitkomstmaat voor geïndividualiseerde beoordelingen van rijprestaties tijdens de gestandaardiseerde rijtest op de weg. De behoefte aan aanvullende uitkomstmaten wordt opnieuw bekeken. Er wordt voorgesteld dat aanvullende uitkomstmaten voor de rijtest op de weg zowel op de bestuurder als op het voertuig betrekking moeten hebben. Dit garandeert dat de veronderstelde stoornis van de rijvaardigheid, zoals die wordt waargenomen bij de metingen voor de bestuurder, gepaard gaat met een functionele stoornis van de rijvaardigheid, zoals die wordt waargenomen bij de metingen voor het voertuig, en vice versa. De potentiële rol van technieken voor machinaal leren voor de integratie van meerdere parameters op basis van de bestuurder en het voertuig wordt besproken. Ook wordt het potentieel voorgesteld van de in **hoofdstuk 7** besproken statistische methoden voor het vaststellen van de grondwaarheid voor *machine learning* benaderingen.

Impact paragraph

In this addendum the societal, clinical, and scientific relevance of the work outlined in this dissertation is discussed.

Societal relevance

Chapter 2 of this dissertation, *driving aptitude and fitness to drive*, was written to be part of the next edition of *Handbook of Forensic medicine* edited by Madea Burkhard. This book is intended as an international compendium of forensic medicine which also covers questions regarding fitness to drive and driver intoxication within the context of the legal domain. The chapter opens the book's section on traffic medicine and therefore provides an introductory overview of the most common methodological approaches to research within this field. In addition, the chapter touches upon considerations regarding retrospectively and prospectively assessing individuals' driving fitness and aptitude that are relevant to the forensic expert. As such, the chapter serves as an orienting introduction to the field of traffic medicine for students and forensic experts, or anyone who takes a personal or professional interest in the field of traffic medicine.

The clinical trials presented in chapters 3 and 4 are an indispensable step in the process for demonstrating a novel medicinal drugs' therapeutic potential in order to get regulatory approval, and subsequently getting it to the patients who could benefit from its use. The presented studies provide a proof-of-principle that the novel pharmaceutical stimulant drug *solriamfetol* has the potential to improve driving performance or narcolepsy and obstructive sleep apnea patients with excessive daytime sleepiness. This proof-of-principle will ultimately contribute to obtaining regulatory approval for the use of solriamfetol in these clinical populations so that patients might benefit from its availability and use.

The analysis presented in chapter 5 does not provide a straightforward conclusion on the causes of observed driving and neurocognitive impairments in long-term users of benzodiazepine receptor agonists. However, the elaborate description of the sample characteristics regarding

complaints and medication use underscores that a one-size-fits-all solution to regulating driving rights while using certain medications might be unjustified in the light of the apparent heterogeneity. In the Netherlands, currently it is prohibited to drive while using (most) benzodiazepine receptor agonists. Policy makers could take from chapter 5 that a more fine grained approach is warranted in order to prevent the unjustified denial of driving rights of members belonging to this patient group.

Overall, the dissertation to a great extent focuses on the intricacies of assessing driving performance. The suitability of assessment tools can differ greatly depending on the question at hand. How the outcomes from these tools relate to real-world accident risk, which is presumed to be predicted by impaired driving performance, is often not firmly established. Hence, the general audience can conclude from this dissertation that the assessment of driving performance as a predictor of accident risk is not an exact science and should be subjected to sufficient scrutiny and debate before reaching unequivocal conclusions that impact legal and clinical policies.

Clinical relevance

Clinicians depend on past and continued research findings for formulating clinical advice and selecting appropriate treatment options. When it comes to the question of the compatibility of certain treatments or medical conditions and the act of driving, it is important to be able to properly value the research methods that yielded the respective findings. For this purpose, clinicians can turn to the book chapter presented in chapter 2 of this dissertation as an orienting reference in the process of interpreting the relevance and quality of research findings in the field of traffic medicine.

The clinical trials presented in chapters 3 and 4 have been presented orally during several international (online) conferences across the globe. The interest in the presented results by the attending clinicians attests to the direct relevance of such trials to clinical practice. The clinical trials demonstrate that the novel pharmaceutical drug solriamfetol has a generally beneficial effect on road-tracking performance during the standardized on-the-road driving test and might be a valid

treatment option to consider when prescribing a stimulant drug for tackling excessive daytime sleepiness associated with sleep apnea and narcolepsy.

Finally, the analysis of long-term benzodiazepine receptor agonist users presented in chapter 5 suggests a negative impact on cognitive and psychomotor performance of long-term benzodiazepine receptor agonist use, although a clear causal relationship remains to be demonstrated. Regardless, the chapter adds to prior studies with similar findings and serves as one of the many reminders for clinicians to maintain the advised short durations of benzodiazepine receptor agonist treatments or to consider alternatives altogether in order to prevent the potential negative consequences of long-term use.

Scientific relevance

A central perspective taken in this dissertation is the methodological viewpoint. Several methodological considerations and questions have been addressed. The analysis on long-term benzodiazepine receptor agonist users of chapter 5 includes two methodological considerations that might be relevant to other researchers. Firstly, the chapter exemplifies a way to deal with the inherent heterogeneity of clinical samples regarding medication use in observational studies through the implementation of pharmacokinetic parameters of the respective drug(s) and characteristics of use (e.g. frequency, dose, time since last use) in previously formulated pharmacokinetic equations for the estimation of blood plasma concentrations. Proposedly, such equations can be used in epidemiological studies that link accident database records to medication database records, or any other study that does not establish actual blood plasma concentrations through sampling. Secondly, the analysis in chapter 5 includes an exploratory factor analysis and subsequent factor extraction to reduce the number of dependent variables. It is proposed that regressing the outcome measures on a smaller number of latent factors is a useful way of reducing the multiple testing problem as an alternative to a multivariate analysis of variance procedure. In addition to reducing the multiple testing problem, exploratory factor analysis has the benefit of potentially providing new qualitative

insights in performance differences across conditions or between individuals based on the observed relationships between variables.

Chapter 6 of this dissertation comprises a fundamental step in scientific research, i.e. replication of findings. It was found that the frequency of occurrences of large lateral deviations during the standardized on-the-road driving test, which were proposed to reflect attentional lapses, could not be replicated in a larger historical dataset. Based on this finding, together with minor conceptual limitations, it was concluded that lane drifts are not a useful additional outcome for the standardized on-the-road driving test. Furthermore, it was concluded that it remains to be validated whether lapses of attention would result in sudden and large deviations in the lateral position. The development and subsequent validation of novel outcome measures are indispensable parts scientific progress.

Finally, the analysis of chapter 7 specifies the thresholds for statistically significant change in the main outcome of the standardized on-the-road driving test for individuals across conditions, i.e. the standard deviation of the lateral position. Up until now, such a threshold had not been established despite its relevance in frequently applied post-hoc McNemar statistical testing. Furthermore, the methods presented can arguably be applied to any continuous variable with sufficient test-retest reliability in order to establish statistically significant change over repeated measures at the individual level. Identifying which individuals exhibit statistically significant change on a primary outcome variable is potentially useful to elucidate the characteristics of those individuals in the sample that do show a significant response to the experimental condition of interest versus those who do not. These insights might yield valuable insights that could aid in clinical care and treatment development. Within the context of driving assessments, the methods described in chapter 7 might prove to be a useful asset for establishing ground truth in the development of machine learning based development of advanced driver monitoring systems, as discussed in chapter 8.

Acknowledgements

It was supposed to be a sprint, but it turned out to be a marathon. Not one in which I finish gracefully with my hands triumphantly raised to the heavens, but one in which I stumble across the finish line, severely dehydrated and with poop running down my leg. Nevertheless, I made it all the way to the end. I could preach something about the importance of perseverance, but honesty urges me to admit that the completion of my PhD has been the product of luck, sunk cost fallacy (and maybe some ego), and the support of many people who I would like to thank in these acknowledgements.

First of all, I would like to thank the assessment committee of this dissertation: Emeritus Professor Wim Riedel, Professor Dr. Tom Brijs, Professor Dr. Luke Downey, Dr. Eef Theunissen, and Dr. Fren Smulders. Thank you for reading through this dissertation and providing me the opportunity to address your questions and remarks during the defense. I am looking forward to receiving your feedback and discussing the contents of the work presented.

Back to the marathon analogy for a minute. Like a professional athlete, I had my own personal coaching team, consisting of my promotor, Prof. Dr. Jan Ramaekers, and my co-promotors, Dr. Annemiek Vermeeren and Dr. Eric Vuurman. This team turned out to be a blessing and a challenge. A blessing because I could draw from the collective experience of three seasoned academics, a challenge because reconciling three (often) different perspectives while including my own turned out to be quite a puzzle at times. However, this is the acknowledgement section where I will not be going on about the endless discussions about phrasing, protocol adjustments that once inserted needed to go again, strong arguments on whether my approach was exploratory or a fishing expedition, and many other topics. No, none of that. This is the part where I look back and acknowledge all the valuable lessons I have learned from you.

Jan, your go-with-the-flow attitude stands in stark contrast to that of the stereotypical stern academic, and serves as a reassuring example that one does not have to consider every little detail before making decisions and moving forward. In Belgium, we once had a prime minister who famously stated: “You only need to start solving problems when they present themselves”. It might as well have been a quote by Professor Ramaekers. I’m sure you disagree a bit, because foresight is indeed valuable to some extent, but you get the point: *Paralysis through analysis* is not your motto. This also shows in your laisser-faire supervising approach, which I greatly appreciated. My-way-or-the-highway authority tends to inspire me to take the highway, but not before turning the ”*my way*” into a classical Belgian road. Yet, I stuck around for almost 6 years (and some more after this) without sabotaging the drugs and driving research line, which shows that you made me feel like I had a decisive voice in deciding whatever turn to take in our projects. You have given me a lot of freedom to find my own way, even if that meant keeping me on board for two more years after my initial contract had ended. At the same time you have always showed involvement and willingness to think along with me when I walked through your door, which was always open (unlike your inbox, it seems). The responsibilities you gave me, and trust that I would faithfully execute these tasks have been of great importance in building my confidence as a researcher. I can imagine that I might not have finished my PhD if I would’ve ended up with a promotor without as much cool and patience as you. Thank you for consistently supporting me while letting me do my thing.

A priori (and post-hoc) consideration of every little detail of a research project is something that only superhumans and Annemiek can do. Annemiek, you are the textbook example of a true academic to me (not the stereotypical one I was alluding to above, though). Your relentless attention to detail has been a source of admiration and frustration to me. You have been the much needed break on my naïve optimism and ignorant (or arrogant) stubbornness. Thank you for providing me with the occasional, yet indispensable, reality checks, as well as your notorious exhaustive feedback on my writings. At the end of my career, I’m certain that I will still often think to myself: “What would Annemiek say?”. I hope that by then, I will have matched your academic drive and discipline, even to

a small extent. Your elective course on psychopharmacology provided me with a new and exciting way to look at the brain and behavior, and has been one of the main reasons why I applied for the research master in neuropsychology, which resulted in me ending up doing a PhD at our department. I can never give you enough credit for setting up such an interesting course, which ended up playing a very significant role in my professional development, as well as in that of many students after me, I'm sure. Thank you.

Eric, I have had the privilege of having you on my promotion team for the first few years, until you had to leave for an unexpected early retirement. Thank you for your valuable insights and direct way of giving feedback. I bet you never realized this, but you too have played a significant role in directing my professional development. I remember when you wrote a script to facilitate the processing of the data presented in chapter 6 of this dissertation. I admired greatly that this senior academic possessed the skills to write computer programs, and I, a digital native, had no clue on how to do such a thing. Since then I have completed several courses on programming in Python, R, and SQL, and have developed strong ambitions to include machine learning technology in future projects. You played a significant role in sparking this keen interest of mine, for which I'm very grateful.

My list of supervisors would not be complete without acknowledging my informal supervisor of the first hour. My journey at NP&PP started with my master's internship, for which the day-to-day supervision was largely delegated to then PhD-student Nick, now Dr. van der Sluiszen. Nick, thank you for taking me on board and teaching me the logistics of running a research project. I thoroughly enjoyed working together from my first day as an intern, as colleagues when I eventually started my PhD, up until the day you left our department. Your hard work has served as a starting point for my journey, and even ended up in chapter 5 of this dissertation (for which a *thank you* is also due to Joke van Dijken, Dr. Janet Veldstra, and Dr. Aurora van de Loo for digging through patient medication profiles, as well as the rest of the team from the University of Groningen and Utrecht University for setting up and carrying out the study). Even now, I still value the occasional sparring about drugs and

driving related topics (and many other topics). Not only do we share academic interests, it turns out we also share a (inappropriate/absurd) sense of humor, a taste for great beers and liquors, and a love for songs about mother's day. Hence, although your journey as a drugs and driving researcher has come to an end (for now), we have plenty in common to keep our friendship alive. I'm happy to have you by my side as a paranymph during the defense, and also as stand-in tall drunk dude at the reception afterwards.

To round of my list of (in)formal supervisors, I would like to thank Dr. Eef Theunissen. Eef, you were my mentor during my master's and I have been lucky to be able to work with you on some projects during my PhD and to be able to continue to do so during my post-doc research. I always felt welcome to run an idea or issue past you if my promotion team happened to be unavailable. I have always enjoyed working with you, not only because we share an accent, but because you are such an easy-going person. It's always a pleasure to get your input. I can only hope we'll have plenty more opportunities to work together in the future.

I would not have been able to pull of this PhD without the indispensable logistical and administrative support I received, for which credit is due to two people in particular. Located in the same office you can find the desks of the two Atlases of the department, together bearing the load of the world we call NP&PP: Anita and Annemie. Anita, thank you for all your help with running the Jazz studies. Despite Nick having shown me the ropes before officially starting my PhD, I still had much to learn. I could not have pulled it off without you. Annemie, your endless patience with my inability to deal with administrative issues is greatly appreciated, and your knowhow about filing paperwork has saved me years of time browsing through the seemingly randomly scattered documentation provided by the University. Together, you two are the backbone of the department, and I want to give you a special thank you for that.

In a thesis concerning experimental driving studies, a special acknowledgement of the driving instructor team is of course due. Irma, thank you for always thinking along when planning issue

number five thousand something popped up. Although this can occasionally get your goat, you always search for solutions to make things work out, which you usually succeed in. Thank you, without your complaisance, I would not have been able to complete the driving studies in this dissertation, as well as those not mentioned in this dissertation. Henk, thank you for your unquestionable professionalism. I always say to people to whom I explain what we do during on-road driving studies that I'm confident that you are able to drive our test vehicles from Maastricht to Paris sitting in the passenger seat. I suspect that's not even an exaggeration. Your excellent safety supervision over the years has been fundamental to continuing on-road driving testing at our department, and therefore also for me to be able to start and complete my PhD. My thanks of course extends to the whole instructor team whom I had the pleasure of working with during my PhD: Hans, Rogé, & Gaston. Thank you for your excellent supervision. The drugs and driving research line depends on it.

The attentive reader (you know, those who also read things other than the acknowledgements) might have noticed that my dissertation contains a fair share of statistics. Although statistics has grown on me a lot throughout my PhD, I still feel very much like a layperson when considering what approach to take. Luckily, there is Dr. Jan Schepers to run to when uncertainty hits so bad it needs to be quantified. Jan, thank you for your time in advising me in several statistical questions that popped up while writing this thesis. Thanks to you, I always know how (un)certain I actually am.

For the collaboration in the projects presented in chapters 3 and 4, I want to thank Dr. Jerryll Asin, Prof. Dr. Sebastiaan Overeem, Prof. Dr. Nico de Vries, Dr. Patty Vonk, Dr. Claire Donjacour, Prof. Dr. Gert Jan Lammers, Dr. Hennie Janssen, Dr. Grace Wang, Dr. Lawrence Carter, Dr. Dan Chen, Dr. Kefei Zhou, and Dr. Joseph Burgon for their help with setting up the studies, executing them, and publishing the findings. I also want to thank the team from Peloton Advantage for their support in writing and publishing the study reports. For these projects, I would like to give a special thanks to

my dear Colleague Dr. Nadia Hutten. When I started my PhD by working on these projects you were also put on the project as a research assistant. Thanks for the swift collaboration during this time. It makes me happy to know that you got the PhD spot you deserved in the end, and even finished before I did (but I guess nearly everybody does.. but still!). It's always a pleasure working with you, or just chatting, and I hope we can do more of that in the future!

Also a special thanks to the IMOB at Hasselt University, and in particular Dr. Judith Urlings, Dr. Veerle Ross, and Prof. Dr. Tom Brijs. Although the study you helped me setup as a collaboration between our two institutions did not work out because of recruitment issues, I enjoyed our collaboration. I hope to explore new opportunities to collaborate in the future.

In addition, I would like to thank the team at the center of addiction medicine at the Massachusetts general hospital. The months I've spend working on the FNIRS project with you have been a wonderful experience. The questions put on my plate during my time in Boston have led to my exploration of statistical methods for identifying individual participants who demonstrate significant impairment or improvement. This has been the basis for the exercise I present in chapter 7 of this dissertation, and will surely be elaborated upon in the future. I look forward to analyzing and reporting on the data collected in the FNIRS project together soon (and perhaps share some more food and drinks, I'll make some more waffles!).

During the rough journey that is completing a PhD, a group of awesome colleagues is a great comfort. In this respect, a special role is reserved for those colleagues who you share an office with. My office crew by faith consists of (clockwise from my point of view sitting at my desk) Lili (can't really see you though), Jessica, and Natasha, now also known as Dr. Kloft, Dr. Bruijel, and Dr. Mason, respectively. I remember vividly at the start of our PhDs that Natasha, whom had also been part of my master's cohort, said that I was a prime example of what she believed a PhD student should be like. Admittedly, I'm a pretty smooth talker and enthusiastic presenter, but that's about it. Almost six years later, I think it's safe to say I have been outclassed by each and everyone of my office mates,

and not only because I'm the last to finish my PhD. I watched in awe how Natasha learned how to run fMRI studies and analyses, although we got very little background on this during our masters (enough to know that it's really complicated, though), while I struggled with logging into Qualtrics. I saw Lili coming into the office three months after the rest of us, but nearly finishing her PhD first (just a few weeks after Natasha, but still in record time) without making any concessions regarding quality (a well-deserved cum laude!). I saw Jessica persistently pull off what I have failed to do in my own work and what should be considered an higher art form: successfully running patient studies (the patients in my studies were not recruited by myself, because that's the hardest part). But most of all, you guys did all of that in silent and contained dedication, while I have spent years swearing at my computer screen when MS Excel was once again making invalid assumptions (what's the logical next step in this sequence: 1-2-3-4-1-2-3-4-? You guessed it: 3,571429), or when some other minor inconvenience occurred. Thank you guys for bearing with me and my visible frustrations. Thank you for letting me vent. Thank you for the many laughs, interesting conversations, food, drinks, and for making me feel relevant by asking me stats questions. I could not have wished for a lovelier group of people to share our office (and office plants) with.

Outside our office, there have been many more colleagues who make working on a PhD a lot easier. This part was the hardest to write because all of you deserve credit and I fear that I might accidentally omit one of you. Also, I already wrote a thesis and I don't have time to write another one to thank everybody individually. I would if I could though! Instead, I'll just buy you guys a bunch of drinks and give a general, but genuine, thank you. I really hope you know I mean it. So, this one goes out to all my colleagues, current and past ones. It's the countless chats at the coffee machine, jokes in the hallways, the great atmosphere at department gatherings, conference visits and accompanying leisure activities (mostly drinking, sometimes hiking), the occasional escaping from rooms, competitive beer pong in some rented house, experimenting with new things (for research purposes of course), some karaoke left and right, the many hilariously goofy speeches (we all know this is about Arjan), or the extremely energetic ones (we all know this refers to Eri), some more drinking,

the occasional slightly burnt jokes and showing genuine interest in how things are going that make it a pleasure to be part of our department. Thank you so much for making my Maastricht experience a fantastic first chapter in my career.

I would explicitly like to NOT thank my friends back home (mentioning no names here, you know if I'm talking about you). Your collective bad influence on my circadian rhythm and alcohol intake has just made this whole endeavor more challenging than it should have been. No, seriously guys, thanks for always keeping me in high regards. You guys really helped me remain confident that I could do this. Et voila!

To my dear parents, this day is also your achievement. Your consistent encouragement (and occasional veto during my adolescent years) has kept me on the path that lead me here today. Although I doubt that doing a PhD was professionally the best option (I bet I would've been a great electrician or gardener), my academic education, from bachelor to PhD, has taught me to take a scientific perspective on life. In my view, being a scientist is not a profession but a philosophy. I will always take my scientific mindset with me, whatever professional turn the future might hold. I hope to pass this perspective on life on to Anneleen, regardless of whatever professional path she wishes to take. Thank you for always enabling me to develop myself and my interests. When I say that the completion of my PhD is largely due to luck, I hope you know that I'm referring to you.

My dear brother, my best man, my daughter's godfather, and now my paranymph, thank you for your continuous interest in what I have been doing over the last few years. You have always viewed me as some sort of genius, although we both know how big of an idiot I can be. Nevertheless, your admiration of what I do has been of tremendous value for my self-esteem, which – believe it or not – came out of this journey with many bruises. Doing a PhD for a pretend genius like me is a lesson in humility. However, you always made me feel like I could take that next hurdle and solve the n^{th} problem, not by just saying comforting words, but by radiating a firm trust in my abilities. Thanks for always keeping me in high regards. I hope you know this goes both ways. Choosing you as my

paranymph was not just a symbolic choice. I also know that you are one of the few people who can fulfil the traditional role of a paranymph and defend my work on my behalf if need be, because you are one of the few people who has always shown a keen interest in my work and whose head does not start to spin whenever things get technical. You would've been a great academic, but I can't blame you chose to do something useful instead.

And finally, Inez, thank you for bearing with me as we build our lives together. When I started my PhD, we were just taking our first steps into our adult lives. Now, almost six years later, we have found our home and welcomed our daughter in it. Making me a father has been the greatest gift anyone could ever give me. Our family has given my life direction and an immense motivation to keep pushing forward. I want you to know that whenever I'm away from home again or whenever I'm working late, that I'm doing all of that for you and Anneleen. You two are the light of my life and without you the completion of this PhD would feel hollow. I love you so much.

About the author

Frederick Vinckenbosch was born in Hasselt, Belgium, on February 24th 1993. He completed his high school education, general secondary education with a focus on Latin and modern languages, at the Sint-Martinusscholen in Herk-De-Stad, Belgium in 2011. Hereafter, he enrolled in the bachelor of Psychology program at the Faculty of Psychology and Neuroscience of Maastricht University in the Netherlands. Frederick completed his bachelor in Psychology in 2014 with a minor in biological psychology. His bachelor thesis focused on exploring a possible link between chronic cannabis abuse and major depressive disorders. hereafter, Frederick enrolled in the master program in Cognitive and Clinical Neuroscience with a specialization in neuropsychology at the same faculty, which he completed in 2016 with a thesis on the effects of polypharmacy on driving performance and related cognitive functioning.

During the last year of his master program, Frederick was offered a position as PhD-student at the department of Neuropsychology and Psychopharmacology at the Faculty of Psychology and Neuroscience of Maastricht University under the supervision of Prof. Dr. Jan Ramaekers, Dr. Annemiek Vermeeren, and Dr. Eric Vuurman. As of September 2016, Frederick started working on his PhD, focusing in medication use and driving performance in clinical populations, with an emphasis on experimental research. Apart from a fair chunk of psychopharmacology, this topic requires the continuous evaluation of methods of assessing driving performance, as well as raising questions about how to identify individuals who exhibit meaningful driving impairment. In addressing these questions, Frederick has developed a keen interest in statistics and machine learning techniques for the identification of impaired individual drivers which he hopes to embed in future projects.

Output

Articles included as part of this dissertation:

Vinckenbosch, F. R. J., Vermeeren, A., Verster, J. C., Ramaekers, J. G., & Vuurman, E. F. (2020). Validating lane drifts as a predictive measure of drug or sleepiness induced driving impairment. Psychopharmacology, 237(3), 877-886.

Vinckenbosch, F. R., Vermeeren, A., Vuurman, E. F., van der Sluiszen, N. N., Verster, J. C., van de Loo, A. J., ... & Ramaekers, J. G. (2021). An explorative approach to understanding individual differences in driving performance and neurocognition in long-term benzodiazepine users. *Human Psychopharmacology: Clinical and Experimental*, *36*(4), e2778.

Vinckenbosch, F., Asin, J., de Vries, N., Vonk, P. E., Donjacour, C. E., Lammers, G. J., ... & Ramaekers, J. G. (2022). Effects of solriamfetol on on-the-road driving performance in participants with excessive daytime sleepiness associated with obstructive sleep apnoea. *Human Psychopharmacology: Clinical and Experimental*, e2845.

Vinckenbosch, F. R. J., Vermeeren, A., Theunissen E.L. & Ramaekers, J. G. (2022). Driving aptitude and fitness to drive. *Handbook of forensic medicine 2nd edition*, 995-1010.

Vinckenbosch F., Lammers G., Overeem S., Chen D., Wang G., Carter L.P., Zhou K., Ramaekers J.G. & Vermeeren A. (under review) Effects of solriamfetol on on-the-road driving in participants with narcolepsy: A randomised crossover trial. Human psychopharmacology: clinical and experimental

Articles not in this dissertation:

van der Sluiszen, N. N., Vermeeren, A., Jongen, S., **Vinckenbosch, F.**, & Ramaekers, J. G. (2017). Influence of long-term benzodiazepine use on neurocognitive skills related to driving performance in patient populations: a review. *Pharmacopsychiatry*, *50*(05), 189-196.

van der Sluiszen, N. N., Wingen, M., Vermeeren, A., **Vinckenbosch, F.**, Jongen, S., & Ramaekers, J. G. (2017). Driving performance of depressed patients who are untreated or receive long-term antidepressant (SSRI/SNRI) treatment. *Pharmacopsychiatry*, *50*(05), 182-188.

Vinckenbosch, F. R., Wingen, M., van der Sluiszen, N. N., Vermeeren, A., Jongen, S., & Ramaekers, J. G. (2019). Langdurig gebruik van tweedegeneratieantidepressiva en rijvaardigheid. *Neuropraxis*, *23*(1), 2-6.

Arkell, T. R., **Vinckenbosch, F.**, Kevin, R. C., Theunissen, E. L., McGregor, I. S., & Ramaekers, J. G. (2020). Effect of cannabidiol and Δ9-tetrahydrocannabinol on driving performance: a randomized clinical trial. *Jama*, *324*(21), 2177-2186.

Hayley, A. C., Shiferaw, B., Aitken, B., **Vinckenbosch, F.**, Brown, T. L., & Downey, L. A. (2021). Driver monitoring systems (DMS): The future of impaired driving management?. *Traffic injury prevention*, *22*(4), 313-317.

Hutten, N., Arkell, T., **Vinckenbosch, F.**, Theunissen, E., Kuypers, K., McGregor, I., & Ramaekers, J. (2021). P. 329 State anxiety induced by THC and CBD does not change after correcting for baseline anxiety levels. *European Neuropsychopharmacology*, *44*, S60-S61.

Mallaroni, P., Mason, N. L., **Vinckenbosch, F. R. J.**, & Ramaekers, J. G. (2022). The use patterns of novel psychedelics: experiential fingerprints of substituted phenethylamines, tryptamines and lysergamides. *Psychopharmacology*, 1-14.

Arkell, T. R., Kevin, R. C., **Vinckenbosch, F.**, Lintzeris, N., Theunissen, E., Ramaekers, J. G., & McGregor, I. S. (2022). Sex differences in acute cannabis effects revisited: Results from two randomized, controlled trials. *Addiction biology*, *27*(2), e13125.

Conference contributions:

Vinckenbosch F., van der Sluiszen N., Veldstra J., van Dijken J., Verster J.C., van de Loo A., Vermeeren A., Brookhuis K., Ramaekers J. (2017). Polypharmacy of ICADTS III drugs and long-term effects on driving. *Fortox 2017, Bled, Slovenia.* Oral presentation

Vinckenbosch, F.R.J., Vermeeren, A., Vuurman, E.F.P.M., van der Sluiszen, N.N.J.J.M., Verster, J.C., van de Loo, A.J.A.E., van Dijken, J.H., Veldstra, J.L., Brookhuis, K.A., de Waard, D. and

Ramaekers, J.G. (2018). An explorative approach to observing the impairing effects of chronic benzodiazepine use on driving in a clinical population. *The International Association of Forensic Toxicologists (TIAFT), Belgium, Ghent*. Oral presentation.

Vinckenbosch, F.R.J., Vermeeren, A., Vuurman, E.F.P.M., van der Sluiszen, N.N.J.J.M., Verster, J.C., van de Loo, A.J.A.E., van Dijken, J.H., Veldstra, J.L., Brookhuis, K.A., de Waard, D. and Ramaekers, J.G. (2019) An explorative approach to observing the impairing effects of chronic benzodiazepine use on driving in a clinical population. *International Council on Alcohol, Drugs and Traffic Safety (ICADTS), Canada, Edmonton*. Oral presentation.

Vinckenbosch F.R.J., Asin J., de Vries N., Vonk P.E. , Donjacour C.E.H.M., Lammers G., Overeem S., Janssen H., Wang G., Chen D., Carter L.P., Zhou K., Vermeeren A. & Ramaekers J.G. (2020). Effects of solriamfetol on driving performance in participants with excessive daytime sleepiness associated with obstructive sleep apnoea. *25th conference of the European Sleep Research Society, Sevilla, Spain.* Oral presentation

Vinckenbosch F., Lammers G., Overeem S., Chen D., Wang G., Carter L.P., Zhou K., Ramaekers J.G. & Vermeeren A.(2020) Effects of Solriamfetol on Driving Performance in Participants with Narcolepsy. *25th conference of the European Sleep Research Society, Sevilla, Spain.* Poster presentation.

Vinckenbosch F.R.J., Asin J., de Vries N., Vonk P.E. , Donjacour C.E.H.M., Lammers G., Overeem S., Janssen H., Wang G., Chen D., Carter L.P., Zhou K., Vermeeren A. & Ramaekers J.G. (2020). Effects of solriamfetol on driving performance in participants with excessive daytime sleepiness associated with obstructive sleep apnoea. *34th annual meeting of the Associated Professional Sleep Societies*. Philadelphia, Pennsylvania, United States of America. Poster presentation.

Vinckenbosch F., Lammers G., Overeem S., Chen D., Wang G., Carter L.P., Zhou K., Ramaekers J.G. & Vermeeren A.(2020) Effects of Solriamfetol on Driving Performance in Participants with Narcolepsy. *34th annual meeting of the Associated Professional Sleep Societies*. Philadelphia, Pennsylvania, United States of America. Poster presentation.

Vinckenbosch F. Vermeeren A., Schepers J. & Ramaekers, J. (2022). A role for lateral position data in the development of driver state monitoring systems. *International Council on Alcohol, Drugs and Traffic Safety (ICADTS), Rotterdam, the Netherlands.* Oral Presentation.

Vinckenbosch F., Theunissen E., Stelling A., Goldenbeld C. & Ramaekers J. (2022). Characteristics of "typical" recreational nitrous oxide use and its use in traffic: an online survey. *International Council on Alcohol, Drugs and Traffic Safety (ICADTS), Rotterdam, the Netherlands.* Oral Presentation.

Vinckenbosch F., Helmerhorst H., Durán Jiménez D., van de Langkruis R., Rurup M., Theunsissen E., Dahan A. & Ramaekers J. (2022). Mind-bending and breathtaking: An experimental crossover trial for uncovering the duration of psychomotor impairment after recreational nitrous oxide use and its relation to exhaled nitrous oxide concentrations. *International Council on Alcohol, Drugs and Traffic Safety (ICADTS), Rotterdam, the Netherlands.* Oral Presentation.

Media appearances:

Hasselaar test rijgedrag met en zonder slaappillen (2018). Het belang van Limburg. https://www.hbvl.be/cnt/dmf20180325_03427958. News paper interview.

Pillen even gevaarlijk als pilsjes in het verkeer. Nieuwe feiten (radio 1) (2018). https://radio1.be/luister/podcast/nieuwe-feiten/1/138. Radio interview

Vermoeid achter het stuur? Het is gevaarlijker dan u denkt (2019). *Omroep Max*. https://www.maxvandaag.nl/sessies/themas/reizen-verkeer/vermoeid-achter-het-stuur-het-is-gevaarlijker-dan-u-denkt/. Television interview.

Slaaptekort bij vrachtwagen chauffeurs (2020). *De Monitor (KRO-NRCV)*, 28 september 2020. https://www.npostart.nl/de-monitor/28-09-2020/KN_1715954. Television interview.