THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Methods for Analysis of Naturalistic Driving Data in Driver Behavior Research

From crash-causation analysis using expert assessment to quantitative assessment of the effect of driver behavior on safety using counterfactual simulation

JONAS BÄRGMAN

Department of Applied Mechanics

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2016

Methods for Analysis of Naturalistic Driving Data in Driver Behavior Research
From crash-causation analysis using expert assessment to quantitative assessment of the effect of driver behavior on safety using counterfactual simulation
JONAS BÄRGMAN
ISBN 978-91-7597-501-6

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4182
ISSN 0346-718X

Department of Applied Mechanics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover image:
We live in a time of change. How will our children relate to driving when they grow up? Driver behavior, the vehicle, and the environment are changing and we need to understand how these changes affect safety. Learn more about the cover image on page 48.

Chalmers Reproservice
Gothenburg, Sweden 2016

To my family

METHODS FOR ANALYSIS OF NATURALISTIC DRIVING DATA
IN DRIVER BEHAVIOR RESEARCH

From crash-causation analysis using expert assessment to quantitative assessment of the effect of driver behavior on safety using counterfactual simulation

JONAS BÄRGMAN
Department of Applied Mechanics
Chalmers University of Technology

Abstract

In the last several years, the focus of traffic safety research—especially when performed in association with the automotive industry—has shifted from preventing injury during a crash to avoiding the crash altogether or mitigating its effects. Pre-crash safety measures include intelligent safety systems (e.g., different levels of automated driving), infrastructure design, behavior-based safety, and policy-making. Understanding driver behavior is crucial in the development and evaluation of such measures. Naturalistic driving data (NDD) can facilitate this understanding by providing information about crash causation and contribute to the evaluation of pre-crash safety measures and the effects of driver behavior on safety. However, NDD's complexity calls for new and better methods to fully exploit its advantages.

This thesis, together with the five included papers, addresses several gaps in current scientific knowledge by presenting novel methods for analyzing NDD that address multiple aspects of the development process for pre-crash safety measures. The chunking method (Paper I) helps to identify and overcome common biases in analysis of everyday-driving time-series data, while the expert-assessment-based crash-causation analysis method (Paper II, supported by Paper III) is a novel approach to studying crash causation through the analysis of NDD with video. Product and prototype development can be improved by utilizing counterfactual simulations, for which the choice of driver behavior model is shown to be crucial (Paper IV)—an awareness that was previously lacking. Being able to compare the effects of drivers' specific behaviors (e.g., driver-vehicle interactions or in-vehicle secondary tasks) on safety could both speed up development of safety measures and improve vehicle designs and design guidelines. Methods to perform such comparisons through the combination of counterfactual glance behavior and pre-crash kinematics had been missing (but are provided in Paper V). This thesis further improves the evaluation of pre-crash safety measures by providing more robust analyses of everyday driving data (Paper I) and by demonstrating the importance of good mathematical models of driver behavior in virtual evaluation (Paper IV).

In summary, these new methods fill important research gaps and have the potential to improve the design of pre-crash safety measures through the use of NDD. Using NDD can augment our understanding of driver behavior and crash causation, important aspects of improving traffic safety and fulfilling Sweden's Vision Zero.

Keywords: naturalistic driving data, driver behavior analysis, safety measures, ADAS, automated driving, safety benefit evaluation, crash causation, counterfactual simulations

Acknowledgments

I would first like to thank my supervisors Prof. Kip Smith, Dr. Marco Dozza, and Prof. Per Lövsund for helping me grow as a scientist, where Marco Dozza deserves a special thanks for the everyday discussions (although not around the coffee machine). I would also like to thank Kristina Mayberry for her thorough language reviews. In addition, some colleagues deserve special thanks. Johan Engström and Trent Victor have been there as sparring partners from the start, nudging me towards important high-impact research—both from an academic and an industry perspective. Other colleagues that have directly or indirectly contributed to my research include Giulio Piccinini, Gustav Markkula, Olle Nerman, Vera Lisovskaja, Julia Werneke, Christian-Nils Boda, Johan Lodin, Daniel Nilsson, Selpi, Alberto Morando, Helena Gellerman, Erik Svanberg, Ines Heinig, Mikael Ljung Aust, and Anna Nilsson-Ehle. Much discussion and many hours with data have taught me to respect all parts of the research chain, from formulating research questions and finding project financing, via getting and processing data, to analysis and writing scientific publications. All of it is crucial to get to the finish line. Prof. John Lee and Carol Flannagan have also been major sources of inspiration. Thank you all!

I would like also to acknowledge the financial and inspirational contributions to this thesis from VINNOVA, FFI, EU, Chalmers, and the SHRP2 program sponsors, through projects such as EFrame, euroFOT, DCBIN, ANNEXT, DREAMi, CRASHED, UDrive, and SHRP2 SO8A.

Finally, Daniela. Without you I would not be where I am today. You are fantastic and have supported me so completely in my challenges, not the least related to the pursuit of my PhD. You are a treasure—thank you! Both our children have been there through it all, brightening my life and just being smart, strong, caring, funny, sensitive, cool, kind, and absolutely wonderful! Thank you also to my friends and family for being there for me.

List of papers with research gaps and scientific contributions

Paper I — Dozza, M., Bärgman, J., & Lee, J.D. (2013). Chunking: A procedure to improve naturalistic data analysis. *Accident Analysis & Prevention*, 58, 309-317. doi: 10.1016/j.aap.2012.03.020

Author contrib.: Dozza and Bärgman developed the method and jointly authored the paper, with support by Lee.

Research gap: Naturalistic data analyses often compare parameters calculated from segments of data with different durations. Such analyses further often assume independence of observations—an assumption that is not always correct. The potential biases introduced as a result are poorly understood, scarcely documented, and seldom acknowledged. Further, a method has been lacking to determine whether data segments have the appropriate duration to provide reliable parameters.

Scientific contrib.: This paper demonstrates how comparing data segments of different duration introduces severe biases in data analysis, and presents a novel methodology to avoid such biases and control for dependent observations. This paper further presents a method to determine the optimal segment data duration to maximize statistical power, minimize bias, and avoid calculating parameters when data segments are too short.

Paper II — Engström, J., Werneke, J., Bärgman, J., Nguyen, N., & Cook, B. (2013). Analysis of the role of inattention in road crashes based on naturalistic on-board safety monitoring data. *Proceedings of the 3rd International Conference on Driver Distraction & Inattention.* Göteborg, Sweden

Author contrib.: All authors jointly and iteratively developed the analysis plan and the qualitative coding schema. Bärgman contributed to authoring the method section, with primary focus on the quantitative data section. Bärgman also contributed to the data processing, analysis, and interpretation of results.

Research gap: Methods for understanding crash causation based on expert assessment are rare, and are typically developed for application to traditional in-depth crash investigation data, without video or other types of pre-crash time-series data (e.g., speed and acceleration).

Scientific contrib.: This paper demonstrates an expert-assessment method designed to be applied to naturalistic driving data with video. Its application to crashes and near-crashes from commercially collected naturalistic driving data has provided insights into light and heavy vehicle crash-causation mechanisms for rear-end and intersection crashes.

Paper III — Bärgman, J., Werneke, J., Boda, C.N., Engström, J., & Smith, K. (2013). Using manual measurements on event recorder video and image processing algorithms to extract optical parameters and range. *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment*, Training, and Vehicle Design, (pp. 177-183). Bolton Landing, N.Y.

Author contrib.: Bärgman developed the method and was the primary author, in addition to designing the study and applying the method jointly with co-authors.

Research gap: Many sets of naturalistic driving data do not include the distance from the instrumented vehicle to a lead vehicle, but forward video is recorded. A pragmatic method to extract lead-vehicle distance from forward video has been missing.

Scientific contrib.: A method to extract lead-vehicle distance from forward video was developed and validated. This methodology also provides relative velocity and optical parameters such as looming.

Paper IV Bärgman, J., Boda, C.N., & Dozza, M., (2016). Counterfactual simulations applied to SHRP2 crashes: The effect of driver behavior models on safety benefit estimations of intelligent safety systems. Submitted to an international scientific journal.

Author contrib.: Bärgman developed the final study design, developed and implemented a majority of the simulations, and was the primary author.

Research gap: Counterfactual simulations are increasingly common in the development of active safety and automated driving. However, such simulations typically include simplistic driver models, and do not consider glance behavior or the driver response process to a threat. In addition, counterfactual simulations have yet to take full advantage of naturalistic data.

Scientific contrib.: This paper demonstrates the importance of the choice of driver model in counterfactual simulations. It further shows how glance behavior and reaction process models and their parameters can affect the safety benefit estimate, allowing counterfactual simulations to benefit (in new ways) from detailed naturalistic driving data.

Paper V Bärgman, J., Lisovskaja, V., Victor, T., Flannagan, C., & Dozza, M. (2015). How does glance behavior influence crash and injury risk? A 'what-if' counterfactual simulation using crashes and near-crashes from SHRP2. *Transportation Research Part F: Traffic Psychology and Behaviour*, 35, 152-169. doi: 10.1016/j.trf.2015.10.011

Author contrib.: Bärgman extracted the data, developed the method, designed the study, applied the method and was the primary author.

Research gap: Glance behavior is critical to safe driving, and secondary tasks and in-vehicle interfaces affect glance behavior. Nevertheless, methods for studying the interaction between glance behavior and pre-crash kinematics have been lacking. The research community has also lacked knowledge on the (dis)similarities of pre-crash kinematics for crashes and near-crashes.

Scientific contrib.: This paper presents a novel method to estimate how glance behavior affects safety in critical situations, using pre-crash kinematics from naturalistic crashes and near-crashes to determine how glance behavior may change the nature (from crash to near-crash) and the severity (impact speed and delta-v) of a crash. The paper further demonstrates that there are only small differences in the pre-crash kinematics between crashes and near-crashes—up until the driver starts to perform an evasive maneuver.

Figure 1 shows the Papers I-V in relation to the context of pre-crash safety measure development:

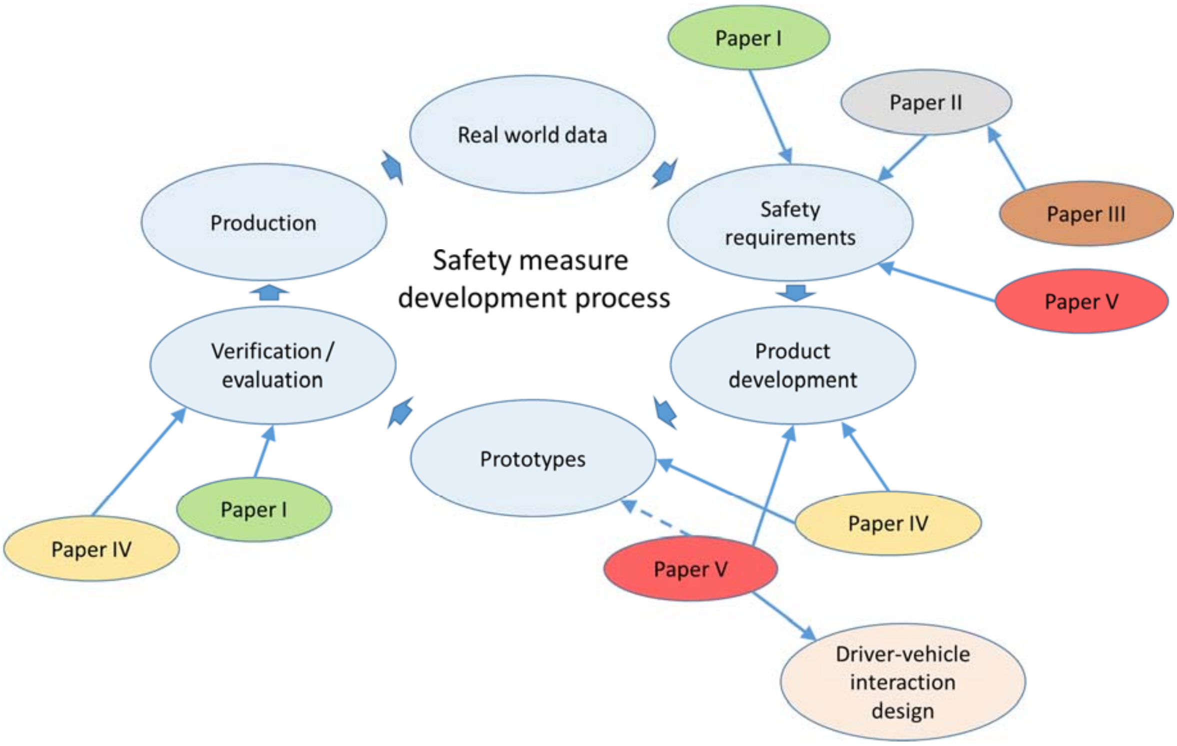

Figure 1: The light blue circles at the center describe Volvo's development process (the 'circle of life') used at the Volvo Car Group since the 1980s (Jakobsson, Lindman, Svanberg, & Carlsson, 2010). Papers I-V are framed around this development process for safety systems. An extended version of this figure, with information on the individual papers' contributions to the development process, is also presented in the Discussion (Section 4).

Definitions and acronyms used in this thesis

Actual severity – The actual outcome of an event, for example: the impact speed with corresponding injury for crashes, and time-to-collision in near-crashes.

AIS – Abbreviated injury scale.

ARIMA – Autoregressive integrated moving average.

Baseline – a term subsuming both the term controls in epidemiology, and segments of everyday driving data in NDD.

CAN – Controller area network.

CNDD – Commercially collected NDD, where (a) the origin of the data is a commercial entity (company), (b) a main incentive of their collection is commercial (making roads safer and saving lives are also important incentives), and (c) the deployment of more instrumented (customer) vehicles impacts company's revenue positively (contrary to NDS of NFOT, where each new unit incurs additional cost).

Counterfactual simulations for safety benefit analysis – Mathematical simulations of counterfactual time-series data of traffic events (often based on real crash kinematics), typically performed both with and without one or several ISS (algorithms and virtual actuators) applied.

Data – Information collected on-road in a naturalistic setting, including quasi-experimental studies (NFOT), observational studies (NDS), and studies using commercially collected NDD.

Delta-V – The change in velocity of a vehicle during a crash event.

DVI – Driver-vehicle interface.

EDR – Event data recorder. "An event data recorder (EDR) is a function or device installed in a motor vehicle to record technical information about the status and operation of vehicle systems for a very brief period of time (i.e., a few seconds) and in very limited circumstances (immediately before and during a crash), primarily for the purpose of post-crash assessment of vehicle safety system performance" (NHTSA, 2012a, p. 74145).

Everyday driving – When drivers go about their everyday lives. Typically recorded in NDD.

FOT – Field operational test. Typically the same as an NFOT (but commonly misused to mean any data collection in real traffic).

FV – Following vehicle

GPS – Global positioning system.

Harm – A metric quantifying the outcome of a crash based on the monetary cost of injuries.

ISS – Intelligent safety systems. The term ISS subsumes all forms of in-vehicle technologies that are active before a crash, and that directly or indirectly are intended to avoid or reduce the severity of crashes. ISS include advance driver assistance systems, active safety systems, cooperative systems, and different levels of automated vehicles.

LV – Lead vehicle

Naturalistic – in a natural setting, without constraints imposed.

NDD – Naturalistic driving data. Data collected unobtrusively in drivers' vehicles.

NDS – Naturalistic driving study. A study where data is collected unobtrusively in drivers' vehicles as they go about their everyday lives. Typically an NDS aims to reveal correlations between traffic events, driver behavior and crash causation.

NFOT – Naturalistic field operational test: A quasi-experimental field study with a treatment/control design, aimed at evaluating one or more (safety) measures by unobtrusively collecting data during participating drivers' everyday driving.

NHTSA – (US) National highway traffic safety administration.

PET – Post-encroachment time.

Potential severity – The potential outcome of an event: what could have happened had something been different. Both crashes and non-crashes have potential severities, for example in terms of estimated crash or injury risk.

Pre-crash safety measure – Physical device or rule that aims to reduce the number of crashes and injury outcomes in traffic by acting before a crash occurs. These include infrastructure design, legislation, policy-making, and intelligent safety systems.

Prospective safety benefit analysis – analysis which estimates the future benefit of safety measures (e.g., ISS) before they have been deployed (e.g., systems in development, or prototype systems)—when statistical data from on-road crashes are not available.

Qualitative analysis of crash causation – The use of human experts who utilize (consider) all available information for a specific event (e.g., crash or near-crash) as a basis for identification of crash causation factors.

Range – The relative distance between two objects on the road. Typically, the distance between a following and a lead vehicle.

Range-rate – The relative speed between two objects on the road. Typically, the relative speed between a following and a lead vehicle.

Retrospective safety benefit analysis – Analysis estimating the safety benefit for safety measures that are already deployed (on market), typically using statistical data found in, for example, crash databases or insurance data.

SCE – safety-critical event (SCE). Events in traffic hypothesized to pose an increased level of risk of a crash, typically including several levels of outcome, such as crashes, near-crashes, and crash-relevant conflicts.

SHRP2 – The second Strategic Highway Research Program. A large US government program (2006-2015) aimed at finding strategic solutions to national transportation challenges. A large naturalistic driving study was conducted in a subprogram on highway safety.

SDLP – Standard deviation of lane position.

Time-series data – Data available at discrete intervals (e.g., 10 Hz) for a segment of time (seconds, minutes or entire trips), including: sensor (e.g., accelerometer), video collected continuously/automatically, and manually annotated video.

TTC – Time to collision.

Table of Contents

1 Introduction

1.1 Background

Crashes in traffic accounted for over 1.24 million fatalities world-wide in 2010. They stand out as the ninth leading cause of death, and the first for men between 15 and 29 years old (WHO, 2013).

Several studies have shown that driver behavior in the pre-crash phase (before impact) is a main contributing factor to traffic crashes. In a comprehensive review of US crashes, Treat et al. (1977) concluded: “Thus, conservatively stated, the study indicates human errors and deficiencies were the cause in at least 64% of crashes, and were the probable cause in about 90-93% of crashes investigated” (p. 28). More recent work has focused on identifying the exact nature of the driver behaviors that end in crashes (Carney, McGehee, Harland, Weiss, & Raby, 2015; Dunn, Hickman, & Hanowski, 2014; Markkula, Engström, Lodin, Bärgman, & Victor, 2016; Victor et al., 2015). Studies of driver behavior are typically aimed at facilitating safer traffic by informing the design of **pre-crash safety measures**[1].

Pre-crash traffic safety measures include infrastructure design (Andersson et al., 2005; Theeuwes & Godthelp, 1995), driver training (Christie, 2001; McGehee, Raby, Carney, Lee, & Reyes, 2007a), legislation and policy-making (Bronrott, 2010; NHTSA, 2013, 2016), and intelligent safety systems (Brännström, Sjöberg, & Coelingh, 2008; Distner, Bengtsson, Broberg, & Jakobsson, 2009). In this thesis the term **intelligent safety systems** (ISS) for traffic safety subsumes all forms of in-vehicle technologies that are active before a crash, which directly or indirectly are intended to avoid crashes or reduce their severity. ISS include advance driver assistance systems, active safety systems, cooperative systems, and different levels of automated vehicles.

1.1.1 Four approaches to studying driver behavior

A review of the recent literature on traffic safety reveals four fundamentally different, but complementary, approaches to the study of driver behavior in traffic and in crashes. Each offers useful insight into various aspects of traffic safety—and in particular, into the role of driver behavior in crash causation.

The first approach, **experimental studies** of driver behavior in specific situations, typically aims to link driver behavior either directly with crashes (Rizzo, McGehee, Dawson, & Anderson, 2001) or with some surrogate measure that is hypothesized to correlate with the occurrence of crashes and crash risk (e.g., Muhrer & Vollrath, 2011; Strayer et al., 2015; Vadeby et al., 2010; Vaux, Ni, Rizzo, Uc, & Andersen,

[1] Bold will be used to highlight important terms.

2010). This highly controlled approach typically uses driving simulators (e.g., Boyle & Lee, 2010; Engström, Johansson, & Östlund, 2005; Wortelen, Baumann, & Lüdtke, 2013), laboratory settings (e.g., Caird & Hancock, 1994; Hancock, Caird, & Johnson, 1991), or test-track experiments (e.g., Bärgman, Smith, & Werneke, 2015; Kiefer et al., 2003; Summala, Lappi, Pekkanen, Lehtonen, & Hietamäki, 2012). Experimental studies require *a priori* specification of the scenario and the driver behavior to be studied (for example, cognitive load, hypothesized to contribute to crashes). This approach facilitates the testing of hypothesized causal relationships between specific driver behaviors (e.g., glance behaviors) and specific traffic scenarios (e.g., a lead vehicle in a highway car-following situation).

The second approach, traditional collection and analysis of **in-depth crash investigation data**, primarily addresses questions related to injury outcome (Fagerlind, Martinsson, & Hagström, 2010; Otte, Krettek, Brunner, & Zwipp, 2003; Seeck et al., 2009). The portion of these data relevant for pre-crash driver behavior research has typically included variables that document vehicle kinematics (e.g., speeds and accelerations) and environmental factors (e.g., weather and road conditions), together with driver and witness accounts of the event (Paulsson, 2005; Sandin & Ljung, 2007; Seeck et al., 2009). The pre-crash kinematics are reconstructed from, for example, tire tracks, post-crash positioning of the vehicles, and vehicle deformations (Niehoff & Gabler, 2006). The data collected through crash investigations can be used in a variety of ways: as input into epidemiological studies of crash occurrence and injuries (Kullgren, 2008; Lefler & Gabler, 2004); as a basis for simulations of vehicle kinematics in crashes; and as a means to identify crash-causation factors. The methods may be based on expert assessment (e.g., Dunn et al., 2014; Habibovic, Tivesten, Uchida, Bärgman, & Ljung Aust, 2013; Sandin & Ljung, 2007; Van Elslande & Fouques, 2007) or different epidemiological methods studying why crashes occur (e.g., Carney et al., 2015; Hickman, Hanowski, & Bocanegra, 2010; Toth, Radja, Thiriez, & Carra, 2003; Victor et al., 2015). Driver and witness accounts, collected through interviews and questionnaires, provide information about driver state (e.g., fatigue). However, this approach cannot provide an in-depth understanding of actual driver behavior (or the detailed interplay of road users) in the few seconds before the crash.

The third approach, **drivers' self-reports** leverages the explicit and tacit information held by drivers about their driving and behavior, the surrounding traffic, the infrastructure, and their vehicle. A wide range of self-report methods have been used in traffic safety research. A literature review found questionnaires and interviews to be the most common self-report tools for eliciting an understanding of driver behavior in the traffic context. Questionnaires are attractive because they enable the collection of a large sample of data relatively quickly. Two examples are the Manchester Driving Behavior Questionnaire, DBQ (Parker, Reason, Manstead, & Stradling, 1995; Reason, Manstead, Stradling, Baxter, & Campbell, 1990), and the Driving Style

Questionnaire, DCQ (French, West, Elander, & Wilding, 1993). Both consist of a series of questions (e.g., How often do you exceed the posted speed limit?) that address aspects of driver behavior (specifically, driving style). The format is forced-choice, with three to five ranges as potential answers, permitting aggregation of the data across respondents and the subsequent use of non-parametric statistical analyses. (For a review on self-reporting tools for research on driving styles, see the informative work by Sagberg, Selpi, Bianchi Piccinini, and Engström (2015).) In contrast, interviews enable the researcher to dig more deeply into a driver's opinions and beliefs. Because they are typically conducted one-on-one, the data can be quite rich and revealing. Given the highly personalized, subjective nature of interview data, they must be treated like observational data rather than experimental data. Examples of the use of interviews include the qualitative descriptions of driving habits by Tillmann and Hobbs (1949) based on detailed interviews of taxi drivers, and the research by Houtenbos (2008) into drivers' views on the expectations and interactions between road users in intersection negotiation. In addition to questionnaires and interviews, self-reports have been a fundamental component in the evaluation of driver acceptance of, and trust in, ISS (Pettersson & Karlsson, 2015; Seppelt, 2009).

Finally, in the last several years a fourth approach has been developed. The collection and analysis of **naturalistic driving data** (NDD) allow the observation and understanding of driver behavior in real traffic (Bronrott, 2010; Carney et al., 2015; Dozza, 2013; Fancher et al., 1998; Hallmark et al., 2011; Hickman et al., 2010; LeBlanc et al., 2006; Neale, Dingus, Klauer, Sudweeks, & Goodman, 2005; Othman, Thomson, & Lannér, 2012; Peng, Boyle, & Hallmark, 2013; Sayer et al., 2010; Tivesten & Dozza, 2014a, 2014b; Uchida, Kawakoshi, Tagawa, & Mochida, 2010; Victor et al., 2010; Victor et al., 2015). The data are **unobtrusively acquired**, and typically include information about the driver, the vehicle, and the driving environment (including other road users). They are typically collected from a variety of sources, such as accelerometers, GPS, radar, and video of the driver, the vehicle, and the surrounding traffic environment. Information this detailed (i.e., dynamic, context-dependent time-series data), was not available until recently. Some of this data may be available from the vehicle's electronic bus system, for example a Controller Area Network (CAN; ISO_11898, 2003).

1.1.2 Naturalistic driving data (NDD)

Until recently there have been only two sources of NDD, **naturalistic field operational tests** (NFOT) and **naturalistic driving studies** (NDS). A third source of NDD has recently become available to some researchers: **commercially collected NDD**. These three sources are described in turn.

Naturalistic Field Operational Tests (NFOT)
NFOT are projects that evaluate some form of pre-crash safety measure, for example, one or more ISS (Bao, LeBlanc, Sayer, & Flannagan, 2012; Benmimoun, Ljung Aust, Faber, & Saint Pierre, 2011; Bezzina & Sayer, 2015; Carsten et al., 2008; Dozza et al., 2010; Fancher et al., 1998; LeBlanc et al., 2006; Ljung Aust, Regan, & Benmimoun, 2011; Mononen et al., 2012; Sayer et al., 2011; Sayer et al., 2010; Viti, Hoogendoorn, Alkim, & Bootsma, 2008). These studies are often empirically rigorous, with both **treatment** and **baseline** (control) phases to enable statistical inference. Many NFOT include sections with descriptive statistics of normal everyday driving (LeBlanc et al., 2006; Sayer et al., 2010). As discussed in Ljung Aust et al. (2011), one drawback of an NFOT of production ISS is that ISS are not always available to consumers as individual products. (For example, forward-collision warning is bundled with adaptive cruise control.) As a result, disentangling the effects of the different systems on safety is not always easy. On the other hand, an advantage of NFOT is that it is probably the method with the highest **ecological validity** capable of evaluating early-to-market and pre-production ISS (Carsten et al., 2008; Ljung Aust et al., 2011; Sayer et al., 2011). This validity is achieved by performing the evaluation through the study of natural behavior in the real world—on real roads, in everyday driving, using the actual system to be evaluated—basically, the definition of a method with high ecological validity (Schmuckler, 2001). Generalizability is, however, still typically limited by, for example, the selection of study participants (Dozza et al., 2010), and the common use of **surrogate (proxy) measures of safety** (e.g., studying near-crashes instead of crashes; this consideration is addressed Section 4.1).

Ambiguity in the literature makes it appropriate to take a stand on issues of NFOT nomenclature. First, the term Field Operational Test (FOT) has been used by some authors to refer to any study conducted on actual roads in traffic (Festag, Le, & Goleva, 2011). In this thesis the acronym NFOT (with the prefix N) is reserved for studies with a treatment/control design which are aimed at evaluating one or more (safety) measures by unobtrusively collecting data during participating drivers' everyday driving. Even the term naturalistic is not always used in conventional ways. It is, for example, not obvious what a "naturalistic driving simulator" (Daza et al., 2011, p. 1199) is. In this thesis naturalistic means in a natural setting, without constraints imposed, in contrast to controlled experimental conditions.

Furthermore, the term **baseline** often appears in literature on NFOT studies as a synonym for an experimental treatment that epidemiology typically calls **control** (Rothman, 2012). However, NFOT baselines are also used in the descriptive analysis (statistics) of drivers' everyday driving, without a comparison to a treatment phase (Othman, Thomson, & Lannér, 2014; Sayer, Devonshire, & Flannagan, 2007; Tivesten & Dozza, 2015). The NDS (Naturalistic Driving Studies) literature typically assigns both meanings to the term *baseline*, so this thesis will also use *baseline* to denote both the epidemiological controls and the segments of everyday driving data. In this thesis **data** refers to information collected on-road in a naturalistic setting, including quasi-experimental studies (NFOT), observational studies (NDS), and studies using commercially collected NDD.

Naturalistic Driving Studies (NDS)

One of the main differences between NDS and NFOT studies is that the latter typically reveal correlations between traffic events, driver behavior and **crash causation** (Blatt et al., 2015; Neale et al., 2005; Utesch et al., 2014), rather than evaluating a specific safety countermeasure (e.g., ISS). Unlike NFOT, NDS are **strictly observational** (Carsten, Kircher, & Jamson, 2013); they do not contrast treatment and baseline conditions as part of the study design. Two of the most common foci of NDS are normal **everyday driving** (Neale et al., 2005; Othman et al., 2014; Sayer, Mefford, & Huang, 2003; Victor et al., 2010), when drivers go about their everyday lives (Carsten et al., 2013), and **safety-critical events** (SCE). The latter are hypothesized to pose an increased level of crash or injury risk (typically used as crash surrogates; Dozza & González, 2013; Paper II; Paper III; Fancher et al., 1998; LeBlanc et al., 2006; Liang , Lee, & Yekhshatyan, 2012; Victor et al., 2010), and typically include several levels of outcomes such as crashes, near-crashes, and crash-relevant conflicts (SHRP2, 2012). Studies often analyze both everyday driving and SCE—at least, the data typically allow for the analysis of both.

As Victor et al. (2010) noted, it is possible to base an observational study on conveniently acquired data from an NFOT. Not only can an NFOT baseline be seen as an NDS, but the treatment phase of an NFOT can be used for research unrelated to the systems the NFOT aims to evaluate (Gordon et al., 2010; LeBlanc, Sivak, & Bogard, 2010; Sayer et al., 2007; Sayer et al., 2003; Tivesten & Dozza, 2015). However, care must then be taken to address the potentially confounding effects of, for example, prototype ISS, or interaction effects among different ISS (Carsten et al., 2013; Ljung Aust et al., 2011).

It is not strictly necessary to distinguish between NDS and NFOT in this thesis or the methods presented in the included papers. However, the possibility of using NFOT for the study of driver behavior (unrelated to the original intent of the NFOT) may not be obvious and warrants mention because it represents a potentially valuable additional data source. For convenience, NDS will be the general term for both NDS and NFOT, unless explicitly stated otherwise.

Commercially collected NDD

Finally, the third source, **commercially collected NDD** (CNDD), is not typically intended to answer a set of research questions, but to generate a profit while improving road safety (Lytx, 2016; SmartDrive, 2016). In this thesis the term *commercial NDD* is used when (a) the origin of the data is a commercial entity (company), (b) their collection is being driven by commercial incentives (however, typically, making roads safer and saving lives are important complementary incentives), and (c) the deployment of additional instrumented (customer) vehicles (collecting the NDD) has a positive impact on the company's revenue. (In contrast, each additional vehicle increases the cost of NDS or NFOT data collection.)

Only a few studies have used commercial NDD to date (Carney et al., 2015; Hickman & Hanowski, 2010; Hickman et al., 2010; Lich & Georgi, 2011; McGehee, Carney, Raby, Lee, & Reyes, 2007b; McGehee, Carney, Raby, Lee, & Reyes, 2007c; McGehee & Carsten, 2010; Soccolich & Hickman, 2014). The differences between commercially collected NDD and NDD collected specifically for research are discussed in Section 4.3.

All three sources of NDD, with their high ecological validity, complement traditional experimental studies (Carsten et al., 2013). Additionally, these sources typically provide more detailed records of the driver, the vehicle, and the environment (both in everyday driving and in the few seconds before a crash or other safety-critical event) than are available from, for example, in-depth crash investigations.

1.1.3 The variety of data in NDD

Video, an important component in most NDD research, is used to understand the driving behavior of individual drivers, as well as the interaction between drivers and their surroundings. Data without video are beyond the scope of this thesis, although the method presented in Paper I could also be applied to NDD without video. For site-based data collection, the video is usually collected from cameras on nearby road infrastructure or buildings (Laureshyn, 2010; Smith, Thome, Blåberg, & Bärgman, 2009), or from mobile towers put in place specifically for the study (Gordon et al., 2012). However, site-based NDD are beyond the scope of this thesis. In this thesis, the term 'NDD' refers to in-vehicle data collected unobtrusively.

Most NDD with video include at least one camera facing forward toward the road ahead and one focused on the driver's face and/or body to identify actions and reactions. However, in some NDD only the forward video is available (Lich & Georgi, 2011; Lich et al., 2012). Additional camera views may include (a) a close-up of the driver's face to capture eyes and facial expressions, (b) a view of the driver's feet to capture brake readiness and reaction times, (c) a rear-facing camera to study the effect of trailing vehicles on driver behavior, and (d) side views to capture road users' actions in complex environments such as intersections (Figure 2). The method presented in Paper III can be applied to NDD with forward video only, but the methods in Papers II and V require video or some other means of monitoring of drivers' visual attention (e.g. eye-trackers; Holmqvist et al., 2011), as well as monitoring of the surrounding traffic scene with, for example, radar (Valldorf &

Gessner, 2003). The counterfactual simulation approach of Paper IV can use vehicle kinematics from NDD without video, but then driver-monitoring data from other sources is needed for full utilization. In addition to video, NDD often include records from the vehicle's internal communication bus (CAN; ISO_11898, 2003). This information may include, but is not limited to, the list shown in Figure 3.

Figure 2: Example of video configuration in NDS. From production system units in the UDrive project (Barnard, Utesch, van Nes, Eenink, & Baumann, 2015). Note that the driver's face is blurred to ensure privacy. Printed with permission from the UDrive project and the experimenter in the video.

The availability of the various types of **CAN** data varies appreciably between NDD sets. Practically speaking, data that include proprietary ISS or derivatives thereof are usually available only to the vehicle manufacturer, the suppliers of the ISS, and trusted research organizations (Victor et al., 2010), since all manufacturers want to avoid reverse engineering and other commercial intellectual property leakage. The methods in all papers in this thesis would benefit from high-quality CAN data (e.g., high-quality speed data, or even range and range rate from forward radars). Although CAN data was used in Papers I, IV and V, they were high-quality only for Paper I.

The measures available on CAN are applicable to a broad range of NDD studies which consider the role of driver behavior in critical events. Some examples of information collected, along with what it is used for, include: accelerometers and brake pedal position to study pre-crash behaviors; information on light condition, windshield wipers, and ambient temperature as potential factors contributing to crashes (indicating visibility, precipitation, and road friction); ISS information for studying driver interaction with such systems and the surrounding traffic (Faber et al., 2011; LeBlanc et al., 2006).

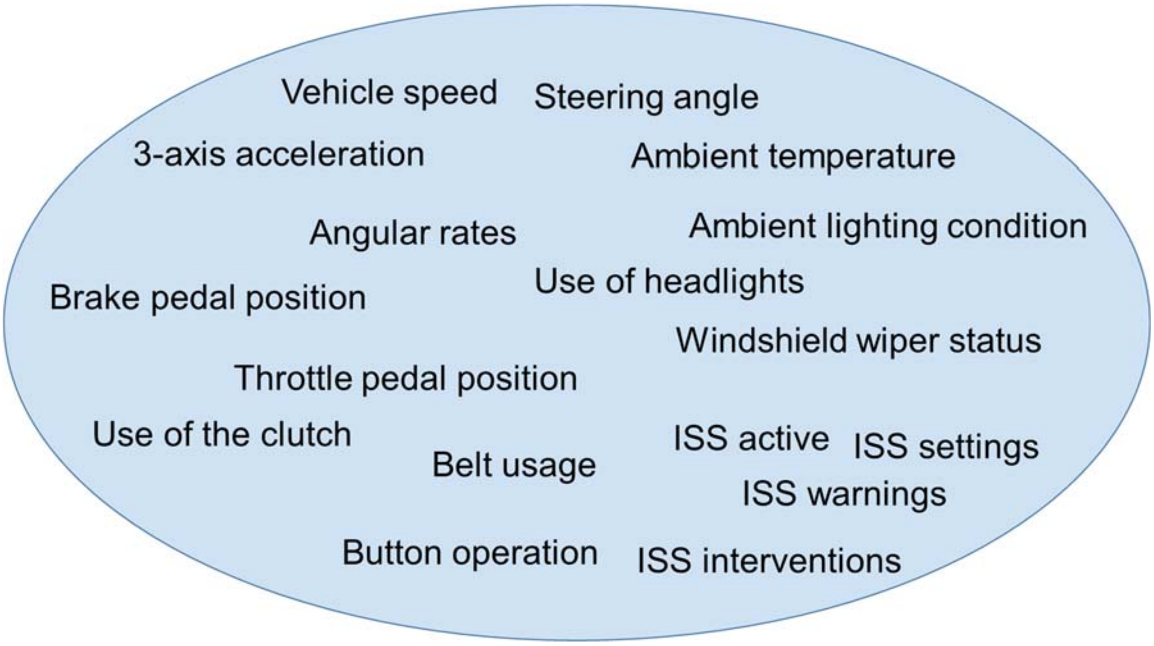

Figure 3: Examples of data that can be available from CAN in NDD.

Additional in- or on-vehicle sensors can be added which are not part of the (individual) CAN configuration, in order to provide specific information about key features of the interaction between the driver, vehicle, and environment—for example: (a) radar for range and range rate with respect to other road users; (b) accelerometers and yaw rate sensors to study drivers' braking, acceleration and steering behavior; (c) GPS to capture the vehicle's position; and (d) camera-based traffic sign recognition. Eye-trackers have also been used in NDD collection, but it has proven difficult to achieve sufficient quality in the capture of driver's gaze direction in naturalistic settings (Victor et al., 2010; Victor et al., 2015). Although CAN may include radar, researchers often obtain range and range-rate data from add-on sensors (e.g., radar), because vehicle manufacturers (or their suppliers) may not be willing to share the CAN data from radar (Blatt et al., 2015). If high-quality range and range rate data to the lead vehicle are unavailable, applying the methods of Papers II, IV and V is more complicated. Fortunately, Paper III provides a method for addressing the lack of such data. The application of the method in Paper I often requires data already available on CAN; otherwise, external sensors must be added to collect the data.

The data from commercial NDD (see Papers II and III, and Section 4.3) are usually limited to one or two video cameras (one forward and one on the driver), two or three accelerometers, and GPS (Lytx, 2016; SmartDrive, 2016).

It is also possible to collect NDD without video, capturing only the vehicle speed and position (e.g., GPS; Mononen et al., 2012)—as is the case with traditional event data recorder (EDR) data, which are included in the NDD category. According to the US national highway traffic safety administration (NHTSA), “An **event data recorder** (EDR) is a function or device installed in a motor vehicle to record technical information about the status and operation of vehicle systems for a very brief period of time (i.e., a few seconds) and in very limited circumstances (immediately before and during a crash), primarily for the purpose of post-crash assessment of vehicle safety system performance.” (NHTSA, 2012a, p. 74145). EDR data can be used as input when applying the methods presented in Papers IV and V.

In NDS and NFOT, **questionnaires** (an example of a self-report tool) are often administered to participating drivers to complement in-vehicle sensor data. In addition to standard questionnaires collecting background information about the drivers (e.g., demographics), commonly used questionnaires include (a) the Driving Style Questionnaire, DSQ (French et al., 1993) and (b) the Manchester Driving Behavior Questionnaire, DBQ (Parker et al., 1995; Reason et al., 1990). They are not explicitly a part of the methods in Papers I-V.

Data from external sources can also be used to complement NDD; for example, **map data** (e.g., road type and number of lanes) can be obtained via map matching through GPS positions or comparison with crash databases with respect to location (LeBlanc et al., 2006; Victor et al., 2010). Applications of the method in Paper I in particular are especially likely to utilize such data.

Having such a plethora of objective data can certainly be useful for traffic safety research. However, appropriate methods for NDD analysis are needed. In their review of two major categories of on-road studies for traffic safety research (controlled on-road experiments and field operational/naturalistic driving studies; see Section 1.1.1), Carsten et al. (2013) assess these methods’ benefits, drawbacks and complementary nature. In discussing NDD, they conclude: “Even though the increase in information density is promising, it is necessary to put effort into developing suitable methods, both for data extraction and data analysis.” (Carsten et al., 2013, p. 172).

2 Methods for analyzing NDD

2.1 Analysis of everyday driving

Most time spent driving is uneventful from a safety perspective (critical events are sparse). Most drivers quickly learn to adapt to the vehicle kinematics, other road users, and the environment in order to travel safely and comfortably. Understanding everyday driving is as important for the development of all types of ISS as it is for other safety measures (Lee, 2008).

Everyday-driving NDD consist of electronic records created while drivers of instrumented vehicles go about their everyday lives. There are two research areas that warrant special attention in the analysis of everyday-driving NDD. First, the data can be analyzed to quantify driving in a variety of ways, including (a) establishing **distributions of driver performance metrics** (LeBlanc et al., 2006; Peng et al., 2013; Sayer et al., 2007), (b) categorizing drivers' behaviors into different **driving styles** (Fancher et al., 1998; Moeschlin, 2007; Sagberg et al., 2015; Yurtsever, Miyajima, Selpi, & Takeda, 2015), and (c) identifying **risk factors** in everyday driving (Barnard et al., 2015; Stutts et al., 2005). Second, everyday-driving NDD are commonly used when evaluating the benefit of in-vehicle ISS in NFOT, by comparing performance metrics from (between): the baseline/control condition when the ISS is inactive and/or not available to the driver, and the treatment condition when it is available (Bao et al., 2012; LeBlanc, Bao, Sayer, & Bogard, 2013; LeBlanc et al., 2006; Sayer, Mefford, Shirkey, & Lantz, 2005b). NFOT is often used to address both of the research areas. For example, studies using NFOT typically include both analyses of driving styles and estimations of safety benefits (Fancher et al., 1998; LeBlanc et al., 2006). In both of these research areas, result stability can be improved by checking for autocorrelation and calculating the performance metrics in a way that minimizes bias. Paper I seeks to do just this.

Strategies for the collection of NDD vary from continuous to the selective collection of short segments. The former retains data from the time the vehicle is started until it is turned off. The latter applies some sampling criteria, such as the acquisition of random segments (e.g. 20 s) of in-vehicle data (Aksan et al., 2013), or the acquisition of data within a fixed distance (e.g. 50 m) of a specific intersection or curve (Gordon et al., 2012; Smith et al., 2009). Both collection strategies are interchangeable for many types of analyses, but some analyses are intrinsically impossible with selective (short segment) collection (e.g., analysis of driver behavior adaptation as a function of context, from start to completion of secondary tasks; Tivesten & Dozza, 2014a). In the early years of NDD collection, data storage capacity was an issue, so the large storage requirements of continuously collected data were a disadvantage. However, with the technological advances in the last decades, this is much less of a limitation (Victor et al., 2010). Today the cost of installation and equipment is typically the limiting factor, and not storage space. Thus if a vehicle is instrumented for NDD collection, the data is typically collected continuously (Sayer et al., 2008; TRB, 2014).

Depending on the everyday driving research area and the strategy for NDD collection, different issues exist in the data analysis. However, basically all analyses share the issue of (some) bias. For example, when analyzing everyday-driving NDD (as well as the equivalent in controlled driving-simulator or on-road studies), it is common to apply some **statistic** (e.g., standard deviation) to segments of data—one example being the calculation of standard deviation of lane position (SDLP; Green, Cullinane, Zylstra, & Smith, 2004). Depending on the segment selection and how sets of segments are aggregated (e.g., means across segments), biases can be introduced. Such biases include **segment-size bias** and **autocorrelation bias**. Segment-size bias can occur when a statistic sensitive to segment size is applied to segments of unequal size, and aggregation is performed across the segments (e.g., mean of the calculated standard deviations (Paper I, and indirectly described in work by Östlund et al. (2004)). Similarly, autocorrelation bias is introduced when samples that are assumed to be independent (e.g., as a prerequisite for a specific statistical analysis) are not actually independent (Box, Jenkins, Reinsel, & Ljung, 2016; Sayer, Devonshire, & Flannagan, 2005a; Paper I). For example, a car's speed on a highway in one instance is highly correlated with its speed in the next few hundred milliseconds, or even the coming seconds and minutes. Few studies acknowledge and address these biases (e.g., Peng et al., 2013; Sayer et al., 2005a; Östlund et al., 2004). Paper I provides a method to address size bias and highlights the importance of controlling for autocorrelation.

2.2 Understanding crash causation

There are two primary approaches in the literature that seek to establish causation from observational data. The first is the **qualitative expert assessment** of observations. The second applies statistical methods such as descriptive statistics, odds ratios, logistic regressions, and induced exposure to the observations, and is typically followed by **scientific reasoning** on crash causation. As both approaches rely on observational data, neither can appeal to the tradition of controlled experimentation (truly randomized, controlled experiments) to claim causation. The issue of inferring causation based on NDD (observational data) is the topic of Section 4.2.

The main difference between (a) qualitative expert assessment and (b) statistical approaches to studying crash causation is that, in the former, experts assess the available information subjectively, identifying causal factors according to some pre-defined process; in the latter, statistical methods are applied to the data at hand, without explicit (subjective) expert assessment. Expert-assessment-based studies typically involve at least one expert in traffic safety who performs the analysis of a set of crashes (or, in some cases, near-crashes;Habibovic et al., 2013). Causation is often studied by organizing causal (or, as others may phrase it depending on their view of causation, 'potentially contributing factors'; see Section 4.2 for a discussion on causation) factors into **causal chains** or some other tree or chain structure (Chang & Wang, 2006; Elvik, 2003; Paper II; Habibovic et al., 2013; Otte, Pund, & Jänsch, 2009; Sandin & Ljung, 2007; Warner & Sandin, 2010).

To date, NDD is seldom used as a basis for expert-assessment crash-causation studies. Notable exceptions are Paper II of this thesis, as well as the work by Habibovic et al. (2013) and Dunn et al. (2014). The Dunn report documented an expert-assessment method for crash causation that aimed "to investigate the crash trifecta concept to determine if the convergence of multiple elements, rather than a single, unitary critical reason, has greater value in explaining the complexities of crash genesis" (p. i). These are believed to be the only three expert-assessment-based crash-causation studies to date which perform a detailed analysis of NDD, including event video—and only Paper II has been applied to commercial NDD.

The second approach to studying crash causation—using direct statistical methods instead of expert assessment—can be further subdivided into studies that use (1) **descriptive statistics** only, when no controls are available with which to contrast the crashes or other SCE (Carney et al., 2015); or (2) **inferential statistics** to establish an association between specific factors and risk (e.g. when controls are available with which the crashes or other SCE can be contrasted; Hallmark et al., 2015; Klauer et al., 2014; Victor et al., 2015). Many studies in this second category make inferences about causation based on the established associations and scientific reasoning. An example of one such method is **induced exposure** (Hautzinger, Pastor, Pfeiffer, & Schmidt, 2007; Lardelli-Claret et al., 2005) which does not use controls in the traditional sense. Instead, some variable that is assumed to be insensitive to what is being studied is used to manage exposure. For example, vehicles/drivers that are categorized as not-at-fault in crashes are used as the controls in a case-control like manner, with the at-fault vehicle/drivers under study as the cases (Lardelli-Claret et al., 2005).

It is important to note, however, that all the analysis methods described in this section have the same basis—their data are observational, so hidden (and uncontrolled) factors may be **confounding the results** (Carsten et al., 2013; Rothman, 2012). For example, talking on the phone while driving has, in studies of NDD (Fitch, Hanowski, & Guo, 2014; Klauer et al., 2014), not been shown to have adverse effects on safety in a rear-end crash scenario. Possibly even more surprisingly, NDD studies have even indicated protective effects of talking on the phone in this scenario (Hickman et al., 2010; Victor et al., 2015). Some researchers have questioned these results, using arguments such as: "[I]t could be possible that drivers only use their telephone in low-risk situations. There might also be an intermediate process, in the sense that drivers who use a mobile phone are aware that they need to exert extra effort, and thus, they even overcompensate by being extra attentive.” (Carsten et al., 2013, p. 167; see also Strayer & Cooper, 2015). This driver adaptation could, potentially, be confounding results. The implications of studies indicating non-risky, or even protective, effects of some behaviors when they are genuinely risky could be devastating to, for example, policy making. However, some researchers argue that it may be vice versa: that established methods (e.g., the way driver reactions are quantified, Markkula et al., 2016) and/or concepts of cognitive load (highly relevant in phone talking) may in fact be incomplete (Engström, Markkula, Victor, & Merat, 2016b), rather than the NDD studies being erroneous.

Another important consideration when seeking to understand crash causation is whether near-crashes (or other non-crash events) should be used as surrogates for crashes in the analysis. Because crashes are rare, the focus of crash-causation research in general (as well as, e.g., in safety benefit evaluations of ISS through NFOT), has often been on using crash surrogates, such as near-crashes, instead, in order to have more data for analysis. The logic to this approach is a **hypothesized relationship between non-crashes and crashes**. Traffic-conflict theory research (Hydén, 1987; Migletz, Glauz, & Bauer, 1985; Svensson, 1992; Svensson, 1998) and the following studies based on NDD (Guo & Fang, 2013; Paper V; Guo, Klauer, McGill, & Dingus, 2010; Wu & Jovanis, 2012) support the existence of this relationship.

There are different ways to define crash surrogates (which include near-crashes) used in crash-causation analysis. As one example, classic conflict-theory researchers primarily classify non-crash events based on human estimates of **time-to-accident** (Svensson, 1998): “Time-to-Accident is the time that remains to an accident from the moment one of the road users takes evasive action calculated assuming that they otherwise had continued with unchanged speeds and directions” (p. i). In contrast, a second, different definition of crash surrogates is used in most NDS (Klauer, Perez, & MacClaggerty, 2011; SHRP2, 2012; Utesch et al., 2014). They typically use a two-stage process to extract safety-critical events (SCE; Utesch et al., 2014)—a term commonly used in NDS literature to describe crash surrogates. The first stage is automated identification of potential conflicts (SCE candidates) using algorithms applied to sensor data (i.e., algorithms that identify deceleration, time-to-collision, or time-to-lane-crossing), and the second is a visual review of videos of the situation, performed according to a set coding schema (SHRP2, 2012). Part of the NDS definition of SCE is often in line with that used by conflict theory (the classical conflict; Svensson, 1998), but other parts are not (e.g., defining a run-off-road or a tire-strike as a crash in SHRP2 (2012)). Current work (in ISO WG5/SC39) aims to improve and harmonize the definition of SCE across NDS by providing a conceptual framework and operationalizing different classes of SCE. See Section 4.1 for a discussion on the use of near-crashes as surrogates for crashes in NDD research.

2.3 Driver behavior in prospective safety-benefit estimates

There are several different methods available for estimating the safety benefit for ISS and other safety measures. This section describes **prospective** methods: they estimate the future benefit of safety measures (e.g., ISS) before they have been deployed (i.e., systems in development, or prototype systems—when statistical data from on-road crashes are not available). In contrast, many benefit estimates are **retrospective**: they estimate the safety benefit for safety measures that are already deployed (on-market). The relevant statistical data for retrospective benefit estimation is typically found in crash databases (NHTSA, 2015) or insurance data (Cicchino, 2016; Isaksson-Hellman & Lindman, 2012), which are not available for prospective benefit estimation. Retrospective safety benefit analyses are excluded from the scope of this thesis.

Examples of methods for prospective safety-benefit estimates include expert-assessment-based methods using in-depth crash data (Strandroth, 2015a; Strandroth, 2015b), combinations of expert assessment and test-track experiments (Lesemann et al., 2011), combinations of test-track experiments and dose-response methods (Bálint, Fagerlind, & Kullgren, 2013), naturalistic field operational tests (Ljung Aust et al., 2011; Sayer et al., 2011), and virtual (counterfactual) simulation-based approaches (Georgi et al., 2009; Lindman & Tivesten, 2006; McLaughlin, Hankey, & Dingus, 2008; Paper IV; Scanlon, Sherony, & Gabler, 2016).

One of the main reasons for using virtual simulations for prospective safety benefit estimates of ISS is to reduce the number of expensive and time-consuming physical experiments in favor of less expensive, easily repeated simulations in a virtual environment (Page et al., 2015). As a testament to the importance of using virtual simulations, developers of passive (injury prevention) safety systems—safety measures to mitigate the consequences of crashes once they have happened—have used mathematical simulations for decades (Prasad & Chou, 2002; Yang et al., 2006). In fact, the automotive industry today considers mathematical simulations of vehicle crashes an absolute necessity for timely and efficient vehicle development. These simulations save money and time by facilitating system optimization and prospective benefit evaluation without the need to crash hundreds of vehicles and vehicle components in the design phase. Developers of passive safety systems rely on models of either crash test dummies (Foster, Kortge, & Wolanin, 1977; Svensson & Lövsund, 1992), or, more recently, finite-element human body models (HBM; Brolin, 2016; Östh, 2014). Typically, researchers in academia develop the dummies and models together with industry. However, to develop and evaluate ISS, which act before the crash, it is not enough to model the physical driver. Mathematical models of driver behavior (Jagacinski & Flach, 2002; Lee, 1976; Lee, 2008; Markkula, Benderius, Wolff, & Wahde, 2012) are needed, rather than mathematical models of the physical driver (i.e., crash test dummies or humans). Further, when developing passive safety systems, the crash is a given. When evaluating ISS, however, the aim is to avoid crashes, or at least mitigate their consequences. This is also the case when the evaluation is performed through mathematical simulations.

Counterfactual simulations for pre-crash safety-benefit analysis are a specific type of virtual simulation: they create events that didn't actually happen, using mathematical simulations based on time-series data from actual real-world events. Typically, counterfactual simulations compare 'what-if' scenarios—events as they would occur with and without ISS (algorithms and virtual actuators). Crashes are usually used (Georgi et al., 2009; Lindman & Tivesten, 2006; Papers IV and V; Scanlon et al., 2016), while near-crashes can also be used (McLaughlin et al., 2008; Paper V). Even everyday-driving normal (lead-vehicle braking) events have been used as a basis for counterfactual simulations (Woodrooffe et al., 2012). Each simulation produces a specific outcome: either a crash (with a specific outcome; e.g. delta-v; Lindman & Tivesten, 2006) or no crash. Paper IV demonstrates how the choice of driver models affects the results, even when all other variables are unchanged. Paper V, however, is different from the typical application of counterfactual analyses—it does not aim to evaluate the safety benefit of an ISS. Instead, Paper V shows how counterfactual simulations can be used to evaluate driver behaviors with respect to safety.

In general, the different approaches of prospective safety-benefit estimation use different types of data and take driver behavior into account in different ways. In FOT and NFOT studies, which often implicitly include driver behavior in the analysis, the actual behavior of the driver is overtly or covertly recorded in the data on which benefit estimations are based. However, counterfactual simulations documented in the literature have typically used simplistic models of driver behavior without considering the implications of the choice of model on benefit estimation (which Paper IV does). Models of driver behaviors are becoming increasingly important—at least in ISS safety measures when the driver is in the loop (Paper IV)—as the automotive industry relies increasingly on counterfactual simulations to study ISS. A testament to the increased focus on counterfactual simulations is the self-funded (members contribute with in-kind time and resources only) European consortium Prospective Effectiveness Assessment for Road Safety (P.E.A.R.S.; Page et al., 2015). It consists of a large number of automotive companies, institutes, governments, and academic institutions developing a "comprehensive, reliable, transparent, and thus accepted methodology for quantitative assessment of these systems by virtual simulation." (Page et. al, p. 1).

While industrial stakeholders aim to achieve pragmatic implementations of counterfactual simulations for the evaluation of ISS, researchers have recently proposed a theoretical framework for counterfactual simulations (Davis, Hourdos, Xiong, and Chatterjee, 2011). They represent crashes and conflicts as the results of the interactions between initial conditions (e.g., kinematics) and evasive actions. The overall statistical framework is also applicable when evasive actions are determined by an ISS rather than a driver. Davis et al. acknowledge the need to understand (a) actual pre-crash kinematics and (b) evasive actions, while highlighting the importance of understanding sampling bias: “Basically, studies of non-crash events can be used to construct proxies for crash probabilities, but the sampling bias that results when attention is restricted to non-crash events can forestall the development of simple predictive relationships.” (Davis et al., 2011, p. 1916).

Understandably, the automotive industry focuses less on academic publishing and theoretical frameworks than on their practical implementations (Page et al., 2015). To date, only a limited number of scientific publications present the method of counterfactual simulation and/or describe its use for estimating the safety benefit of ISS (Erbsmehl, 2009; Funke, Srinivasan, Ranganathan, & Burgett, 2011; Georgi et al., 2009; Kusano & Gabler, 2010; Lindman & Tivesten, 2006; Markkula, 2015; McLaughlin et al., 2008; Scanlon et al., 2016; Woodrooffe et al., 2012). Further, a literature review found only two studies using NDD as the source for the kinematics in the counterfactual estimation of prospective safety benefits: McLaughlin et al. (2008) used data from the 100-car NDS (Neale et al., 2005), while Scanlon et al. (2016) used event data recorder (EDR) data. However, there are few articles actually utilizing counterfactual simulations in which the model of driver behavior is addressed—beyond, for example, the application of simple reaction times and constant decelerations as evasive actions. Exceptions include a recent PhD thesis by Markkula (2015) and Paper IV, which argues for the need for validated driver models to achieve good safety benefit estimates using counterfactual simulations.

Although the scope of this thesis excludes the explicit development of mathematical models of driver behavior, they are intrinsic to counterfactual simulations in which the ISS require drivers’ reactions (Markkula, 2015; Paper IV), or when the effect of the driver behavior itself is being studied (Paper V). However, the term ‘driver model’ has many different meanings in the literature. Markkula (2015) categorizes driver models into three main groups: conceptual, statistical, and process models. Examples of **conceptual models** include those that outline how drivers adapt to risks in driving (Bärgman et al., 2015; Näätänen & Summala, 1974; Vaa, 2007; Wilde, 1982), and those that describe how drivers interact with information and the world (e.g., predictive processing, Engström et al., 2016a; hierarchical models of driving, Michon, 1985; and information processing,Wickens, Hollands, Banbury, & Parasuaman, 2016). **Statistical models**, on the other hand, are statistical descriptions of driver behavior, with distributions of reaction times (Green, 2000; Olson & Sivak, 1986) as one of the most commonly used statistical models. Other statistical driver models describe drivers’ safety margins (e.g., comfort-zone boundaries; Bärgman et al., 2015; Lübbe, 2015). These models can also be components of ISS algorithm implementations (Brown, Lee, & McGehee, 2001). Finally, **process models**, in Markkula’s (2015) terminology, are models that allow for computer (mathematical) simulation: input is typically processed moment by moment (using current or historical data), and an action or actions are performed based on the input. Process models of driver behavior originate in different research domains (e.g., neuroscience, Benderius (2014); and control theory, Jagacinski and Flach (2002)). These models describe driver behaviors at different levels of abstraction, from body movement (Georgopoulos, 1986) to lateral (Fajen & Warren, 2003; Salvucci & Gray, 2004) or longitudinal (Lee, 1976) control while driving, for example. Note that while these three groups of models are different, they are not mutually exclusive. Many of the conceptual models have been developed into process models (actually, most process models are based on conceptual models), implemented algorithmically for use in a computer simulation (Salvucci, 2006). Statistical models, too, are often an integral part of process models (e.g., Paper IV and any process model of driver behavior that includes reaction time distributions). In addition, there are models of driver behavior that arguably do not fit into Markkula's three categories—for example, when the models are implicitly constructed as part of driver-behavior detection systems (Al-Sultan, Al-Bayatti, & Zedan, 2013; Kuge, Yamamura, Shimoyama, & Liu, 2000). A review of models of driver behavior in critical situations can be found in the 2012 paper by Markkula et al.

Driver behavior models to be applied to counterfactual simulations can be used in two different ways: for pre-crash event generation and for simulated driver **reaction**. **Pre-crash event generation** is the application of counterfactual drivers' behaviors to pre-crash kinematics time-series (typically) from real-world events (crashes or near-crashes), in order to produce counterfactual pre-crash event time-series data (see Figure 4). These data are then used in the counterfactual simulations as if they were recorded data from real events to provide counterfactual events during the pre-crash phase. Pre-crash event generation is thus applied in the preparation phase of counterfactual simulations (Figure 4). **Simulated driver reaction**, on the other hand, is the application of process models of driver reaction to some stimuli, to simulate when and how the driver reacts (Figure 5). For all ISS where the driver is to react to an ISS (e.g., forward collision warning), the simulated driver reaction is at the core of the counterfactual simulations. Both pre-crash event generation and simulated driver reaction can use both statistical and process models (Figure 4 and 5, and Papers IV and V).

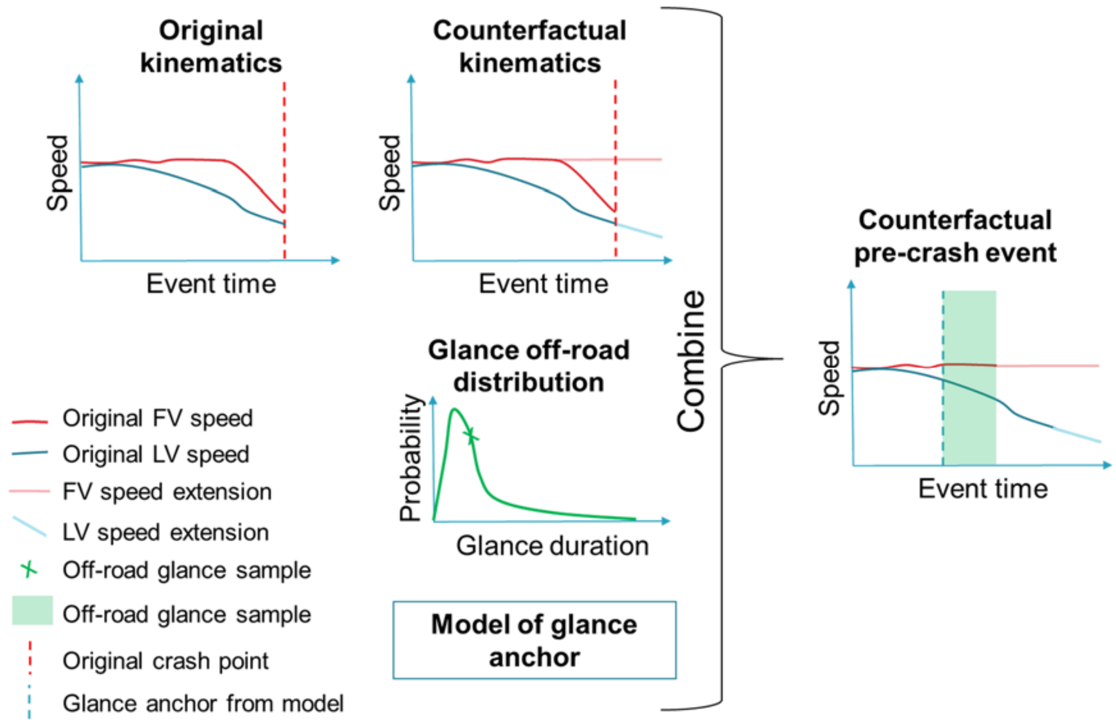

Figure 4: A demonstration of pre-crash event generation where counterfactual glance behavior is applied to original pre-crash kinematics (as in Papers IV and V). The original kinematics come from an actual crash (for demonstration purposes a sketch of an actual crash is used here). The counterfactual kinematics are created by finding the start of following vehicle's (FV) evasive maneuver (braking), after which the speed is set constant. The lead vehicle's (LV) deceleration is extended beyond the crash point. By combining the counterfactual kinematics, a sample from a distribution of off-road glance duration, and a model a model of when that glance occurs, counterfactual pre-crash event data are created. This counterfactual pre-crash event data can then be used in a counterfactual simulation, as if both the counterfactual kinematics and the glance behavior had actually been recorded.

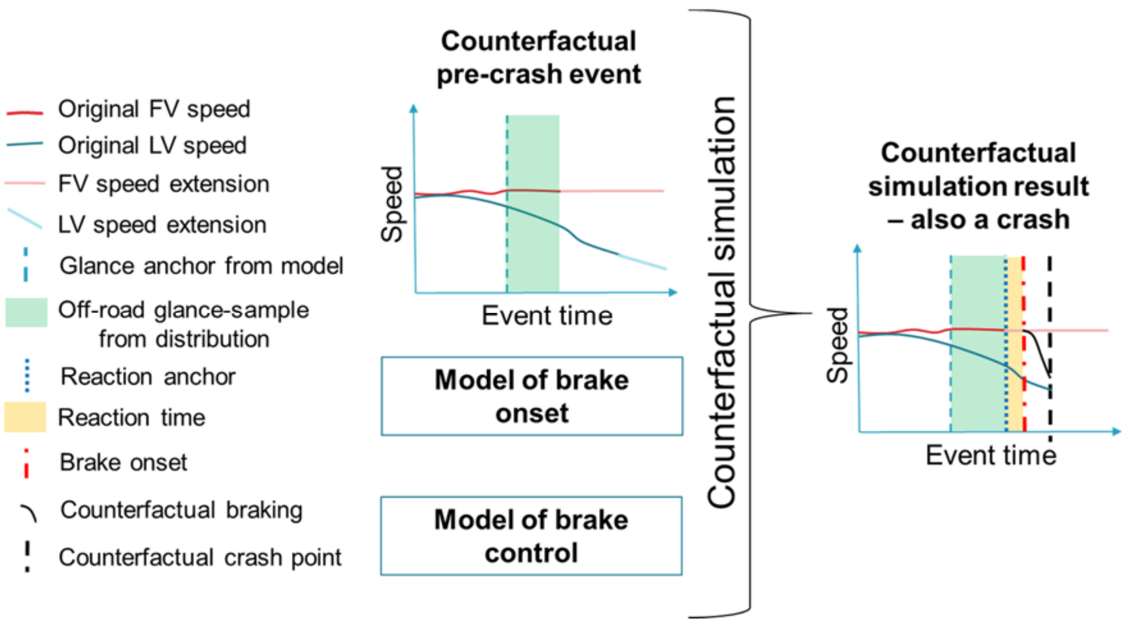

Figure 5: A demonstration of how the counterfactual pre-crash event data from pre-crash event generation (see Figure 4) is combined with a model of brake onset (when the driver

starts to apply the brakes; here, a constant reaction time added to the time of the glance back on road), and brake control (how the driver applies the brakes).

In addition to appropriate, validated driver models, other validated models are needed for the following: sensors (e.g., range and field-of-view limitations), the environment (e.g., friction, occlusion, and weather conditions), the vehicle (e.g., vehicle dynamics and actuators such as braking and steering systems), and ISS algorithms (e.g., their actual implementations). However, these models, typically more mature than models of driver behavior (Markkula, 2015), are beyond the scope of this thesis, as are the details of the development of the driver models used in the thesis.

2.4 Understanding the effect of driver behavior on safety

Research seeking to understand how driver behavior affects traffic safety is multi-faceted and dates back many decades (Gibson & Crooks, 1938; Lee, 1976). As a result, methods for estimating the risks related to drivers' behaviors vary widely. The methods include using hypothesized relationships between specific performance metrics and crash risk from controlled experiments and NDD (e.g., standard deviation of lane position and reaction times after the precipitating event; Boyle & Lee, 2010; Engström, Aust, & Viström, 2010; Engström et al., 2016b; Gordon et al., 2009; Green et al., 2004). Additionally, simulator and test-track studies have compared, for example, have compared an outcome metric (e.g., the number of crashes or near-crashes) or a severity metric (e.g., time-to-collision; SAE, 2015) between control and treatment phases in empirical studies that impose different tasks for the driver (Horrey, Lesch, & Garabet, 2009; Lee, Caven, Haake, & Brown, 2001; Markkula, Benderius, Wolff, & Wahde, 2013; Strayer et al., 2015). Furthermore, as NDD have emerged, researchers have applied different epidemiological approaches to NDD (e.g., comparing talking on the phone or texting to baseline in a case-control study; Hickman et al., 2010), and in-depth analysis of time-series pre-crash data (Victor et al., 2015). The most common epidemiological approach for studying the effect of driver behavior on safety that has been applied to NDD is the use of (crude) odds ratios or logistic regression to calculate, for example, the odds of different factors (Hickman et al., 2010; Klauer et al., 2014; Victor et al., 2015). A key aspect to minimizing bias in such analyses is the selection of appropriate controls to be compared with the crashes and near-crashes (Kidd & McCartt, 2015; Victor et al., 2015).

Observations and epidemiological approaches based on NDD are only useful if the specific tasks to be studied are actually available in the data. Drivers are constantly bringing new devices into their vehicles (e.g., smart phones with different apps) and vehicle designers develop new **driver-vehicle interfaces** (DVI; e.g., new infotainment or voice-control systems, Merat, Jamson, Lai, Daly, & Carsten, 2014; Reimer & Mehler, 2013). As a result, the risks associated with these novel interactions also need to be constantly evaluated and updated. The evaluations can in turn inform appropriate legislation, develop governmental design guidelines (European_Communities, 2007; NHTSA, 2012b), and enable the vehicle industry to develop safe and sound DVI (Merat et al., 2014; Park & Kim, 2015). For example, the US NHTSA has developed visual-behavior guidelines for evaluating DVI (NHTSA, 2013), which include (but are not limited to) driving simulator studies to acquire drivers' glance behaviors related to the DVI being evaluated. The glance behavior data are then evaluated with respect to a set of performance indicators. The results of the evaluation inform the decision to pass or fail the DVI with respect to glance behavior. It should be noted that the NHTSA guidelines and corresponding simulator study procedures have been criticized for being too sensitive to study design and sub-optimization (Heinrich, 2015; Ljung Aust, Dombrovskis, Kovaceva, Svanberg, & Ivarsson, 2013; Rydström, Ljung Aust, Broström, & Victor, 2015).

Using pre-crash event generation and simulated driver reaction (see Section 2.3), Paper IV presents a new method that can be used by designers of DVI. The method could also be an alternative (or, likely, a complement) to that which is proposed in, for example, NHTSA guidelines. Using this method, mathematical (virtual) counterfactual simulations can evaluate drivers' glance behaviors during specific tasks, perform prospective risk analyses on the introduction of specific DVI, and evaluate other tasks which drivers perform in vehicles. Much of the literature for counterfactual simulations for ISS is also relevant for these (driver-behavior-focused) counterfactual simulations. However, no other counterfactual method has been found in the literature which uses real-traffic crash and near-crash data as a basis for estimating the risks associated with specific driver behaviors.

2.5 Aim and scope

2.5.1 General aim

This thesis aims to provide and demonstrate methods for analysis of NDD to understand and evaluate driver behaviors in relation to traffic safety and crash causation. These methods can build a foundation for the development of pre-crash traffic safety measures, as well as enable the evaluation of the effect of driver behavior on safety on, for example, the design of driver-vehicle interfaces.

2.5.2 Specific aims

The methods presented and demonstrated in this thesis aim to enable researchers to:

- obtain stable, reliable and comparable results in the analysis of continuous NDD (Paper I).
- identify crash-causation factors/mechanisms through a structured expert-assessment using NDD (Papers II-III).

- calculate pre-crash interaction kinematics (e.g., range and range rate) and optical parameters (e.g., optical expansion rate and tau) in NDD rear-end crash and near-crash data (Paper III).
- realize the importance of selecting appropriate driver models in counterfactual safety benefit evaluation of intelligent systems for traffic safety (paper IV).
- calculate the effects of different driver behaviors on traffic safety through counterfactual simulations using NDD (Paper V).
- compare different driver-vehicle interfaces with respect to safety (Paper V).

In addition to addressing these aims, the thesis discusses three topics relevant to NDD analysis:

- The use of near-crashes as surrogates for crashes
- The use of statistics and expert assessment to infer causation when analyzing NDD
- The use of commercially collected NDD.

2.5.3 Scope

The research conducted in Papers I-V encompasses a wide range of NDD analyses and a variety of NDD sources. The scope is thus relatively broad. However, it is somewhat narrowed by the application of the methods primarily to rear-end crash scenarios in which an NDD-instrumented following vehicle follows and impacts (or could have impacted) a lead vehicle. (This does not, however, rule out the application of these methods to other crash scenarios.) The methods include a primarily quantitative analysis of NDD data (Papers I and III-V) and analysis using expert assessment (Paper II), but exclude qualitative analyses based on subject self-reporting (e.g., questionnaires). Explicit mathematical modeling of driver behavior is also excluded from the thesis, although the use of such models is a prerequisite for the successful application of the methods presented in Papers IV & V.

3 Summary of papers

3.1 Paper I

Title: Chunking: A procedure to improve naturalistic data analysis
Objective: To develop and present a method, *chunking*, that facilitates robust and comparable results in the analysis of continuous NDD.
Background: At the time of publication (2013), NDD had been available for only a limited time and to only a few researchers. Methods to facilitate appropriate analysis are needed now that NDD are beginning to be more readily available. Traditional methods neglect the impact of segment length and its potential to significantly bias results. Segment length is the section of continuous data upon which a statistic (e.g. mean or standard deviation) is applied to extract a performance indicator (e.g. standard deviation of lane position).
Method: The chunking method reduces this source of bias. It is appropriate both for basic driver behavior research and for the development and evaluation of Advanced Driver Assistance Systems (ADAS); in fact, it was developed and first applied in the evaluation of ADAS, applying chunking to segments of baseline and treatment alike. Chunking is designed to be used when the aim of analysis is to create aggregate measures of continuous NDD across trips or conditions (e.g., all data above 70 km/h). Chunking divides the datasets into equivalent subsets of data (called chunks) before other calculations are applied. The main benefit of the method is the robust and consistent calculation of parameters when analyzing continuous NDD, although care has to be taken to account for sample dependencies and auto-correlation. Examples of the application of chunking are presented, and results are compared with traditional methods of analysis.
Results: Large biases in statistical results with traditional methods were reduced using the chunking approach. For example, in a set containing 399 driving hours of NDD, 75% <u>of the data</u> were in segments longer than 100 s although 75% <u>of the segments</u> were shorter than 100 s. The choice of appropriate chunk size was a key factor in obtaining robust results.
Conclusion: Prior to this paper, little research had been aimed specifically at addressing the methodological issues of analyzing continuous NDD. The results show that neglecting chunking can introduce large biases in statistical results. Using chunking is advisable in many cases, but the effects of auto-correlation and sample dependencies must be still be considered.
Application: Chunking is appropriate wherever aggregation across segments in NDD are being considered, particularly when extracting an indicator (e.g. standard deviation) that is sensitive to segment length.

3.2 Paper II

Title: Analysis of the role of inattention in road crashes based on naturalistic on-board safety monitoring data

Objective: To investigate the role of driver inattention in crossing-path intersection crashes and rear-end crashes.

Background: The role of drivers' inattention in traffic crashes is not well understood. Data from traditional methods (e.g., post-crash reconstructions) do not have the requisite fidelity in the pre-crash phase to allow detailed analysis of driver inattention. NDD (e.g., time-series data with video) have emerged as a promising source for studying this phase; however, most naturalistic driving studies record only a few crashes, if any. In contrast, quite a few actual crashes are recorded by on-board safety management systems (OBSM), typically mounted in commercial vehicles and collecting GPS, accelerometer data, video of the forward roadway (and, often, video of the driver), for a period of time around crashes and other safety-critical events.

Method: A total of 133 real-traffic crash events involving vehicles instrumented with the DriveCam (now Lytx) OBSM system were analyzed. The events were drawn from two scenarios: 70 rear-end crashes where the OBSM-instrumented vehicle was the striking vehicle, and 63 crossing-path intersection crashes where the instrumented vehicle intended to drive straight through the intersection. Annotation of the 133 video sequences included (a) driver actions and visual behavior, (b) whether there was occlusion of the other (conflicting) vehicle, and (c) reconstruction of optical parameters (distance and relative speed between the two vehicles) in the rear-end events (as described in Paper IV). An event-coding scheme, partly based on work by Habibovic et al. (2013) and the EU-US inattention taxonomy group presented in Engström et al. (2013), was developed for the qualitative expert assessment of factors contributing to the crashes. The experts applying the scheme used both qualitative information about the event (e.g. video and narratives) and quantitative data extracted from accelerometers, GPS and video annotation. When all 133 events had been coded, all codings were aggregated.

Results: Inattention, especially in terms of eyes off the forward roadway, was identified as a primary contributing factor to rear-end crashes. In contrast, for intersection crashes (where the study vehicle was going straight), occlusion of the conflicting vehicle and insufficient safety margins were identified as key factors.

Conclusion: Commercially collected event-based data (with video) can provide an unprecedented, detailed view of the unfolding of the few seconds before a crash, providing new information about pre-crash behavior. A key conclusion is that the role of inattention as a factor contributing to crashes strongly depends on the scenario and crash type.

Application: A deeper understanding of the role of inattention in crashes could improve legislation, as well as the design of behavior-based safety systems and other safety measures.

3.3 Paper III

Title: Using manual measurements on event-recorder video and image-processing algorithms to extract optical parameters and range

Objective: To develop and evaluate a method for extracting optical parameters related to a lead vehicle, from the perspective of the driver of the following vehicle. The method is applied to video collected using event-based NDD.

Background: Traditionally, research into crash causation has primarily focused on post-crash reconstruction, using interviews and controlled follow-up experiments to study suspected crash causation factors and mechanisms. Commercially collected NDD with video have recently become available to researchers. Such data permit the analysis of time-series data with video from the pre-crash phase of relatively large number of actual crashes. These analyses will probably facilitate further understanding, both qualitative and quantitative, of the factors and mechanisms contributing to crashes.

Method: Manual measurements obtained through annotation of commercially collected NDD video were used to reconstruct optical parameters and range to a lead vehicle, as seen from a following vehicle. Video data of the rear of a stationary passenger car of known width were collected for a set of 14 different ranges, by means of the forward-looking camera of a Lytx (2016) (previously DriveCam) system. The optical parameters of the camera were extracted, and a model was created that rectifies the intrinsic distortion in the image. For method validation, twenty participants manually measured the on-screen size of the vehicle, and the errors between coders for different ranges were analyzed. The model was used to predict the range between the cars, and the results were compared to the actual range data.

Results: The results indicate that the method is useful when the ranges between the two vehicles are relatively short: when they are less than 10 m apart, the range estimate is within 10cm of the actual range. Incorrect estimation of the lead-vehicle width contributes significantly to range estimation errors. For optical parameters, on the other hand, optical rectification errors are likely to be the main source of error.

Conclusion: Parameter estimates using the proposed method are good for short ranges, while at longer ranges it should be used with care. The method is relevant for event-based NDD with video when image processing competence and tools are not available, and could be used as validation for other methods.

Application: This method could play an important role in research exploring why lead-vehicle conflicts occur, and thus in the development of safety measures to reduce the number and severity of rear-end crashes.

3.4 Paper IV

Title: Counterfactual simulations applied to SHRP2 crashes: The effect of driver behavior models on safety benefit estimations of intelligent systems
Objective: To demonstrate (a) the importance of the choice of driver behavior model in safety benefit estimations using counterfactual simulations, and (b) how counterfactual simulations can be used for parameter-sensitivity analysis.
Background: With the rapid increase in the development and deployment of pre-crash safety measures, the need for computer-based simulations to evaluate the potential benefit of pre-crash safety measures is also increasing. The counterfactual simulations compare the results obtained with and without a specific pre-crash safety measure applied to the pre-crash kinematics, obtained from one of a variety of data sources. These simulations require models of the relevant sensors, the vehicle, the driver, and the environment. However, to date, the models of driver behaviors included in such simulations have been rudimentary; furthermore, the effect of the choice of driver models has not been documented.
Method: Counterfactual simulations of a forward collision warning and an automatic emergency braking system were performed, using real-crash pre-crash kinematics from the SHRP2 naturalistic driving study as the basis. Three aspects of driver behavior were compared, with respect to the percent of crashes avoided by the FCW and the AEB and the estimated impact speed: glance off-road, reaction, and braking. Counterfactual pre-crash kinematics with glance off-road behavior were first created in a Monte Carlo fashion. Simulations were then run with and without FCW and AEB. Sensitivity analyses on driver behavior and FCW and AEB parameters were also conducted.
Results: The results reveal that the choice of driver behavior model has a substantial impact on the percentages of crashes avoided and the estimated impact speed for FCW. Changes to the distributions of driver off-road glances have less of an effect on the FCW safety benefit estimation than the positioning of those glances in relation to pre-crash kinematics. The effects of the driver model on the combined effect of FCW and AEB is small, while the proportion of FCW and AEB are radically different across different driver models and model parameter settings. The safety benefit estimations using counterfactual simulations for both FCW and AEB differ greatly from the retrospective analyses of FCW and AEB based on real-crashes. More research is needed to understand these differences.
Conclusion: Users of counterfactual simulations that include the driver in the loop must choose their driver models carefully when evaluating the safety benefits of pre-crash safety measures.
Application: Until driver models are good enough, estimates of the real-world benefits of pre-crash safety measures is likely to be error-prone. However, with validated driver models, safety measure development can be supported by using counterfactual simulations to compare algorithms and parameter settings.

3.5 Paper V

Title: How does glance behavior influence crash and injury risk? A 'what-if' counterfactual simulation using crashes and near-crashes from SHRP2
Objective: To develop and demonstrate a method for estimating the crash and injury risk of specific off-road glance behaviors, for both crashes and near-crashes; to provide a means for evaluating the safety impact of glance behaviors resulting from specific vehicle designs (e.g., infotainment systems) or secondary tasks (e.g., tuning the radio or texting on a mobile phone).
Background: The role of distraction as a cause of crashes has been more firmly established in the last few years. Meanwhile, the automotive industry regularly introduces new driver vehicle interfaces (DVI; also known as HMI), and new types of secondary tasks (e.g., text messaging) while driving appear and become prevalent. To guide DVI design, policies and legislation on, for example, distractions, evaluation needs to be done without waiting until post-hoc crash analyses can be performed. Evaluation methods can even become part of guidelines. Models of driver behavior in critical situations (e.g., glance behavior), together with pre-crash kinematics (e.g., from the SHRP2 NDD dataset used here), can facilitate the development of methods for evaluating the effect of driver behavior on safety. In addition, there is a current debate on the relevance of near-crashes for crash-causation research. Methods are needed that provide insights into the similarities and differences of crashes and near-crashes, asking questions such as: Are the initial conditions of pre-crash kinematics different between crashes and near-crashes, or are the drivers' actions the primary determinant of the actual outcome?
Method: This paper introduces a two-step approach to calculate model-based injury and crash risks through mathematical counterfactual (what-if) simulations. The method is demonstrated through its application on 37 lead-vehicle crashes and 186 lead-vehicle near-crashes from the SHRP2 naturalistic dataset.
Results: The main result of this paper is its method for evaluating the effect of glance behaviors on crash and injury. The analysis demonstrates how crash and injury risk can be calculated as a continuous metric across both crashes and near-crashes, for a given driver model. Another important result is the demonstration of the differential influence of percent-on-road glances, glance-off-road distribution, and total-task time on crash and injury risk. Insight into the kinematic similarities between crashes and near-crashes before a driver starts any evasive maneuver is yet another result.
Conclusion: Counterfactual simulations can be used to understand the effect of driver glance behavior parameters on safety. Near-crashes and crashes have similar pre-crash kinematics–it is primarily the actions of the driver that affect the outcome.
Application: The methods supports the tuning and safety optimization of in-vehicle interface designs, and provides a means for evaluating new secondary tasks (e.g., interacting with mobile devices).

4 Discussion

As naturalistic driving studies and other sources of NDD have become more numerous, their data have become available to a large community of researchers. In this thesis (and elsewhere, e.g., Dunn et al., 2014; McDonald et al., 2013; Sayer et al., 2007; Victor et al., 2015), new analysis methods have been developed to exploit this wealth of information in an ongoing effort to improve the design and evaluation of pre-crash safety measures. The development of new methods can also highlight the need to discuss and resolve some fundamental aspects of the associated research domain.

This chapter presents a discussion on three general topics of NDD (near-crashes vs. crashes, crash causation, and commercially collected NDD (CNDD)). These topics can be considered a foundation for the work in Papers I-V and this thesis. The topics are noteworthy because they highlight important issues surrounding the current and future use of NDD. In addition, the five methods associated with the analysis of NDD described in this thesis are related to each other and the literature. Benefits, drawbacks, and obstacles that remain are discussed in the context of the research gaps the methods address. First, however, the five included papers are framed in the context of the development process of pre-crash safety measures.

As stated in the aim (Section 2.5), the methods presented in Papers I-V can all contribute to the development of pre-crash safety measures, and Paper V can also contribute to the design of driver-vehicle interfaces (DVIs). Each paper addresses at least one part of the development process (Figure 6 and Table 1).

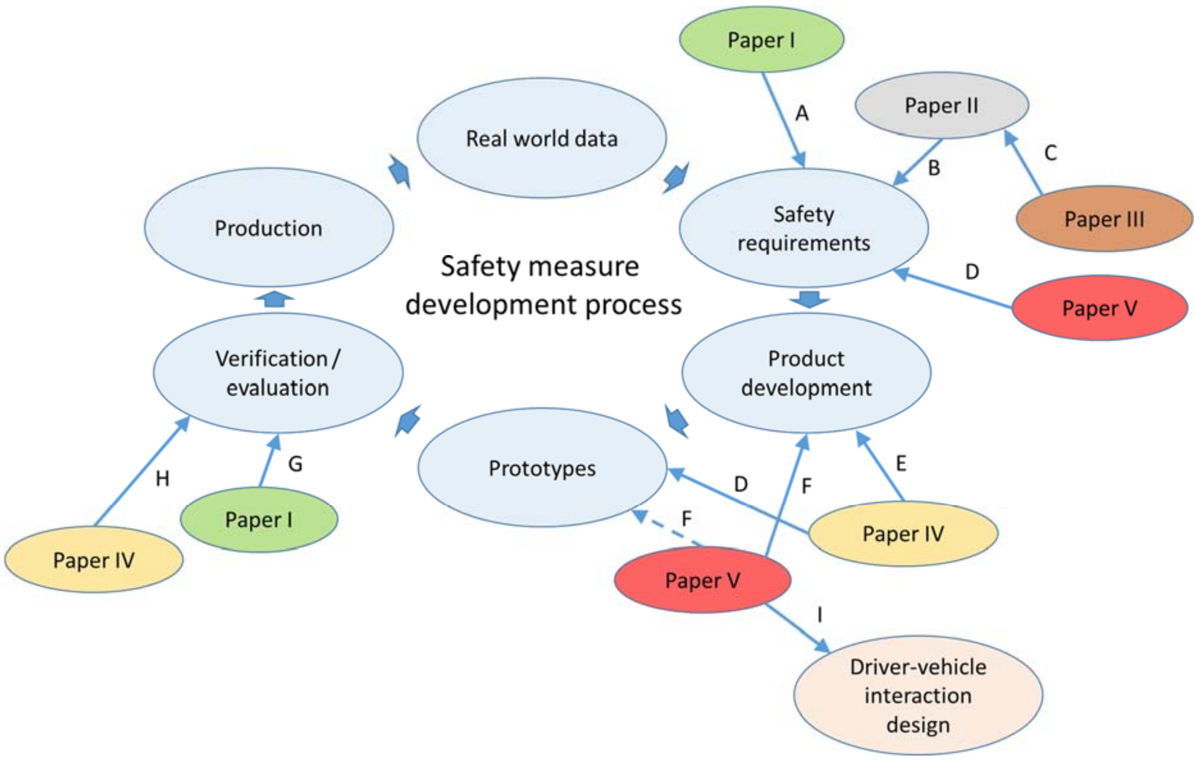

Figure 6: Papers I-V framed in the real-world-driven development process for safety systems used at the Volvo Car Group since the 1980s (Jakobsson et al., 2010). The light blue circles at the center describe Volvo's original development process. Paper V also supports the development of driver-vehicle interfaces (not necessarily related to safety measures). This is shown as a separate item (I; bottom). The letters A-I refer to Table 1, below.

Table 1: Each paper's contribution(s) to the safety measure development process above (Figure 6; A-I).

A. The method in Paper I can provide a more robust analysis of everyday driving NDD—when setting safety requirements, for example.
B. The expert-assessment method for crash causation in Paper II can help identify safety problems—a prerequisite to developing good safety requirements.
C. Paper III demonstrates how the method in Paper II can be applied to commercially collected NDD (CNDD) with video when other measures (range, range rate, and optical parameters) are unavailable.
D. The method in Paper V, if extended to other crash scenarios and refined for the rear-end scenario, could facilitate the development of more detailed safety requirements with respect to driver behaviors' effects on safety.
E. Paper IV demonstrates the importance of the correct models of driver behavior in counterfactual (mathematical) simulations of an ISS during initial development, as well as in the prototype phase.

F. The counterfactual simulation method in Paper V can be used to understand the effect of drivers' specific behaviors on safety during product and prototype development of, for example, automated vehicles—helping to speed up the development process.
G. The method in Paper I can improve the robustness of results when evaluating the effects of prototype safety measures on safety – chunking applies to segments of baseline and treatment alike.
H. Paper IV shows the importance of correct driver behavior models when performing counterfactual simulations to evaluate the effect of prototypes (or production systems) on safety.
I. The method in Paper V supports the development and evaluation of driver-vehicle interaction designs with respect to safety, and could also be used to refine glance behavior guidelines for driver-vehicle interactions.

The real-world data (top blue circle in Figure 6) are traditionally a variety of conventional crash data (e.g., in-depth crash investigations, or crash statistics). Different forms of NDD—including CNDD—can complement traditional data with more detailed information on the performance of future safety measures.

4.1 Near-crashes vs. crashes

The use of near-crashes as surrogates for crashes is the first of the three general topics to be discussed. Such a discussion is warranted, given the current debate (Bärgman, Guo, Jovanis, & Knipling, 2016) in the research community on the topic. Furthermore, Paper V addresses the kinematic similarities between crashes and near-crashes in rear-end scenarios, and, as near-crashes are used in Papers II-IV, it is relevant to frame the use of near-crashes in the context of crashes. The use of near-crashes has its critics (e.g., Carsten et al., 2013; Knipling, 2015).

In order to address the criticisms of the use of near-crashes in NDD research, a recent panel debate at the Transportation Research Board 2016 annual meeting in Washington D.C. (Bärgman et al., 2016) discussed the relevance and validity of SCE from different perspectives. Specifically, based on a previous publication (Knipling, 2015), Ronald Knipling highlighted issues with the use of datasets of mixed SCE—for example, SCE across rear-end, run-off-road, and intersection scenarios—from NDD in any analysis. I agree with his assessment: SCE should be stratified, for example on a per-scenario basis (e.g., only analysing rear-end crash scenario data from an NDD, as in Victor et al., 2015). The take-away message of the panel can be summarized by three points: (1) Do not use mixed SCE unless you can make a very good case for doing so, (2) Take care when using near-crashes as surrogates for crashes in order to avoid bias, and (3) Develop and apply methods that promote generalizability of the SCE analysis.

A common argument against the use of non-crash SCE in NDD analysis is that the link between crashes and non-crash SCE is not sufficiently established. This is partially due to the lack of data (crashes) to establish such relationships. In particular, too few crashes (specifically high-severity crashes) are available in NDD to enable such linkage (Knipling, 2015). With the advent of CNDD, this is likely to change. In a few years there will probably be a vast amount of pre-crash data, with detailed records of the few seconds before the crash. Analyses of these data should contribute to a greater understanding of the relationship between low-criticality (including non-crash) SCE and severe crashes (Paper II). For now the research community should acknowledge near-crashes as different from crashes (e.g., Paper II, contrasting the crashes and near-crashes in crash causation analysis; Jonasson & Rootzén, 2014; Knipling, 2015). Meanwhile, additional research is needed to investigate the link between near-crashes and crashes (e.g., Paper V, showing that the pre-evasive-maneuver kinematics between rear-end crashes and near-crashes are similar; Dozza, 2016; Victor et al., 2015).

Another criticism relates to generalizability. The typical (kinematic-trigger based) operationalization of SCE by definition would exclude some events of particular interest, such as driver drowsiness or falling asleep at the wheel (Knipling, 2015), because the majority of kinematic triggers are based on the driver performing an evasive maneuver. However, it is hard for drivers to perform evasive maneuvers when they are asleep. This valid criticism should be taken into account when analyzing data. In essence, the critique comes from the fact that SCE triggers relying on certain driver actions fail to capture crash-causation mechanisms that preclude those actions in non-crash SCE. As in all research, the researchers need to be aware of the limitations of their data (Carsten et al., 2013). A challenge is to identify such blind spots and biases. However, as mentioned previously, the large number of crashes (potentially) available from CNDD (Carney et al., 2015; Hickman et al., 2010) could shed light on these issues and identify the limitations of using, for example, near-crashes as crash surrogates in traffic safety research. After the relationship between crashes and near-crashes is more firmly established (or debunked), more informed choices can be made about using surrogates in future studies.

Furthermore, different (automatic) kinematic SCE triggers (in the first phase of SCE extraction from NDD) will vary in their relevance to specific crash scenarios. For example, road-departure-based triggers (e.g. triggered by lane-tracker signals; Hallmark et al., 2011) are probably quite relevant for run-off-road crash scenarios. In contrast, hard-braking SCE triggers (Bagdadi & Várhelyi, 2013; Hallmark et al., 2011) are relevant for scenarios in which the driver brakes (and may produce an overemphasis on lead-vehicle conflicts; Knipling, 2015). Depending on the selection strategy for triggers, the proportion of SCE in different crash types (scenarios) will differ widely, without necessarily reflecting the underlying distribution of crash types in actual crashes. Knipling (2015) heavily criticizes the use of mixed-SCE analysis without weighting, stating it is “like cooking without a recipe” (p. 201). The issue of limited representativeness and generalizability of the NDD selected is discussed in Papers II, IV and V. The research community needs improved methods which can produce representative sets of SCE—or use weighting to produce more generalizable results. Current efforts in that direction are ongoing (Imberg, Kovaceva, Bärgman, & Nerman, 2016; Imberg, Liskovskaja, Selpi, & Nerman, 2016).

Evaluating and correcting biases in data has long been a challenge in traditional traffic safety research, which uses national crash statistics and databases of in-depth post-hoc crash analysis (Cryer et al., 2001; Gabler, Hampton, & Roston, 2003; Niehoff & Gabler, 2006; Tivesten, Jonsson, Jakobsson, & Norin, 2012; Yamamoto, Hashiji, & Shankar, 2008). Even so, as pointed out by Knipling (2015), a literature review revealed only a few peripheral studies and discussions on how SCE selection bias affects the generalizability of NDD results (Jonasson & Rootzén, 2014; Wu & Jovanis, 2012). Obviously, further research on selection bias and the generalizability of NDD is needed.

Going forward, traffic safety research benefits in many ways from studies that use naturalistically collected SCE data—which include crashes, near-crashes, and other forms of SCE. Understanding the relationships between different types of SCE will allow researchers to take full advantage of these data in many ways. Analysis of SCE from NDD has the potential to improve ISS, driver training, infrastructure, policy-making, and legislation design. However, as with all data, care has to be taken to establish under what conditions, including boundary conditions, specific data can be used (Papers I-V), and how surrogates for crashes can and should be used (Paper V). While such care has not always been taken, it is clearly possible to do so (e.g. limiting generalization). Actually, it should be possible to develop methods for using NDD (including CNDD) to complement in-depth crash data and crash statistics on a broader scale than that of today. NDD do provide unparalleled insight into driver behavior.

4.2 Inferring causation from NDD

Inferring causation from NDD—the second of three general topics of NDD analysis—is directly relevant to Paper II. The paper’s aim is to develop and demonstrate an expert-assessment method that can support the identification of crash-causation factors, using an observational (NDD) approach. This approach complements stringent epidemiological or experimental approaches that also seek to establish causation. As one of the largest criticisms of the use of NDD is the mantra that “association does not mean causation” (Carsten et al., 2013)—an inherent problem for observational studies—a discussion on this topic is clearly warranted. It should be noted that, in general, causation is a difficult concept (Beauchamp & Rosenberg, 1981; Fair, 1979; Rothman, 2012; Wicksteed & Cornford, 2 vols., 1929).

“What is a ‘cause’?” is how Blower and Campbell (2005) begin their report on the methodology of the Large Truck Crash Causation Study (LTCCS), an observational study of crash causation; inferring causation from observational data is hard, but not impossible (Rothman, 2012). In line with epidemiology (Rothman, 2012), it would seem that the only way to identify real-world crash causation is through observational studies. Now that vehicle video is increasingly available in NDD, crash reconstruction using NDD—and the subsequent analysis into causes (e.g., driver behavior mechanisms contributing to the occurrence of crashes)—can in many respects be more refined than crash reconstruction in traditional crash-causation studies, which is based only on reconstruction and interviews. Refinement is possible through the availability of objective time-series observation of the pre-crash phase (Paper II). However, NDD have drawbacks unknown to traditional crash-causation studies. These include: (a) current datasets allow low generalizability (risk of overgeneralization, currently likely to be the main disadvantage of NDD); (b) obtaining information about driver-state factors (e.g., drowsiness) is hard without post-crash interviews (or similar); (c) post-crash injury and long-term disability follow-ups are typically not done for NDD crashes; and (d) the rarity of crashes means that crash surrogates are used (which should be done with care; see 4.1). Thus, NDD and traditional crash-causation studies will need to be conducted in parallel for the foreseeable future—as complementary tools for understanding crashes and crash causation. Using NDD to analyze crash causation should be considered at least as valid as more traditional analyses using in-depth crash data or data from other crash databases, with the caveat of avoiding overgeneralization (with currently available datasets).

When NDD analysis of causation is being discussed, the “correlation does not imply causation” mantra is often invoked (Boyle & Lee, 2010; Carsten et al., 2013). The difficulty of inferring causation with observational data is actually present in most traditional (e.g., in-depth analysis, or crash-database-based) crash-causation studies. There is no reason why a well-constructed NDD case-crossover study with carefully chosen controls (baselines)—for example, using only crashes—should be less valid in terms of inferring causation (or not) than a study using induced exposure applied to in-depth crash data (Lie, Tingvall, Krafft, & Kullgren, 2006; Rizzi, Strandroth, & Tingvall, 2009). Of course, each method has its benefits and drawbacks.

Similarly, in traditional expert-assessment studies, causation is often inferred without employing a strictly controlled experimental approach. Actually, in expert-assessment studies, controls are typically unavailable. For example, no controls were used in Paper II, or one particularly important study in the traditional crash-causation literature on the influence of drivers’ (actually drivers’ and bikers’) behaviors and their expectations in crash causation (Räsänen & Summala, 1998). Their study was based on data from four different crash investigation teams (each including a police officer, a vehicle engineer, a traffic engineer, and a physician) which conducted in-depth bike crash studies in four cities in Finland during the 1990s. Based on the experts' reconstructions of the crashes, erroneous expectations were found to be a main contributing factor, particularly when the bike came from a bike lane on the right when the driver was turning right. Note that the study partially relied on drivers' and riders' memories (i.e., post-hoc reconstructions) of the events. While the authors used no controls, causal (contributing) factors were presented and strongly argued for. A review of the literature has found no criticism of their paper from a "claiming causation" perspective. Indeed, the statement, "The level of reconstruction in these data provides a fairly reliable description of what actually happened" (Räsänen & Summala, 1998, p. 659) is probably accurate. Based on theory and scientific reasoning, the authors identified causal mechanisms without a stringent, random, controlled-experiment approach. This basis is similar to the expert-assessment method presented in Paper II (as well as similar methods; Dunn et al., 2014; Habibovic et al., 2013). Different methods have different benefits and drawbacks: to date NDD analyses often have an issue with generalizability and some fundamental biases, while traditional crash-causation analysis may have issues with data quality (e.g., derived from e.g. interviews or from lower-quality police reporting; Cryer et al., 2001) and other types of biases that studies using NDD don't have. However, from a causation-claiming perspective, both (arguably) have more or less the same issues.

One of the most well-known studies of crash risk is the Indiana Tri-Level study of crash causation (Treat et al., 1977), which defined causation as: "a factor necessary or sufficient for the occurrence of the crash; had the factor not been present in the crash sequence, the crash would not have occurred" (p. 16). This definition is highly consistent with epidemiological views of causation (Rothman, 2012). In the expert-assessment method presented in Paper II, a similar definition of causation was used as the basis for the identification of factors contributing to crashes: analysts applying the method are to consider all the available information on a crash (or near-crash), and for each factor in a structured coding schema ask the question, "If this factor had not been present, would there have been a crash?" This will, of course, introduce potentially biased subjectivity into the procedure. With stringent operationalization of such procedures, and with the method based on a model of how crashes occur, the subjectivity is minimized. However, in the further development of methods based on expert-assessment, validation—including studies of inter-rater reliability—should be a cornerstone (Warner & Sandin, 2010).

After observational studies (e.g., NDD) have provided insights into likely crash causation mechanisms, controlled experiments can (and should) be designed to dissect those findings to detail how driver behavior contributes to crash causation. However, NDD and driving simulators have so far mainly been utilized in isolation from each other, rather than benefiting from each other's strengths (Boyle & Lee, 2010; Tarko, Boyle, & Montella, 2013). It is worth emphasizing that controlled experiments and NDD should not seen as competing; they are complementary.

Rather, when results from NDD and controlled experiments are in disagreement (e.g, the effect of the role of taking on the phone providing different results and interpretations in the simulator and NDD studies; Strayer & Cooper, 2015), the details of both methods should be scrutinized to identify the reason for the discrepancy. Together, the two methods can help researchers better understand crash causation and ultimately help save lives.

Finally, “inaction is also an action” (C. Flannagan, personal communication, June 22, 2016). When findings from observational driving studies (e.g., expert-assessment methods; see Paper II), or epidemiological studies (Victor et al., 2015) are dismissed, or their results considered less useful than controlled experiments (Carsten et al., 2013; Kircher, Patten, & Ahlström, 2011; Strayer & Cooper, 2015), there is a risk that the research focus is directed toward aspects of traffic safety that are less than optimal for reducing injuries and saving lives. The concept of “inaction is also an action” is comparable to the debate in the 1950-60s on the relationship between smoking and cancer. It took a long time for researchers to provide statistically irrefutable evidence of smoking causing cancer – although there was overwhelming evidence that showed a clear association between the two. The respected statistician Ronald Fisher wrote articles arguing that association is not enough to establish causation (see “The alleged relationship between smoking and cancer” (Fisher, 1957)). If researchers repeatedly find evidence in observational studies that point in one direction, not acting on such results is an action (inaction) that can have consequences for safety and public health. These consequences apply to traffic safety research with respect to the use of observational data (e.g., NDD) in carefully conducted studies. As previously stated, if experimental studies and observational studies do not corroborate each other’s results, the methods used in experimental studies should be scrutinized, as should the NDD methods. In general, researchers must know their data, their methods, and the limitations thereof (Carsten et al., 2013; Papers I-II).

4.3 Commercially collected NDD as a traffic-safety research tool

The potential of commercially collected NDD (CNDD) as a traffic-safety research tool—the third general topic on NDD analysis—is directly relevant for Papers II and III, as both use CNDD. It should be noted that the methods of Papers IV and V can also be advantageously applied to CNDD, and I argue that a shift towards doing so in the future is likely.

This discussion topic is rooted in the distinction between NDD collected with the intention of performing research on driver behavior (i.e., conducting an NDS) and CNDD. In the latter, there are two parallel aims: the main commercial aim is to generate a profit stream, while promoting the common good (saving lives) through improved traffic safety is the main humanitarian aim. In contrast, most NDS focus on issues of traffic safety by answering a set of empirically falsifiable questions and interpreting and applying the results—by, e.g., developing an ISS, designing infrastructure, or making policy. The results from analysis of NDS can also, however, be used for commercial purposes. For CNDD to be collected, incentives are needed for both a product/service provider and a buyer (for example, a trucking firm). The buyer's incentives are typically to (a) have an objective means of proving innocence and reducing legal costs in the event of a crash, (b) reduce the fleet's stand-still costs by reducing the number of crashes, (c) reduce the repair costs for crashes, and (d) reduce vehicle maintenance costs by diminishing wear-and-tear (for example, on brakes and tires) (Lytx, 2016; SmartDrive, 2016; Victor et al., 2010). The providers' incentive is naturally to make money, but they also want to make roads safer (Lytx, 2016).

Most of these motives differ from those that traditionally drive NDD research—with the exception that both include the aim to make roads safer (and save lives). To date, bodies that fund NDS have overwhelmingly been government agencies, at the federal (e.g., European Union or US. Department of Transportation) or regional (national funding in Europe or state funding in the US) levels. The financing is often motivated by the opportunity to improve traffic safety by (a) identifying traffic safety concerns (Boyle et al., 2009; Sayer et al., 2007; Uchida et al., 2010; Utesch et al., 2014); (b) quantifying relationships between safety and a specific factor, such as a driver, vehicle or environmental condition (Boyle et al., 2009; Papers I and V; Victor et al., 2015); or (c) addressing known traffic safety concerns, by means of ISS product development or evaluation (Benmimoun et al., 2011; Bezzina & Sayer, 2015; Papers IV and V; LeBlanc et al., 2006; Sayer et al., 2011).

CNDD has received little attention from the research community to date; only a few studies have used them as a basis for traffic safety research (Carney et al., 2015; Eiríksdóttir, 2016; Hickman & Hanowski, 2010; Hickman et al., 2007; Lich & Georgi, 2011; Lich et al., 2012; McGehee et al., 2007b; Olson, Hanowski, Hickman, & J., 2009; Rose, Carter, Pentecost, & Voitel, 2013). A literature review revealed that, of the studies that do use CNDD, none applied structured expert-assessment-based crash-causation methods to the data. This is a research gap that Paper II addresses. The closest methodological match is a prevalence assessment of teen drivers (Carney et al., 2015); it, too, examines CNDD video, extracting a set of behaviors and factors. The difference between the two papers is that Paper II performs an expert-assessment-based crash-causation analysis, while Carney et al. perform only the prevalence assessment. Further, no studies were found that document methods for extracting kinematics and optical parameters from CNDD video; Paper III addresses this second research gap.

To clarify their benefits and drawbacks, NDS and CNDD are discussed in terms of their utility as valuable sources of data for future traffic safety research. The benefits of data from research-focused NDS are: (a) sensing suites, connectivity, data acquisition, and other vehicle instrumentation are usually extensive (e.g., CAN bus and radar; Dingus et al., 2006; Selpi et al., 2012; TRB, 2014); (b) sample frequency and sensor fidelity are typically high; (c) both everyday driving (baseline) and SCE are usually captured (e.g., Benmimoun et al., 2011; LeBlanc et al., 2006; Sayer et al., 2010; Victor et al., 2010); and (d) researchers can define vehicle and driver selection and criteria for SCE extraction (e.g., Barnard et al., 2015; Klauer et al., 2011; Sayer et al., 2010). On the other hand, the main drawback with research (non-commercial) NDS data collection is that it is expensive. As a result, datasets from research NDS typically contain only a small number of crashes (and other SCE), and data analysis tends to receive a relatively small share of the budget. Although the expert-assessment method presented in Paper II can yield useful results even when applied to datasets with relatively few crashes, access to a larger set of crashes across all severities, as (potentially) facilitated by CNDD, is highly desirable. With larger datasets, stratification into smaller and more homogenous crash-scenario typologies will provide a foundation for understanding details and nuances in drivers' behaviors in the pre-crash phase.

There are several major benefits to CNDD. First, because data collection does not need to be government-sponsored (McGehee et al., 2007b; McGehee et al., 2007c; Papers II-III), there is a potentially large cost savings for researchers and society. Commercial data collection is driven by a positive relationship between the number of vehicles equipped with the data acquisition system and the company's revenue, rather than each additional unit on the road increasing cost. Second, because each vehicle that has the data collection system installed generates income for the providers, any company collecting CNDD strives for a large number of vehicle installations. As a result, a large number of crashes (and other types of SCE) are likely to be observed and recorded. Finally, in CNDD data collection is a continuous process—in contrast to NDS, which has a finite data collection period defined by the specific research (or data collection) project (Blatt et al., 2015; LeBlanc et al., 2006; Neale et al., 2005). Crashes from these large commercial datasets can be sub-categorized (stratified) into more detailed crash scenarios for analysis (Paper II). This sub-categorization can provide opportunities to address concerns with NDD analysis, for example: (a) investigate differences in causal mechanisms between crashes of different severities within the same crash scenario, and (b) understand the relationship(s) between lower severity-level SCE (e.g. near-crashes) and more severe (e.g., fatal) crashes.

Although it has many advantages, CNDD has its drawbacks. A literature review revealed five main disadvantages to CNDD compared to traditional NDD. First, vehicle selection cannot be controlled by the researchers. Second, the initial event selection procedures are not set by the researchers (Hickman et al., 2010; Paper II). Third, the fidelity of the data is, at least to date, usually lower than researchers would prefer, with limited sensing equipment and low sample rates (Paper III). Fourth, the drivers studied are likely to be taking part in a behavior-based safety program and thus their behavior may not be representative of the general population (Paper III). Finally, since a traditional control group is not likely to be part of the providers' business model, baseline data from normal driving are usually not collected. Thus, in order to perform risk calculations, researchers must use non-traditional controls if epidemiological methods for analysis are to be applied. For example, to calculate risk, Hickman, Soccolich, Fitch, and Hanowski (2015) used events kinematically triggered by the in-vehicle system, but not classified as safety-related (e.g., driving over train tracks or potholes), as controls. This is an approach that merits further examination, but additional approaches should be developed as well, to establish effective baseline-selection methods and validate their use in risk calculations when CNDD are used.

In spite of the drawbacks, the uniqueness of the data means CNDD must be considered—as long as the research questions do not extend beyond what the data can support (e.g., as long as the results of CNDD analysis are generalized with care). It is important for researchers and governments to show the companies collecting CNDD that collaborating adds value to their business model. Without some incentive, it is unlikely that the data will be made available to external researchers, whose work could help traffic safety research, reducing injuries and fatalities in traffic more extensively than the CNDD company's safety services alone. Paper III demonstrates the relevance of CNDD for developing behavior-based safety services, and may support the development of other safety solutions as well—supporting safer driving on our roads. The analysis in Paper II takes us one step closer to convincing CNDD providers of the benefit of releasing data for traffic safety research. Furthermore, methods such as that presented in Paper III (as well as Meng and Wang (2016), a follow-up of the Paper III method) are needed to facilitate the use of CNDD for research. As researchers continue to explore the great potential of CNDD, it is important to thoroughly understand the data's advantages and disadvantages.

4.4 Improving everyday-driving analysis of NDD by chunking

Paper I presents a procedure for reducing biases and errors in the analysis of normal everyday driving using vehicle-based NDD. Scientific publications and technical reports seldom provide detailed descriptions of their calculations in the analysis of NDD or other data sources, and may even omit the descriptions altogether. For example, Green et al. (2004, p. 20) state: "Unfortunately, there is no official, standard, or even well accepted definition of the standard deviation of lane position [SDLP], and, in fact, it is extremely rare for research reporting results to define it." Since it is rare for papers and technical reports to define SDLP and other driver performance metrics, it is difficult to estimate to what extent previous studies are affected by bias—such as size bias error (introduced when a function is applied to segments of different sizes/durations; Paper I addresses this problem). In some studies it is relatively clear how the metrics were calculated, so it is possible to infer whether size bias played a role (and if so, whether it was compensated for). For example, in LeBlanc et al. (2006) it can be deduced (although other interpretations are possible) that the SDLP was calculated on the entire pool of data (samples) of segments of lane position data for a specific stratification (e.g., all data across all trips with/without ISS or between different road types for each of the four weeks; LeBlanc et al., 2006). For this and similar aggregate analyses that do not apply a statistic (e.g., mean or standard deviation) to individual segments of data, the chunking method of Paper I does not provide much benefit (as there is nothing to chunk).

In contrast, some studies clearly addressed the problem of size bias appropriately, even when dealing with shorter segments. For example, Peng et al. (2013) calculate SDLP on one 3s segment (chunk) per secondary task duration, and Sayer et al. (2007) apply SDLP to 5s segments (see Sayer et al., 2005a for more details). However, Sayer et al. (2005a) and Sayer et al. (2007) also used autoregressive integrated moving average (ARIMA; McDowall, McCleary, Meidinger, & Hay, 1980) methods to avoid sample dependencies and the size bias issue. There is yet another potential issue with data segmentation: short segment choices. The analysis in Paper I shows that for chunks shorter than approximately 60s, the SDLP is highly sensitive to size changes (see Figure 7, Paper I), and the variance in such an unstable condition may be large. Unfortunately, in studies such as Sayer et al. (2005a), Sayer et al. (2007), and Peng et al. (2013), the researchers could not choose chunks of longer duration, due to the short duration of the tasks they were studying.

Paper I addresses the assumption of independent observations in many statistical methods, and stresses the importance of identifying autocorrelation in data (e.g., through autocorrelation analysis) and addressing it (e.g., carefully choosing an appropriate analysis method and deciding if chunking should be used). In the Peng et al. (2013) study, observation independence was achieved by selecting only one 3 s chunk per task, even if the segment available was much longer. This selection method may be suitable if high autocorrelation is found, but statistical power is sacrificed when chunking is not (or cannot) be applied (e.g., when high autocorrelation does not allow for it). Thus, if autocorrelation can be shown to be low, the use of chunking is beneficial. Note that the problem with autocorrelation also exists in many other NDD analysis methods (as well as in the analysis of controlled experiments): for example, when pooling all data points for all segments (e.g., trips in NDD data) and then performing a mathematical operation (LeBlanc et al., 2006). Although metrics can be calculated on these aggregate distributions of data (e.g., the individual data points within each individual segment pooled, on which a statistic such as standard deviation is applied), if the data are highly autocorrelated then care must be taken not to violate conditions under which specific statistical techniques can be used. Paper I highlights the issue with autocorrelation; future work should continue to explore innovative methods to address it, such as Sayer et al. (2005a) use of the residuals from ARIMA-based methods.

Finally, chunking can also, in some cases, reduce bias when applied to data from experiments, such as the SDLP calculations in the Driver Workload Metrics Project test-track study (Angell et al., 2006) and the Maciej and Vollrath (2009) simulator experiment. (As the studies are similar, the study design is described only for the former.) The drivers were asked to perform 22 tasks, which were later evaluated using different performance metrics, including SDLP. The tasks lasted approximately 10 s to 120 s, and SDLP was calculated over the task duration. According to Paper I, the SDLP results from both studies would be affected by size bias due to the different, and short, task durations. Similar biases may be present in several other studies as well, but without detailed descriptions of the calculations it is hard to know. In one such study Östlund et al. (2004), the size bias issue was at least marginally acknowledged.

4.5 Expert-assessment based crash causation analysis using NDD

The expert-assessment-based methods presented in Paper II address the identification of factors and mechanisms that may be associated with crash causation. Paper III facilitates the analysis of Paper II by providing kinematic and optical parameters from crashes and near-crashes. The term ‘expert assessment’ should be understood as a qualitative assessment by human experts who utilize all available information for a specific event (e.g., crash or near-crash) in order to identify crash-causation factors and mechanisms. This identification process is a foundation of Paper II. The information considered may include quantitative data (e.g., acceleration and speed), texts (e.g., written narratives), and video (e.g., forward and driver views). The factors and mechanisms identified include driver behaviors and aspects of the traffic environment that contribute to the occurrence of crashes and other SCE.

Traditional expert-assessment-based crash-causation analysis uses interviews and road-users’ memories to understand even very time-critical causation factors (e.g., visual occlusion and glance behaviors; Sandin, 2009a; Seeck et al., 2009). However, complementing these analyses with NDD means that time-series kinematics and video records can provide invaluable data for the last few seconds (or more) before the critical moment, in both crashes and near-crashes. For example, determining the role of visual occlusion in intersection crashes previously meant relying on drivers’ memories (Sandin, 2009a), while the method in Paper II can be applied to address the role of visual occlusion in relation to SCE occurrence and the presence of other road users. Objective data about occlusion is thus available, taken from a perspective similar to that of the driver of the instrumented vehicle. This approach facilitates a more detailed study of how an SCE unfolds over time.

In the near future NDD-based crash-causation analysis (Bärgman & Piccinini, 2016) may challenge the infamous “looked-but-failed-to-see” category of intersection-crash causes (Herslund & Jørgensen, 2003; Koustanaï, Boloix, Van Elslande, & Bastien, 2008; White & Caird, 2010). Research to date has few insights into what this category actually comprises. This category may have been invented (or at least much overused) due to a lack of information about what actually happened in the few seconds before the crash—an information gap that NDD and expert-assessment-based crash-causation analysis will, at least partially, eliminate (e.g., using the method in presented in Paper III). Such insights into crash causation were previously unthinkable. Currently, there are still only a few methods using qualitative analysis of time-series data and video in NDD to identify the factors and mechanisms that contribute to crashes and other SCE (notable exceptions inlcude Dunn et al. (2014); Habibovic et al. (2013)), and none (with the exception of the method in Paper II) have yet been applied to CNDD. Thus, Paper II makes a scientific contribution to traffic safety research by providing a starting point (together with the work by Habibovic et al., 2013) for structured expert-assessment approaches identifying factors and mechanisms that contribute to the occurrence of SCE using detailed NDD. Traditional methods alone are less suited to addressing these aspects of crash causation because there is some critical information they simply cannot provide. However, traditional approaches in turn can contribute information that typical NDD do not have (e.g., information on drivers' states or traits from post-crash interviews)—the two approaches are complementary.

Understanding the influence of other road users on the occurrence of SCE is another example of the usefulness of crash-causation analysis methods that use NDD with video. Traditional crash databases and in-depth crash studies cannot identify when or how other road users' presence may be contributing to crashes (Sandin, 2009b; Van Elslande & Fouques, 2007); the data are simply not recorded, or may be unavailable to researchers in traditional studies. In contrast, vehicle-based NDD with video provide detailed (frame-by-frame) visual information about 1) when and for how long the driver looks towards other road users, and 2) how and when other road users may occlude the principal other vehicle (Paper II). Because occlusion and drivers' gaze behaviors have been identified as important factors for crash causation (Dozza & Werneke, 2014; Klauer et al., 2014; Räsänen & Summala, 1998; Sandin, 2009a; Victor et al., 2015), NDD-based insights are important, and will be for a long time to come.

As previously described, NDD provide a wealth of information for the study of both everyday driving and SCE. However, even with the data from the large SHRP 2 (TRB, 2014) study, in which over 1000 crashes were collected, after the crashes are stratified into specific crash scenarios and low-severity crashes are excluded, the number of crashes in each group is still relatively small. To address this deficit of data, data from nontraditional sources (e.g., CNDD, used in Papers I and II), can be used as a complement. Although there are some limitations, CNDD lends itself readily to the quantitative analysis of crash causation. For example, a recent study of

Chinese (Shanghai) CNDD SCE (Piccinini, Engström, Bärgman, & Wang, 2016) concluded that the main reason for the occurrence of SCE in China is quite different from the main reason in the US (as reported in Paper I), at least for professional Chinese drivers. It seems that distraction, the main reason in the U.S., is much less of an issue in China. Basically, drivers are less distracted in China. If you are distracted in China, you have a very high probability of ending up in a crash, contrary to the US, where there is 'just' an increased risk associated with distracting activities— not a close-to-certainty of ending up in a crash. Instead, Piccinini et al. (2016) study indicates that drivers' choice of (short) headways is the main crash causation factor in China.

Due to the large amounts of data in NDD analysis, the difficult tasks of quality assurance and data validation (should) consume relatively large portions of the available funding. In particular, range and range-rate data from radar deserve special attention. Radar has been available in several NDS studies; however, the quality and availability of range and range-rate data have been quietly debated in the research community. In the study by Victor et al. (2015) of SHRP2 data, the radar data was considered so poor (in terms of availability and quality) that the method presented in Paper III was used to extract range and range-rate data from video instead—to avoid excluding a large portion of the available crashes from the analysis. At the time that project was being run, the radar data had not been processed enough to provide range and range-rate data of sufficient quantity (missing radar data for entire events) and quality (e.g., gaps and errors in the data, and inconsistencies in tracking) for the analysis at hand. (Note that at least one project has been conducted to improve SHRP2 data (Gorman, Loren, & Hankey, 2015)). In addition, information about range and range-rate is not always part of NDD datasets; for example, current CNDD typically does not include range or range-rate (Carney et al., 2015; Hickman et al., 2010; Papers II and III). As a result, pragmatic methods for extracting this information, vital to the analysis of crashes, are needed. Paper III fills this need, enabling the extraction of these variables from video when datasets do not contain range and range-rate data.

4.6 ISS safety-benefit evaluation using counterfactual simulation

The terms actual severity and potential severity are important for framing the application of counterfactual simulations to safety-critical events in traffic. In this thesis **actual severity** is defined as the actual outcome of an event: for crashes the associated metric is typically delta-V (with corresponding injury severity), and for near-crashes it could be time-to-collision (TTC; SAE, 2015), depending on what is relevant in the particular crash scenario. Actual severity describes what actually happened at a particular time. Conversely, **potential severity** is what could have happened had something been different in an event (Figure 7). The latter is thus the counterfactual ('what-if') outcome if, for example, an ISS had been available in an event (Paper IV), or if a driver had exhibited a different glance behavior (Paper V).

Studies analyzing only the actual outcome of crashes, for example, delta-V (Gabauer & Gabler, 2006; Toth et al., 2003; Viano & Ridella, 1996) fall into the actual-severity domain.

When there is a crash, it could be that a very small change in any of a set of variables would have produced a near-crash instead of a crash (Davis et al., 2011; Papers IV and V). It could also be that the particular event was very rare—some combination of factors that do not commonly occur made the event into a crash. Further, when a crash is compared with a near-crash it is possible that the crash was a very low-speed event, with a very low probability of occurring (see Figure 8), while the near-crash was a high-speed event, with a high probability of becoming a (high severity/injury risk) crash. Using the actual severity concept, even a 1 km/h bumper touch would be considered more severe than a 90km/h straight crossing path near-crash (Tijerina, Chovan, Pierowicz, & Hendricks, 1994) with a margin between the vehicles of 100 milliseconds—because the actual outcome was more severe. In contrast, in terms of potential severity, the near-crash would likely be considered more severe—because the potential for damage and injury was much greater.

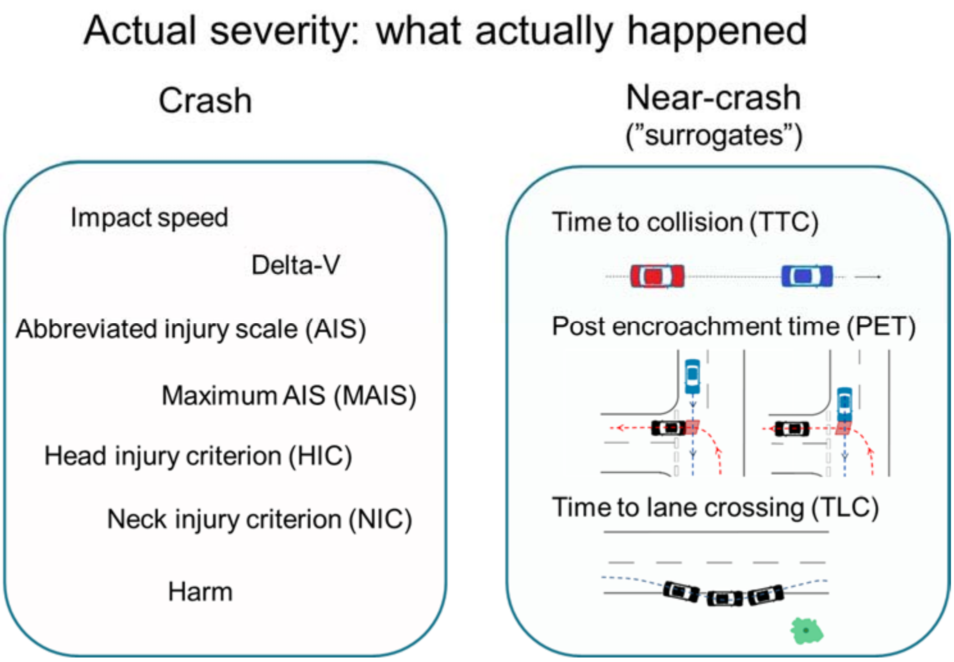

Figure 7: Actual severity metrics for crashes and near-crashes. For crashes the metrics typically start with impact speed, and may end with some injury criteria (NIC; Boström et al., 2000; AIS, MAIS;Gennarelli & Wodzin, 2006; HIC; Margulies & Thibault, 1992), or cost (Harm; Malliaris, Hitchcock, & Hedlund, 1982) metrics. For near-crashes the metrics vary, depending on the specific crash scenario.

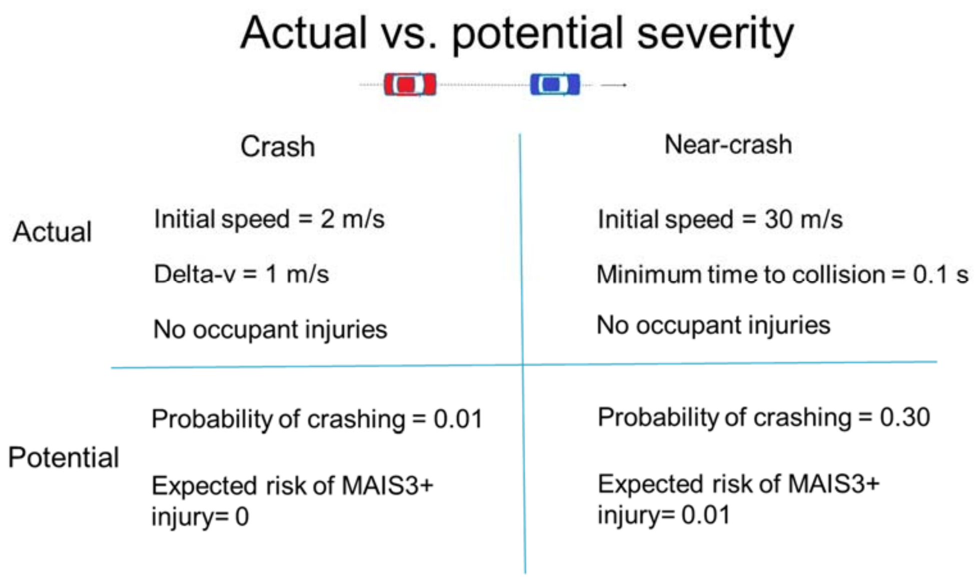

Figure 8: A 2x2 matrix compares actual and potential severities for a crash and a near-crash. A crash with very low initial and impact speeds is contrasted to a near-crash with high initial speed, but very low time-to-collision. The actual severity reflects exactly what happened. The potential severity indicates the probability of crashing and the risk of injury in each particular scenario: with a specific driver behavior (e.g., glance off-road), vehicle (e.g., braking system performance), and environment (e.g., friction and distance to the lead vehicle) configuration. In this example, there is a much higher risk of an occupant crashing and suffering a severe injury (e.g., MAIS3+; Gennarelli & Wodzin, 2006) injury in the near-crash than in the crash.

Perhaps surprisingly, Paper V shows that, on average, near-crashes and crashes, at least for rear-end scenarios, have similar kinematics before the driver of the following vehicle starts performing an evasive maneuver. The timing and execution of the maneuver (e.g. jerk and brake level; Paper IV; Davis et al., 2011) determine the outcome. If ISS (including autonomous vehicles) can do a better job than a human at getting that timing and execution right, they would provide higher overall safety and should be promoted, for example through car safety assessment programs (Euro_NCAP, 2015).

Consequently, each crash should be seen as an individual instance along a continuum of possible outcomes (as shown in Paper V and as discussed by Davis et al., 2011). Other instantiations produce more (or less) severe crashes or even no crash (i.e., a near-crash). From a traffic-safety perspective, claiming that crashes are always to be considered more severe than near-crashes is problematic—except when talking specifically about actual severity.

The operationalization of potential severity using counterfactual simulations would be a great contribution to traffic safety research (Papers IV and V), as well as industry and government. The successful mathematical modeling of crash-causation mechanisms and driver behavior is, however, a prerequisite to improving the prospective analysis of risks and the understanding of both ISS and crash-causation mechanisms through counterfactual simulations. This is grounded in the theoretical basis for counterfactual simulations presented by Davis et al. (2011), and two counterfactual simulation demonstrations are presented in Papers IV and V. As has been noted, published research on counterfactual simulations is rare to date, even though they are of great interest to the automotive industry (Page et al., 2015). It is hoped that the near future will bring improvements to the mathematical models and more research confirming the potential of counterfactual simulations, with appropriate attention given to models of driver behavior.

The use of counterfactual evaluation methods to estimate the safety benefits of ISS is not new. In fact, a range of methods is subsumed under the term counterfactual simulations: from purely expert-assessment based counterfactual evaluation (experts do best-guess estimates of the benefit, based on limited information), to highly mathematical counterfactual simulations with advanced models of crash causation and driver behavior. Paper IV and V are far to the latter side of this range. One example of a method that is somewhere in the middle is presented in a study by Strandroth (2015a); the researcher examined in-depth data from individual crashes and applied what could be called an expert-assessment-based counterfactual simulation, supported by basic models of crash kinematics, but without explicit consideration of detailed driver behavior aspects of crash causation. This approach was further validated in a subsequent paper (Strandroth, 2015b). Although the Strandroth (2015b) method is shown to be good at estimating a rough, ISS-generic benefit, it is probably not the best choice for addressing detailed aspects of ISS designs, specifically with respect to the impact of driver behavior, since the method does not include detailed models of crash causation mechanisms, or time-series of the pre-crash kinematics and driver behavior. On the other hand, counterfactual simulations of ISS for safety benefit analysis which use mathematical simulations and models of driver behavior as demonstrated in Paper IV, can fill an important gap; they facilitate detailed, rapid development and optimization of ISS at relatively low cost. As Paper IV shows, appropriate, validated models of driver behavior and crash causation mechanisms become very important.

Researchers considering ISS may be asking the following: Why not run NFOT studies as part of the ISS development, optimization, and early evaluation process, instead of using counterfactual simulations? There are, after all, several advantages to conducting an NFOT (Benmimoun et al., 2011; Bezzina & Sayer, 2015; LeBlanc et al., 2006; Sayer et al., 2010): (a) performance indicators of everyday driving can be studied, (b) the ecological validity is considerably higher than in counterfactual simulations, and (c) drivers' subjective experience of ISS interfaces and interactions (e.g. nuisance warnings; Smith & Källhammer, 2010) can be studied directly (counterfactual analysis cannot evaluate drivers' subjective assessment of an ISS). The simple answer is that NFOT studies are costly in terms of money and time.

Consider the development of in-crash passive safety systems: it is expensive to crash cars to develop a new airbag, just as it is expensive to conduct large studies of safety benefit evaluations for ISS. The creation of counterfactual simulations with models of crash-causation mechanisms and driver behavior (as well as other included models) which are all well validated reduces both time-to-market and development costs significantly (as well as costs associated with estimating and evaluating benefits). Counterfactual simulations with large parameter spaces can be created before running a smaller set of NFOT to validate the simulations and evaluate drivers' acceptance (which cannot be done through simulations). Furthermore, Papers IV and V both demonstrate the feasibility of multi-dimensional sensitivity analyses, which can be used by the automotive industry in the design and evaluation stages.

The future use of counterfactual simulations to evaluate the benefit of ISS lies in the hands of the developers (and validators) of mathematical models of driver behaviors, as well as in the hands of the researchers providing a detailed understanding of crash-causation mechanisms. As the development and deployment of intelligent (i.e. autonomous and connected) vehicles increase, methods for evaluating their impact on safety will be crucial. Counterfactual simulations can be one piece of the puzzle.

4.7 Using counterfactual simulations to understand the effects of driver behavior

A literature review revealed no other documented methods that use mathematical simulations combining pre-crash kinematics and models of driver behavior to evaluate the effect of driver behavior on safety. Paper V is pioneering the research in this area, focusing particularly on assessing risks related to driver behaviors involving new driver-vehicle interfaces, and other in-vehicle driver behaviors (e.g., using smart phones in different ways while driving).

The method presented in Paper V can be used by vehicle designers to evaluate specific interface designs with respect to safety. Similarly, the method could be used in design guidelines; in the last few years the rapid increase of new in-vehicle infotainment features (e.g., navigation systems and menu-based music selection) and interfaces between the driver and the vehicle (e.g., touch screens) has prompted governments to provide additional visual-manual task guidelines (see Regan, Lee, & Young, 2008, for a review of such guidelines). However, it is difficult to develop these guidelines. The current US visual-manual guidelines (NHTSA, 2013) have been criticized for having methodological problems of repeatability (Rydström et al., 2015) and being overly strict (Heinrich, 2015; Ljung Aust et al., 2013); they may also be using thresholds and requirements that are not entirely based on state-of-the-art research (Victor et al., 2015). The problem with designing such guidelines is probably rooted in the difficulty of detailing a quantitative relationship between driver (glance) behavior, crashes, and crash outcomes (Ljung Aust et al., 2013; Victor et al., 2015). It is particularly difficult to evaluate the net safety effect of a vehicle design that includes infotainment systems as well as ISS, since the latter improve safety and may radically mitigate, for example, glances off the roadway related to infotainment systems (Victor et al., 2015). The need for methods that can evaluate DVIs and new ISS in combination is clearly stated in the following excerpt from a technical report on the analysis of SHRP2 NDD, which focused on quantifying the relationship between glances off-road and crash risk:

> “Regarding human-machine interaction design, distraction guidelines, and other regulatory agency countermeasures, the results emphasize the need to tackle the distraction problem as a joint probability problem. Risk can most effectively be reduced by removing the timing mismatch of eyes off road and lead-vehicle closure rates (inverse TTC change rate). A reduction of both sides of the equation—reducing eyes-off-road occurrence and reducing closure rates—is recommended.” (Victor et al., 2015, p. 106)

A counterfactual simulation approach that combines the evaluation of ISS (Paper IV) and the evaluation of DVI designs (Paper V) is one way to achieve the combined evaluation suggested by Victor et al. (2015).

Typically, design guidelines, such as the US Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (NHTSA, 2013) use thresholds on metrics from normal driving, such as time to complete task (i.e., total task time), total eyes-off-road time, and longest single glance duration. The metrics are usually obtained from normal (baseline) driving in a driving simulator, rather than from critical events. A possible reason for this is that exposing study participants repeatedly to a (the same simulated) critical event (i.e., unexpected lead-vehicle braking) is (at best) questionable with respect to adaptation and ecological validity (Engström et al., 2010).

In contrast, counterfactual simulations (Paper V) allow crash and injury risk to be evaluated as counterfactual outcomes (e.g., impact speed and injury risk) on a continuous scale, rather than by means of thresholds on everyday driving metrics. The advantages of counterfactual simulations are: (a) they can capture the continuous nature of risk increase (rather than the somewhat binary use of thresholds on driver performance metrics in everyday driving data from simulators), and (b) they address crash (and possibly injury) risk directly (rather than via metrics only indirectly affecting safety—since critical kinematics and glances off-road often need to coincide for there to be a rear-end crash (Victor et al., 2015)).

Paper V demonstrates that counterfactual-based approaches could replace (or, at least, complement) threshold values for everyday driving metrics in the guidelines. The replacement consists of pre-crash kinematics and a mathematical driver reaction model, onto which a glance behavior parameterization (including total task time, percent eyes-off-road, and an eyes-off-road distribution) is applied (Paper V). This type of approach makes it possible to evaluate glance behavior in a variety of scenarios with different ISS (Victor et al., 2015). The ability to apply ISS is essential because the automotive industry introduces new ISS in parallel with new in-vehicle systems (e.g., infotainment systems), and what is actually important is the net effect on safety of these combined introductions. Of course, only risks that the implemented models take into account will be part of the results—particular care has to be taken, as the occurrence of counterfactual crashes is based on the crash-causation models (e.g., glances off-road at an inopportune time, or having too short headway). What is not modeled will not produce crashes, and thus not be part of the benefit evaluation.

For further development of the counterfactual method for evaluating DVIs, validated models of driver behavior (e.g. reaction models) are critical. Development of (real-world) validated mathematical models of the effect of cognitive load on safety is of particular importance, as is extending the application to other crash scenarios beyond rear-end crashes (e.g. intersections and run-off-road). There is also a clear need to further generalize counterfactual simulations, while using appropriate driver models. Crash database reconstructions (Erbsmehl, 2009; Georgi et al., 2009; Lindman & Tivesten, 2006) and EDR (Scanlon et al., 2016) are typically much more generalizable than the typical NDD—the former being developed over many years to facilitate weighting to population level generalization for a region or nation. However, crash database reconstructions and EDR do not include detailed (time-series) information about driver behaviors (e.g., glance behavior). By merging kinematics from more generalizable data (i.e., crash database reconstructions or EDR) with time-series driver behavior from NDD (e.g., see Figure 4), it may be possible to create reasonably generalizable counterfactual simulations that include driver behavior. There is, however, a need for more research into understanding the generalizability of driver models and driver behavior in crash causation.

4.8 This thesis and the future of automated driving

Highly automated vehicles (HAV) will provide new, large naturalistic datasets, increasing the opportunities and demands for counterfactual analyses. The introduction of HAV also increases the need for an understanding of their new crash-causation mechanisms. The new NHTSA guidelines (NHTSA, 2016) require HAV to record enough driving data (from the HAV system and the human driver, if in control) to reconstruct any event of interest. An **event of interest** includes crashes, incidents, and "**positive outcome events**", i.e. events in which the HAV system correctly detects a safety-relevant situation, and successfully avoids an incident. Further, the NHTSA guidelines propose that these data be shared. In the likely scenario in which naturalistic data become increasingly available, the methods proposed in this thesis can have a large impact.

NHTSA guidelines also set a high priority on the validation of HAV that, arguably, only counterfactual analyses can meet. In fact, NHTSA requires that all components of the HAV system (including each software update) are validated before they are implemented in the real world. As NFOTs or real-world controlled experiments would be unfeasible because they would both take too long and not be economically justifiable, simulations would most likely dominate the validation process for automated driving. Running counterfactual analyses, such as the ones proposed in this thesis, may become the standard tool for automated driving validation, paving the way for new applications and further development of the methodology introduced here. Many counterfactual simulations will validate only the technical performance (e.g., sensor limitations and failures), without explicit models of driver behavior (Nilsson, 2014).

However, while automated driving will limit the driver's involvement in the driving task, it will not eliminate the need for behavior analyses and road-user (including the driver) models. For example, for many automated vehicle implementations, drivers are expected to take over control from the vehicle under specific conditions (transition of control; Gold, Damböck, Lorenz, & Bengler, 2013; Seppelt & Victor, 2016). In addition, interactions between HAV and other road users are particularly critical for automated driving, especially in mixed-traffic situations.

Thus, more research is clearly needed to develop models of driver behavior in automated vehicles, including HAV, as well as to continue the development of the methods presented in this thesis. A starting point for the former is the understanding of, for example, glance behaviors and reaction processes in different contexts (e.g., with and without autonomous driving); the work by Morando, Victor, and Dozza (2016) is an example of a step in the right direction. This paper can be complemented by studies similar to those by Markkula et al. (2016) and Victor et al. (2015), with the scope extended to other crash scenarios (including interactions with vulnerable road users, such as pedestrians and bicyclists). Further, models of drivers' task-switching performance (Gold et al., 2013) will also be an integral part of counterfactual simulations of automated vehicles.

The front-cover image

We live in a time of change. Vehicles are becoming more and more automated, and the behavior of drivers (or, eventually, riders) is likely to change rapidly in the next several decades due to automation. How will the children of today think of vehicles when they retire? It is not very likely that ISS will make riders of future vehicles so completely safe that passive safety systems are not needed, so that vehicle occupants can act like those in the cover image—completely carefree, without worrying about the outside environment or any risk of crashing. Or, can human transportation eventually be that safe? If so, how far off is this reality?

It is not only automation that changes driver behavior. For example, the “need” to constantly interact with new devices and communicate digitally is affecting our behavior and changing the landscape of distraction. Is it possible to legislate away the problem in the long term? Can we expect our children, who are growing up with constant access to digital media and new ways to interact, to resist these temptations while driving (an extreme case is demonstrated on the cover image)? Or, do we need to focus on developing ISS and roadways that allow for such behavior? I believe we do. We are already moving in that direction (towards full automation), but—as long as the driver has some responsibility in the vehicle, or other human road users are present in the traffic environment—we need to develop methods to understand and evaluate driver and other road-users behavior in relation to safety, as well as to ISS. Addressing this need is what this thesis is all about.

(The car in the cover image was parked safely (off-course) with the ignition off, and both parents of the children had given their consent to the use of the image/photo for this thesis cover – thank you Ina and Erik.)

5 Conclusions

In accordance with its aim to develop methods for naturalistic driving data (NDD) analysis, this thesis provides a complementary, innovative set of methods capable of taking advantage of the vast store of information in NDD for traffic safety research and pre-crash safety system development. The methods address five specific scientific knowledge gaps. First, potential biases due to segment duration and autocorrelation in the analysis of NDD have rarely been acknowledged or understood (Paper I). Second, expert-assessment methods for understanding crash causation are rare and have typically not been optimized for analyzing NDD with video (Paper II). Third, a pragmatic method to extract range and range-rate in rear-end crash scenarios has been lacking (Paper III). Fourth, there has been a dearth of understanding regarding the importance of the choice of driver model in safety benefit analyses using counterfactual simulations (Paper IV). Fifth, and finally, a method for studying the effect of the combination of pre-crash kinematics and driver behavior has been missing (Paper V).

The methods can be further developed in the future, and applied in the automotive industry and government. Traffic safety research can be improved by the contribution from this thesis, particularly in terms of (Paper I) facilitating robust analysis of continuous NDD and identification of biases in NDD analysis, (Paper II) providing tools to understand why crashes and near-crashes occur, (Paper III) enabling extraction of range, range rate and optical parameters from commercially collected NDD (CNDD), (Paper IV) demonstrating the importance of driver models in counterfactual safety benefit evaluation, and (Paper V) enabling the estimation of crash and injury risks related to different glance behaviors.

The methods support different parts of the pre-crash safety measure development process. For example, the methods presented in Papers I-III and V can help set requirements for pre-crash safety measures such as intelligent safety systems (ISS, e.g., different levels of automated driving), infrastructure design, behavior-based safety, and policy-making. The methods can also be used in the development of pre-crash safety measures (e.g., specific ISS products or prototypes; Papers IV and V), and can contribute to benefit evaluation, either directly (Paper I, enabling robust analyses) or indirectly (e.g. Paper IV, providing insights into methodological prerequisites). Further, this thesis also provides a novel framework for estimating the combined effects of driver behavior and pre-crash kinematics on safety (Paper V)—facilitating rapid evaluation of driver-vehicle interaction designs and emerging driver behaviors in the real world. The methods of Papers I-V rest on a foundation of general methodological aspects related to the analysis of NDD. Thus, the thesis also provides a discussion on three topics relevant to the analysis of NDD but which have seldom been discussed in the scientific community: (a) the use of near-crashes as surrogates for crashes—near-crashes have several uses in traffic safety research, but care needs to be taken when using near-crashes them as surrogates for crashes;

(b) inferring causation based on NDD—observational data of some kind are needed to study crash causation, and NDD provide an unprecedented level of detailed information about the pre-crash phase, including detailed data of driver behavior (e.g., video and time-series of vehicle kinematic); (c) the use of CNDD—foreseen to play a larger role in future research on driver behavior because of their large datasets, with detailed records of the driver, the vehicle, and the environment in the last few seconds before a crash.

As the importance of driver behavior increases in the design of vehicles (and, in particular, safety-measure design), utilizing the methods in this thesis is likely to result in safer vehicles and better safety measures developed in a shorter time—helping to reduce the number of crashes and injuries on our roads and reach the Vision Zero target.

References

Aksan, N., Schall, M., Anderson, S., Dawson, J., Tippin, J., & Rizzo, M. (2013). Can intermittent video sampling capture individual differences in naturalistic driving? *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment Training and Vehicle Design.* Bolton Landing, NY. (pp. 135–141). Retrieved from http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/022_Aksan_0.pdf

Al-Sultan, S., Al-Bayatti, A. H., & Zedan, H. (2013). Context-aware driver behavior detection system in intelligent transportation systems. *IEEE Transactions on Vehicular Technology. 62*(9), 4264-4275. doi:10.1109/TVT.2013.2263400

Andersson, J., Lannér, G., Wink, W., Halleman, B., Naing, C. L., Thomson, R., . . . Hill, J. (2005). *Roadside infrastructure for safer European roads: D08 European best practice for roadside design: guidelines for maintenance and operations of roadside infrastructure - deliverable 08 in the European poject Roadside Infrastructure for Safer European Roads (RISER)*. Chalmers University of Technology. Gothenburg, Sweden. Retrieved from https://dspace.lboro.ac.uk/2134/2204

Angell, L., Auflick, J., Austria, P. A., Kochhar, D., Tijerina, L., Biever, W., . . . Kiger, S. (2006). *Driver Workload Metrics Task 2 Final Report*. (DOT HS 810 635). National Highway Traffic Safety Administration. Washington, DC. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/Driver%20Distraction/Driver%20Workload%20Metrics%20Final%20Report.pdf

Bagdadi, O., & Várhelyi, A. (2013). Development of a method for detecting jerks in safety critical events. *Accident Analysis & Prevention. 50*, 83-91. doi:10.1016/j.aap.2012.03.032

Bálint, A., Fagerlind, H., & Kullgren, A. (2013). A test-based method for the assessment of pre-crash warning and braking systems. *Accident Analysis & Prevention. 59*, 192-199. doi:10.1016/j.aap.2013.05.021

Bao, S., LeBlanc, D., Sayer, J., & Flannagan, C. (2012). The impact of integrated crash warning system on heavy truck drivers' time headway maintenance. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 54*(3) doi:10.1177/0018720812439412

Barnard, Y., Utesch, F., van Nes, N., Eenink, R., & Baumann, M. (2015). The study design of UDRIVE: the naturalistic driving study across Europe for cars, trucks and scooters. *European Transport Research Review. 8*(2) doi:10.1007/s12544-016-0202-z

Beauchamp, T. L., & Rosenberg, A. (1981). *Hume and the problem of causation*. New York: Oxford University Press.

Benderius, O. (2014). *Modelling driver steering and neuromuscular behaviour.* (Doctoral dissertation), Chalmers University of Technology, Göteborg, Sweden. (ISBN:978-91-7597-090-5).

Benmimoun, M., Ljung Aust, M., Faber, F., & Saint Pierre, G. (2011). *Safety analysis method for assessing the impacts of advanced driver assistance systems within the European large scale field test "euroFOT"*. Paper presented at the ITS European Congress, Lyon, France. Retrieved from http://www.eurofot-ip.eu/download/papersandpresentations/its_european_congress/mohamed_benmimoun.pdf

Bezzina, D., & Sayer, J. (2015). *Safety pilot model deployment: Test conductor team report*. (DOT HS 812 171). National Highway Traffic Safety Administration. Washington, DC. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2015/812171-SafetyPilotModelDeployDelTestCondrTmRep.pdf

Blatt, A., Pierowicz, J. A., Flanigan, M., Lin, P.-S., Kourtellis, A., Lee, C., . . . Hoover, M. (2015). *SHRP2 safety research - Naturalistic driving study: Field data collection*. Transportation Research Board at the National Academy of Sciences. Washington, DC. ISBN: 978-0-309-27393-0. Retrieved from http://www.trb.org/Main/Blurbs/170888.aspx

Blower, D., & Campbell, K. L. (2005). *Methodology of the large truck crash causation study*. (FMCSA-RI-05-035). Federal Motor Carrier Safety Administration. Washington, DC. Retrieved from http://ntl.bts.gov/lib/51000/51300/51398/FMCSA-RI-05-035.pdf

Boström, O., Fredriksson, R., Håland, Y., Jakobsson, L., Krafft, M., Lövsund, P., . . . Svensson, M. Y. (2000). Comparison of car seats in low speed rear-end impacts using the BioRID dummy and the new neck injury criterion (NIC). *Accident Analysis & Prevention. 32*(2), 321-328. doi:10.1016/S0001-4575(99)00105-0

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Boyle, L. N., & Lee, J. D. (2010). Using driving simulators to assess driving safety. *Accident Analysis & Prevention. 42*(3), 785-787. doi:10.1016/j.aap.2010.03.006

Boyle, L. N., Lee, J. D., Heyens, K. M., McGehee, D. V., Hallmark, S., & Ward., N. J. (2009). *S02 integration of analysis methods and development of analysis plan: Phase i report*. University of Iowa. doi:10.17226/22847

Brolin, K. (2016). Human body models. In G. Tiwari & D. Mohan (Eds.), *Transport planning and traffic safety - making cities, roads, and vehicles safer* (pp. 175-186). Boca Raton, FL: CRC Press.

Bronrott, B. (2010). *FMCSA uses of naturalistic driving data*. Paper presented at the 2nd International Naturalistic Driving Research Symposium, Blacksburg, VA. Retrieved from https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/docs/Naturalistic-Driving-Research.pdf

Brown, T. L., Lee, J. D., & McGehee, D. V. (2001). Human performance models and rear-end collision avoidance algorithms. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 43*(3), 462-482. doi:10.1518/001872001775898250

Brännström, M., Sjöberg, J., & Coelingh, E. (2008, 4-6 June 2008). A situation and threat assessment algorithm for a rear-end collision avoidance system*. Proceedings of the Intelligent Vehicles Symposium, 2008 IEEE.* Eindhoven, Netherlands. (pp. 102-107). doi:10.1109/IVS.2008.4621250.

Bärgman, J., Guo, F., Jovanis, P., & Knipling, R. R. (2016). Using naturalistic driving study data for road safety analysis: Connecting crashes to safety-critical events. Washington, DC: Panel debate at the Transportation Research Board (TRB) 96th Annual Meeting.

Bärgman, J., & Piccinini, G. (2016). Glance behavior for teen drivers in left turn across path / opposite direction crashes *Manuscript in preparation*

Bärgman, J., Smith, K., & Werneke, J. (2015). Quantifying drivers' comfort-zone and dread-zone boundaries in left turn across path/opposite direction (LTAP/OD) scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour. 35*, 170-184. doi:10.1016/j.trf.2015.10.003

Caird, J. K., & Hancock, P. A. (1994). The perception of arrival time for different oncomming vehicles at an intersection. *Ecological Psychology. 6*(2), 83-109. doi:10.1207/s15326969eco0602_1

Carney, C., McGehee, D., Harland, K., Weiss, M., & Raby, M. (2015). *Using naturalistic driving data to assess the prevalence of environmental factors and driver behaviors in teen driver crashes*. AAA Foundation for Traffic Safety. Washington, DC. Retrieved from https://www.aaafoundation.org/sites/default/files/2015TeenCrashCausationReport.pdf

Carsten, O., Fowkes, M., Lai, F., Chorlton, K., Jamson, S., Tate, F., & Simpkin, B. (2008). *ISA - UK Intelligent speed adaptation. Final report* University of Leeds. Leeds, UK. Retrieved from http:/www.dft.gov.uk/pgr/roads/vehicles/intelligentspeedadaptation/fullreport.pdf

Carsten, O., Kircher, K., & Jamson, S. (2013). Vehicle-based studies of driving in the real world: The hard truth? *Accident analysis and prevention. 58*, 162-174. doi:10.1016/j.aap.2013.06.006

Chang, L.-Y., & Wang, H.-W. (2006). Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis & Prevention. 38*(5), 1019-1027. doi:10.1016/j.aap.2006.04.009

Christie, R. (2001). The effectiveness of driver training as a road safety measure: an international review of the literature*. Proceedings of the Road Safety Research, Policing and Education Conference.* Melbourne, Victoria, Australia

Cicchino, J. B. (2016). *Effectiveness of forward collision warning systems with and without autonomous emergency braking in reducing police-reported crash rates*. Insurance institute for highway safety. Retrieved from http://www.iihs.org/frontend/iihs/documents/masterfiledocs.ashx?id=2124

Cryer, P. C., Westrup, S., Cook, A. C., Ashwell, V., Bridger, P., & Clarke, C. (2001). Investigation of bias after data linkage of hospital admissions data to police road traffic crash reports. *Injury Prevention. 7*(3), 234-241. doi:10.1136/ip.7.3.234

Davis, G. A., Hourdos, J., Xiong, H., & Chatterjee, I. (2011). Outline for a causal model of traffic conflicts and crashes. *Accident Analysis & Prevention. 43*(6), 1907-1919. doi:10.1016/j.aap.2011.05.001

Daza, I. G., Hernandez, N., Bergasa, L. M., Parra, I., Yebes, J. J., Gavilan, M., . . . Sotelo, M. A. (2011, 5-7 Oct. 2011). Drowsiness monitoring based on driver and driving data fusion*. Proceedings of the 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC).* Washington, DC. (pp. 1199-1204). doi:10.1109/ITSC.2011.6082907.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., . . . Jermeland, J. (2006). *The 100-car naturalistic driving study, Phase II – results of the 100-car field experiment*. (Contract No. DTNH22-00-C-07007). National Highway Traffic Safety Administration. Washington, DC.

Distner, M., Bengtsson, M., Broberg, T., & Jakobsson, L. (2009). City safety - a system addressing rear-end collisions at low speeds*. Proceedings of the 21st International technical conference on the enhanced safety of vehicles (ESV); Paper nr. 09-0371.* Stuttgart, Germany.

Dozza, M. (2013). What factors influence drivers' response time for evasive maneuvers in real traffic? *Accident Analysis & Prevention. 58*, 299-308. doi:10.1016/j.aap.2012.06.003

Dozza, M. (2016). What is the relation between crashes from crash databases and near-crashes from naturalistic data? *Manuscript in preparation*

Dozza, M., Bimpas, M., Kessler, C., Petersson, M., Amditis, A., & Tattegrain, H. (2010). *Eurofot: Constraints and trade-offs in testing hypotheses*. Paper presented at the Transport Research Arena Europe conference, Brussels, Belgium. Retrieved from Retrieved from: http://eurofot-ip.eu/download/events/TRA/papers/eurofot_constrains_and_trade_offs_in_testing_dozza.pdf

Dozza, M., & González, N. P. (2013). Recognising safety critical events: Can automatic video processing improve naturalistic data analyses? *Accident Analysis & Prevention. 60*, 298-304. doi:10.1016/j.aap.2013.02.014

Dozza, M., & Werneke, J. (2014). Introducing naturalistic cycling data: What factors influence bicyclists' safety in the real world? *Transportation Research Part F: Traffic Psychology and Behaviour. 24*, 83-91. doi:10.1016/j.trf.2014.04.001

Dunn, N. J., Hickman, J. S., & Hanowski, R. J. (2014). *Crash trifecta: A complex driving scenario describing crash causation* (14-UI-025). National Surface Transportation Safety Center for Excellence. Blacksburg, VA. Retrieved from https://vtechworks.lib.vt.edu/bitstream/handle/10919/49685/NSTSCE%20Final%20Report_Crash%20Trifecta_FINAL.pdf;sequence=1

Eiríksdóttir, H. (2016). *Quantitative analysis of rear-end crash causation mechanisms based on naturalistic crash data.* (Masters thesis), Chalmers University of Technology, Göteborg, Sweden. (ISSN 1652-8557).

Elvik, R. (2003). Assessing the validity of road safety evaluation studies by analysing causal chains. *Accident Analysis & Prevention. 35*(5), 741-748. doi:10.1016/S0001-4575(02)00077-5

Engström, J., Aust, M. L., & Viström, M. (2010). Effects of working memory load and repeated scenario exposure on emergency braking performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 52*(5), 551-559. doi:10.1177/0018720810381072

Engström, J., Bärgman, J., Nilsson, D., Seppelt, B., Markkula, G., Piccinini, G., & Victor, T. (2016a). Great expectations: A predictive processing account of automobile driving. Submitted for publication

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F. 8*(2), 97-120. doi:10.1016/j.trf.2005.04.012

Engström, J., Markkula, G., Victor, T., & Merat, N. (2016b). Effects of cognitive load on driving performance: The cognitive control hypothesis. Submitted for publication

Engström, J., Monk, C. A., Hanowski, R. J., Horrey, W. J., Lee, J. D., McGehee, D. V., . . . Yang, D. (2013). *A conceptual framework and taxonomy for understanding and categorizing driver inattention.* US-EU ITS Cooperation, Driver Distraction and HMI Working Group. Retrieved from http://ec.europa.eu/digital-agenda/en/news/us-eu-inattention-taxonomy-report

Erbsmehl, C. (2009). Simulation of real crashes as a method for estimating the potential benefits of advanced safety technologies *Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles (ESV); Paper nr. 09-0162.* Stuttgart, Germany.

Euro_NCAP. (2015). Study confirms high effectiveness of low speed autonomous emergency braking (AEB). Retrieved from http://www.euroncap.com/sv/press-media/pressmeddelanden/study-confirms-high-effectiveness-of-low-speed-autonomous-emergency-braking-aeb/

European_Communities. (2007). *Commission recommendation on safe and efficient in-vehicle information and communication systems; update of the european statement of principles on human machine interface*. (2007/78/EC; CELEX nr: 32007H0078). Commission of the European Communities. Retrieved from http://eur-lex.europa.eu/legal-content/AUTO/?uri=CELEX:32007H0078&qid=1477908644261&rid=1

Faber, F., Jonkers, E., Ljung Aust, M., Benmimoun, M., Regan, M., Jamson, S., & Dobberstein, J. (2011). *Deliverable D6.2: Analysis methods for user related aspects and impact assessment on traffic safety, traffic efficiency and environment*. euroFOT consortium. Brussles, Belgum. Retrieved from http://www.eurofot-ip.eu/download/library/deliverables/eurofotsp620110628v10dld62_analysis_methods.pdf

Fagerlind, H., Martinsson, J., & Hagström, L. (2010). INTACT - national initiative for enhanced crash investigations*. Proceedings of the 4th International Conference ESAR - Expert Symposium on Accident Research.* Hannover, Germany.

Fair, D. (1979). Causation and the flow of energy. *Erkenntnis. 14*(3), 219-250. doi:10.1007/bf00174894

Fajen, B. R., & Warren, W. H. (2003). Behavioral dynamics of steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance. 29*(2), 343-362. doi:10.1037/0096-1523.29.2.343

Fancher, P., Ervin, R., Sayer, J., Hagan, M., Bogard, S., Bareket, Z., . . . Haugen, J. (1998). *Intelligent cruise control field operational test. Final report*. (DOT HS 808 849). The University of Michigan Transportation Research Institute. Ann Arbor, MI.

Festag, A., Le, L., & Goleva, M. (2011). Field operational tests for cooperative systems: A tussle between research, standardization and deployment. *Proceedings of the 8th ACM international workshop on Vehicular inter-networking.* Las Vegas, NV. (pp. 73-78). doi:10.1145/2030698.2030710.

Fisher, R. (1957). Alleged dangers of cigarette-smoking. *British Medical Journal. II*, 43, 297-298

Fitch, G. M., Hanowski, R. J., & Guo, F. (2014). The risk of a safety-critical event associated with mobile device use in specific driving contexts. *Traffic Injury Prevention. 1*, 1-34. doi:10.1080/15389588.2014.923566

Foster, J., Kortge, J., & Wolanin, M. (1977). Hybrid iii-a biomechanically-based crash test dummy. *Proceedings of the 21st Stapp Car Crash Conference.* New Orleans, LA. doi:10.4271/770938.

French, D. J., West, R. J., Elander, J., & Wilding, J. M. (1993). Decision-making style, driving style, and self-reported involvement in road traffic accidents *Ergonomics. 36*(6), 627-644. doi:10.1080/00140139308967925

Funke, J., Srinivasan, G., Ranganathan, R., & Burgett, A. (2011). Safety impact methodology (SIM): application and results of the advanced crash avoidance technologies (ACAT) program. *Proceedings of the 22nd International Technical Conference on the Enhanced Safety of Vehicles (ESV). Paper nr. 11-0367-O.* Washington, D.C.

Gabauer, D., & Gabler, H. (2006). Comparison of delta-v and occupant impact velocity crash severity metrics using event data recorders. *Association for the Advancement of Automotive Medicine 50th Annual Proceedings*, 57-71. PCMID:3217488

Gabler, C., Hampton, C., & Roston, T. (2003). Estimating crash severity: Can event data recorders replace crash reconstruction? *Proceedings of the 18th International Conference on Enhanced Safety of Vehicles (ESV.); Paper nr. 490.* Nagoya, Japan.

Gennarelli, T. A., & Wodzin, E. (2006). AIS 2005: A contemporary injury scale. *Injury. 37*(12), 1083-1091. doi:10.1016/j.injury.2006.07.009

Georgi, A., Zimmermann, M., Lich, T., Blank, L., Kickler, N., & Marchthaler, R. (2009). New approach of accident benefit analysis for rear end collision avoidance and mitigation systems *Proceedings of the 21st International Conference on Enhanced Safety of Vehicles (ESV); Paper nr. 09-0281. .* Stuttgart, Germany.

Georgopoulos, A. P. (1986). On reaching. *Annual Review of Neuroscience. 9*, 147-170. doi:10.1146/annurev.ne.09.030186.001051

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile driving. *The American Journal of Psychology. 51*(3), 453-471. doi:10.2307/1416145

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). “Take over!” How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* (pp. 1938-1942). doi:10.1177/1541931213571433. Retrieved from http://pro.sagepub.com/content/57/1/1938.abstract

Gordon, T., Bareket, Z., Kostyniuk, L., Barnes, M., Hagan, M., Kim, Z., . . . Vayda, A. (2012). *Site-based video system design and development*. (S2-S09-RW-1). Transportation Research Board at the National Academy of Sciences. Washington, DC. Retrieved from http://www.trb.org/Publications/Blurbs/166323.aspx

Gordon, T., Blankespoor, A., Barnes, B., Blower, D., Green, P., & Kostyniuk, L. (2009). Yaw rate error - a dynamic measure of lane keeping control performance for the retrospective analysis of naturalistic driving data. *Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles (ESV); Paper nr. 09-0326.* Stuttgart, Germany.

Gordon, T., Sardar, H., Blower, D., Ljung Aust, M., Bareket, Z., Barnes, M., . . . Theander, H. (2010). *Advanced crash avoidance technologies (ACAT) program – final report of the Volvo-Ford-UMTRI project: Safety impact methodology for lane departure warning – method development and estimation of benefits*. (DOT HS 811 405). US Department of Transportation. National Highway Traffic Safety Administration. Washington, DC. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2010/811405.pdf

Gorman, T., Loren, S., & Hankey, J. (2015). *S31: NDS data dissemination activities. Task 1.6: Radar post-processing*. Virginia Tech Transportation Institute. Blacksburg, VA.

Green, M. (2000). "How long does it take to stop?" Methodological analysis of driver perception-brake times. *Transportation Human Factors. 2*(3), 195-216. doi:10.1207/STHF0203_1

Green, P., Cullinane, B., Zylstra, B., & Smith, D. (2004). *Safety vehicles using adaptive interface technology: Typical values for driving performance with emphasis on the standard deviation of lane position: A summary of the literature*. SAVE-IT consortium. Retrieved from http://www.umich.edu/~driving/publications/litrev_3a.pdf

Guo, F., & Fang, Y. (2013). Individual driver risk assessment using naturalistic driving data. *Accident Analysis & Prevention. 61*, 3-9. doi:10.1016/j.aap.2012.06.014

Guo, F., Klauer, S. G., McGill, M. T., & Dingus, T. A. (2010). *Evaluating the relationship between near-crashes and crashes: can near-crashes serve as a surrogate safety metric for crashes?* (DOT HS 811 382) Washington, DC. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NVS/Human%20Factors/Safety%20Problem%20Identification/DOT%20HS%20811%20382.pdf

Habibovic, A., Tivesten, E., Uchida, N., Bärgman, J., & Ljung Aust, M. (2013). Driver behavior in car-to-pedestrian incidents: An application of the Driving Reliability and Error Analysis Method (DREAM). *Accident Analysis & Prevention. 50*, 554–565. doi:10.1016/j.aap.2012.05.034

Hallmark, S., Hsu, Y.-Y., Boyle, L. N., Carriquiry, A., Tian, Y., & Mudgal, A. (2011). *Evaluation of data needs, crash surrogates, and analysis methods to address lane departure research questions using naturalistic driving study data*. (S2-S01E-RW-1) Washington, DC. Retrieved from http://onlinepubs.trb.org/onlinepubs/shrp2/SHRP2_S2-S01E-RW-1.pdf

Hallmark, S., Oneyear, N., Tyner, S., Wang, B., Carney, C., & Mcgehee, D. (2015). *Analysis of naturalistic driving study data: Roadway departures on rural two-lane curves*. (S2-S08D-RW-1.). Transportation research board at the National Academy of Sciences. Washington, DC. ISBN: 978-0-309-27418-0. Retrieved from http://www.trb.org/Main/Blurbs/171254.aspx

Hancock, P. A., Caird, J. K., & Johnson, S. B. (1991). The left turn*. Proceedings of the International Ergonomics Association.* Paris, France. (pp. 1428-1430). Retrieved from http://www.peterhancock.ucf.edu/Downloads/ref_con_pubs/Hancock_Caird_Johnson_1991.pdf

Hautzinger, H., Pastor, C., Pfeiffer, M., & Schmidt, J. (2007). *TRACE Deliverable 7.3: Analysis methods for accident and injury risk studies*. (TRACE D7.3). Traffic Accident Causation in Europe (TRACE) consortium. Brussles, Belgium. Retrieved from http://www.trace-project.org/publication/archives/trace-wp7-d7-3.pdf

Heinrich, C. (2015). Fighting driver distraction – recent developments 2013 - 2015*. Proceedings of the 24th Internatioal Technical Conference on enhanced safety of vehicles (ESV); Paper nr. 15-0359.* Göteborg, Sweden.

Herslund, M.-B., & Jørgensen, N. O. (2003). Looked-but-failed-to-see-errors in traffic. *Accident Analysis & Prevention. 35*(6), 885-891. doi:10.1016/S0001-4575(02)00095-7

Hickman, J. S., & Hanowski, R. J. (2010). *Evaluating the safety benefits of a low-cost driving behavior management system in commercial vehicle operations*. (FMCSA-RRR-10-033). U.S. Department of Transportation - Federal Motor Carrier Safety Administration. Blacksburg, VA. Retrieved from http://media.cygnus.com/files/cygnus/whitepaper/VSPC/2011/AUG/drivingbehaviormanagementsyste_10336417.pdf

Hickman, J. S., Hanowski, R. J., & Bocanegra, J. (2010). *Distraction in commercial trucks and buses: Assessing prevalence and risk in conjunction with crashes and near-*

crashes. (Contract No. DTMC75-09-J-00045). FMCSA. Washington, DC. Retrieved from http://ntl.bts.gov/lib/51000/51200/51287/Distraction-in-Commercial-Trucks-and-Buses-report.pdf

Hickman, J. S., Knipling, R. R., Hanowski, R. J., Wiegand, D. M., Inderbitzen, R. E., & Bergoffen, G. (2007). *Commercial truck and bus safety - impact of behavior-based safety techniques on commercial motor vehicle drivers: A synthesis of safety practice*. (ISBN: 978-0-309-42269-7). Transportation Research Board. Washington, DC. doi:10.17226/23193.

Hickman, J. S., Soccolich, S., Fitch, G., & Hanowski, R. (2015). *Driver distraction: Eye glance analysis and conversation workload*. (FMCSA-RRR-14-001). U.S. Department of Transportation: Federal Motor Carrier Safety Administration. Retrieved from http://ntl.bts.gov/lib/56000/56200/56289/14-001-Cognitive_Distraction-FINAL-NOV_2015.pdf

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.

Horrey, W. J., Lesch, M. F., & Garabet, A. (2009). Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. *Journal of Safety Research. 40*(1), 7-12. doi:10.1016/j.jsr.2008.10.011

Houtenbos, M. (2008). *Expecting the unexpected.* (Doctoral dissertation), Delft University of Technology, Delft. (ISBN:978-90-73946-03-3).

Hydén, C. (1987). *The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique*. (Bulletin 70. B0007B9WR4). Lund University, Department of Traffic Planning and Engineering. Lund, Sweden.

Imberg, H., Kovaceva, J., Bärgman, J., & Nerman, O. (2016). Stratified sampling for assessment and weighted statistical analysis of safety critical events in naturalistic driving data. *Manuscript in preparation*

Imberg, H., Liskovskaja, V., Selpi, & Nerman, O. (2016). Optimal auxilary variable assisting two-phase sampling designs for naturalistic driving applications. *Manuscript in preparation*

Isaksson-Hellman, I., & Lindman, M. (2012). The effect of a low-speed automatic brake system estimated from real life data. *Annals of Advances in Automotive Medicine / Annual Scientific Conference. 56*, 231-240. PCMID:3503425

ISO_11898. (2003). *Road vehicles -- Controller area network (CAN) -- Part 1: Data link layer and physical signalling*. (ISO 11898-1:2003). International Organization for Standardization. Retrieved from http://www.iso.org/

Jagacinski, R. J., & Flach, J. M. (2002). *Control theory for humans*. New Jersey: Lawrence Erlbaum Associates, Inc.

Jakobsson, L., Lindman, M., Svanberg, B., & Carlsson, H. (2010). Real world data driven evolution of volvo cars' side impact protection systems and their effectiveness. *Annals of Advances in Automotive Medicine / Annual Scientific Conference. 54*, 127-136. PCMID:3242537

Jonasson, J. K., & Rootzén, H. (2014). Internal validation of near-crashes in naturalistic driving studies: A continuous and multivariate approach. *Accident Analysis & Prevention. 62*(0), 102-109. doi:10.1016/j.aap.2013.09.013

Kidd, D. G., & McCartt, A. T. (2015). The relevance of crash type and severity when estimating crash risk using the SHRP2 naturalistic driving data*. Proceedings of the 4th International Conference on Driver Distraction and Inattention.* Sydney, Australia.

Kiefer, K. J., Cassar, M. T., Flannagan, C., LeBlanc, D., Palmer, M. D., Deering, R. K., & Shulman, M. A. (2003). *Forward collision warning requirements project: Refining the camp crash alert timing approach by examining "last-second" braking and lane change maneuvers under various kinematic conditions*. (DOT HS 809 574). Office of

Advanced Safety Research. Washington, DC. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/2003/HS809574Report.pdf

Kircher, K., Patten, C., & Ahlström, C. (2011). *Mobile telephones and other communication devices and their impact on traffic safety: A review of the literature*. (VTI rapport 729A). VTI. Linköping, Sweden.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine. 370*(1), 54-59. doi:10.1056/NEJMsa1204142

Klauer, S. G., Perez, M. A., & MacClaggerty, J. (2011). Naturalistic driving studies and data coding and analysis techniques. In B. Porter (Ed.), *Handbook of traffic pshycology* (pp. 73-86). Waltham, MA: Elsevier

Knipling, R. R. (2015). Naturalistic driving events: No harm, no foul, no validity *Proceedings of the 8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* Snowbird, Salt Lake City, UT. (pp. 197-203).

Koustanaï, A., Boloix, E., Van Elslande, P., & Bastien, C. (2008). Statistical analysis of "looked-but-failed-to-see" accidents: Highlighting the involvement of two distinct mechanisms. *Accident Analysis & Prevention. 40*(2), 461-469. doi:10.1016/j.aap.2007.08.001

Kuge, N., Yamamura, T., Shimoyama, O., & Liu, A. (2000). *A driver behavior recognition method based on a driver model framework*. doi:10.4271/2000-01-0349

Kullgren, A. (2008). Dose-response models and EDR data for assessment of injury risk and effectiveness of safety systems*. Proceedings of the IRCOBI Conference.* Bern, Switzerland.

Kusano, K., & Gabler, H. (2010). Potential occupant injury reduction in pre-crash system equipped vehicles in the striking vehicle of rear-end crashes. *Ann Adv Automot Med. 54*, 203-214. PCMID:3242552

Lardelli-Claret, P., Jimenez-Moleo, J. J., Luna-del-Castillo, J. d. D., Garcıa-Martın, M., Moreno-Abril, O., & Bueno-Cavanillas, A. (2005). Comparison between two quasi-induced exposure methods for studying risk factors for road crashes. *American Journal of Epidemiology. 163*(2) doi:10.1093/aje/kwj015

Laureshyn, A. (2010). *Application of automated video analysis to road user behaviour.* (Doctoral dissertation), Lund University, Lund, Sweden. (ISBN:978-91-628-8003-3).

LeBlanc, D., Bao, S., Sayer, J., & Bogard, S. (2013). Longitudinal driving behavior with integrated crash-warning system: Evaluation from naturalistic driving data. *Transportation Research Record: Journal of the Transportation Research Board Annual Meeting. 2365* doi:10.3141/2365-03

LeBlanc, D., Sayer, J., Winkler, C., Ervin, R., Bogard, S., Devonshire, J., . . . Gordon, T. (2006). *Road departure crash warning system field operational test: Methodology and results*. (UMTRI-2006-9-1). The University of Michigan Transportation Research Institute. Ann Arbor, MI. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/2006/RDCW-Final-Report-Vol-1_JUNE.pdf

LeBlanc, D., Sivak, M., & Bogard, S. (2010). *Using naturalistic driving data to assess variations in fuel efficiency among individual drivers* The University of Michigan sustainable worldwide transportation Ann Arbor, MI. Retrieved from https://deepblue.lib.umich.edu/handle/2027.42/78449

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception. 5*(4), 437–459. doi:10.1068/p050437

Lee, J. D. (2008). Fifty years of driving safety research. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 50*(3), 521-528. doi:10.1518/001872008X288376

Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 43*(4), 631-640. doi:10.1518/001872001775870340

Lefler, D. E., & Gabler, H. C. (2004). The fatality and injury risk of light truck impacts with pedestrians in the United States. *Accident Analysis & Prevention. 36*(2), 295-304. doi:10.1016/S0001-4575(03)00007-1

Lesemann, M., Zlocki, A., Aparicio, A., Jacobson, J., Eriksson, H., Bruzelius, F., & Nordström, L. (2011). *Deliverable D 5.2. Final report. Testing and evaluation methods for ict-based safety systems*. (eVALUE-101231-D52-V10-FINAL). Retrieved from http://transport-research.info/sites/default/files/project/documents/20121024_101220_97514_evalue-080930-d21-v04-final.pdf

Liang , Y., Lee, J. D., & Yekhshatyan, L. (2012). How dangerous is looking away from the road? Algorithms predict crash risk from glance patterns in naturalistic driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 54*(6), 1104-1116. doi:10.1177/0018720812446965

Lich, T., & Georgi, A. (2011). Use of drive event recorder data for evaluation of rear end collision avoidance and mitigation systems. *Proceedings of the 1st International Symposium on Future Active Safety Technology towards zero-traffic-accidents.* Tokyo, Japan. (pp. 1-5).

Lich, T., Mettler, A., Georgi, A., Danz, C., Doopyo, L., Sooncheol, L., & Sul, J. (2012). Benefit analysis of predictive rear end collision avoidance and mitigation systems for South Korea using video documented accident data. *Proceedings of the The 11th International Symposium on Advanced Vehicle Control.* Seoul, South Korea.

Lie, A., Tingvall, C., Krafft, M., & Kullgren, A. (2006). The effectiveness of electronic stability control (ESC) in reducing real life crashes and injuries. *Traffic Injury Prevention. 7*(1), 38-43. doi:10.1080/15389580500346838

Lindman, M., & Tivesten, E. (2006). A method for estimating the benefit of autonomous braking systems using traffic accident data. *SAE Technical Paper* doi:10.4271/2006-01-0473

Ljung Aust, M., Dombrovskis, S., Kovaceva, J., Svanberg, B., & Ivarsson, J. (2013). An empirically based suggestion for reformulating the glance duration criteria in NHTSA's visual-manual interaction guidelines. *SAE 2013 World Congress & Exhibition - SAE Technical Paper* doi:10.4271/2013-01-0444

Ljung Aust, M., Regan, M., & Benmimoun, M. (2011). *Disentangling the effects of advanced driver assistance system functions in field operational tests: Recommendations from the European "euroFOT" project*. Paper presented at the ITS World Congress 2011, Lyon, France. Retrieved from http://www.eurofot-ip.eu/download/papersandpresentations/its_european_congress/iffstar.pdf

Lübbe, N. (2015). *Integrated pedestrian safety assessment: A method to evaluate combinations of active and passive safety.* (Doctoral dissertation), Chalmers University of Technology, Göteborg, Sweden. (ISBN:978-91-7597-296-1).

Lytx. (2016). Lytx - Delivering insights. Driving results. Retrieved from http://www.lytx.com/

Maciej, J., & Vollrath, M. (2009). Comparison of manual vs. speech-based interaction with in-vehicle information systems. *Accident Analysis & Prevention. 41*(5), 924-930. doi:10.1016/j.aap.2009.05.007

Malliaris, A. C., Hitchcock, R., & Hedlund, J. (1982). A search for priorities in crash protection*. Proceedings of the SAE International Congress and Exposition* Detroit, MI. (pp. 1-33). doi:10.4271/820242.

Margulies, S. S., & Thibault, L. E. (1992). A proposed tolerance criterion for diffuse axonal injury in man. *Journal of Biomechanics. 25*(8), 917-923. doi:10.1016/0021-9290(92)90231-O

Markkula, G. (2015). *Driver behavior models for evaluating automotive active safety: From neural dynamics to vehicle dynamics.* (Doctoral dissertation), Chalmers University of Technology, Göteborg, Sweden. (ISBN:978-91-7597-153-7).

Markkula, G., Benderius, O., Wolff, K., & Wahde, M. (2012). Quantitative models of driver behavior in near-collision situations: A review of models suitable for computer simulation. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 54*(6), 1117-1143.

Markkula, G., Benderius, O., Wolff, K., & Wahde, M. (2013). Effects of experience and electronic stability control on low friction collision avoidance in a truck driving simulator. *Accident Analysis & Prevention. 50*, 1266-1277. doi:10.1016/j.aap.2012.09.035

Markkula, G., Engström, J., Lodin, J., Bärgman, J., & Victor, T. (2016). A farewell to brake reaction times? Kinematics-dependent brake response in naturalistic rear-end emergencies. *Accident Analysis & Prevention. 95, Part A*, 209-226. doi:10.1016/j.aap.2016.07.007

McDonald, A. D., Lee, J. D., Aksan, N. S., Dawson, J. D., Tippin, J., & Rizzo, M. (2013). The language of driving. *Transportation Research Record: Journal of the Transportation Research Board. 2392*(1), 22-30. doi:10.3141/2392-03

McDowall, D., McCleary, R., Meidinger, E., & Hay, R. (1980). *Interrupted time series analysis*. London, UK: Sage publications Inc.

McGehee, D., Raby, M., Carney, C., Lee, J. D., & Reyes, M. L. (2007a). Extending parental mentoring using an event-triggered video intervention in rural teen drivers. *Journal of Safety Research 38*, 215–227. doi:10.1016/j.jsr.2007.02.009

McGehee, D. V., Carney, C., Raby, M., Lee, J. D., & Reyes, M. (2007b). Extending parental mentoring using and event-triggered video intervention in rural teen drivers. *Journal of Safety Research. 38* doi:10.1016/j.jsr.2007.02.009

McGehee, D. V., Carney, C., Raby, M., Lee, J. D., & Reyes, M. L. (2007c). The impact of an event-triggered video intervention on rural teenage driving*. Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment.* Stevenson, Washington. (pp. 565-571). Retrieved from http://drivingassessment.uiowa.edu/DA2007/PDF/091_McGeheeeCarney.pdf

McGehee, D. V., & Carsten, O. (2010). Perception and biodynamics in unalerted precrash response *Annals of advances in automotive medicine. 54*, 315-332. PCMID:3242562

McLaughlin, S., Hankey, J., & Dingus, T. (2008). A method for evaluating collision avoidance systems using naturalistic driving data. *Accident Analysis & Prevention. 40*, 8-16. doi:10.1016/j.aap.2007.03.016

Meng, L., & Wang, Y. (2016). *Tool development and quantitative analysis for naturalistic Left Turn Across Path/Opposite direction (LTAP/OD) driving scenarios.* (Masters thesis), Chalmers University of Technology, Göteborg, Sweden. (ISSN 1652-8557).

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour. 27, Part B*, 274-282. doi:10.1016/j.trf.2014.09.005

Michon, J. A. (1985). Critical review of driver behaviour models: What do we know? What should we do? In L. A. Evans & R. C. Schwing (Eds.), *Human Behaviour and Traffic Safety* (pp. 487–525). New York: Plenum Press. .

Migletz, D. J., Glauz, W. D., & Bauer, K. M. (1985). *Relationships between traffic conflicts and accidents*. (FHWA/RD-84/042). US Department of Transportation, Federal Highway Administration. McLean, VA.

Moeschlin, F. (2007). *A method for achieving adaptive incident detection in automotive safety.* (Masters thesis), Chalmers University of Technology, Göteborg, Sweden.

Mononen, P., Franzen, S., Pagle, K., Morris, A., Innamaa, S., Karlsson, M., . . . Fruttaldo, S. (2012). *Field operational test of aftermarket and nomadic devices in vehicles. Final report*. (INFSO-ICT 224067 Final report). TeleFOT Consortium and the European Commission

Morando, A., Victor, T., & Dozza, M. (2016). Drivers anticipate lead-vehicle conflicts during automated longitudinal control: Sensory cues capture driver attention and promote appropriate and timely responses. *Accident Analysis & Prevention. 97*, 206-219. doi:10.1016/j.aap.2016.08.025.

Muhrer, E., & Vollrath, M. (2011). The effect of visual and cognitive distraction on driver's anticipation in a simulated car following scenario. *Transportation Research Part F: Traffic Psychology and Behaviour. 14*(6), 555-566. doi:10.1016/j.trf.2011.06.003

Neale, V. L., Dingus, T. A., Klauer, S. G., Sudweeks, J., & Goodman, M. (2005). An overview of the 100-car naturalistic study and findings*. Proceedings of the 19th International Technical Conference on the Enhanced Safety of Vehicles (ESV); Paper nr. 05-0400.* Washington, DC.

NHTSA. (2012a). *Federal motor vehicle safety standards: Event data recorders*. National highway traffic safety administration at US Department of transportation. Washington, DC. Retrieved from http://www.gpo.gov/fdsys/pkg/FR-2012-12-13/pdf/2012-30082.pdf

NHTSA. (2012b). *Visual-manual nhtsa driver distraction guidelines for in-vehicle electronic devices*. (NHTSA-2010-0053). National Highway Traffic Safety Administration as the US Department of Transportation. Washington DC. Retrieved from http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Distraction_NPFG-02162012.pdf

NHTSA. (2013). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices*. (NHTSA-2010-0053). National highway traffic safety administration at US Department of transportation. Washington, DC. Retrieved from http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Distraction_NPFG-02162012.pdf

NHTSA. (2015). *National automotive sampling system (NASS) general estimates system (GES) analytical user's manual 1988-2014*. (DOT HS 812 215). National Highway Traffic Administration at the Department of Transportation. Washington, DC. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812215

NHTSA. (2016). *Federal automated vehicles policy: Accelerating the next revolution in roadway safety*. (12507-091216-v9). National highway traffic safety administration at the US department of transportation. Washington, DC. Retrieved from https://www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20guidance%20PDF.pdf

Niehoff, P., & Gabler, H. C. (2006). The accuracy of winsmash delta-v estimates: The influence of vehicle type, stiffness, and impact mode. *Annual Procedings Association Advancment of Automotive Medicine. 50*, 73–89. PCMID:3217484

Nilsson, J. (2014). *Computational verification methods for automotive safety systems.* (Doctoral dissertation), Chalmers University of Technology, Göteborg, Sweden. (ISBN:978-91-7385-973-8).

Näätänen, R., & Summala, H. (1974). A model for the role of motivational factors in drivers' decision-making. *Accident Analysis & Prevention. 6*, 243-261. doi:10.1016/0001-4575(74)90003-7

Olson, P. L., & Sivak, M. (1986). Perception-response time to unexpected roadway hazards. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 28*(1), 91-96. doi:10.1177/001872088602800110

Olson, R. L., Hanowski, R. J., Hickman, J. S., & J., B. (2009). *Driver distraction in commercial vehicle operations*. (FMCSA-RRR-09-042). Department of Transportation, Federal Motor Carrier Safety Administration. Retrieved from http://www.distraction.gov/

downloads/pdfs/driver-distraction-commercial-vehicle-operations.pdf
Othman, S., Thomson, R., & Lannér, G. (2012). Using naturalistic field operational tests (fot) data to identify horizontal curves. *Journal of Transportation Engineering. 138*(9), 1151–1160. doi:10.1061/(ASCE)TE.1943-5436.0000408
Othman, S., Thomson, R., & Lannér, G. (2014). Safety analysis of horizontal curves using real traffic data. *Journal of Transportation Engineering. 140*(4), 1-9. doi:10.1061/(ASCE)TE.1943-5436.0000626
Otte, D., Krettek, C., Brunner, H., & Zwipp, H. (2003). Scientific approach and methodology of new in-depth investigations study in Germany called GIDAS*. Proceedings of the 18th Interntional Technical Conference on The Enhanced Safety of Vehicles (ESV.); Paper nr. 161.* Nagoya, Japan.
Otte, D., Pund, B., & Jänsch, M. (2009). A new approach of accident causation analysis by seven steps acass *Proceedings of the International Technical Conference on the Enhanced Safety of Vehicles (ESV); Paper nr. 245.* Stuttgart, Germany. Retrieved from http://www-nrd.nhtsa.dot.gov/pdf/esv/esv21/09-0245.pdf
Page, Y., Fahrenkrog, F., Fiorentino, A., Gwehenberger, J., Helmer, T., Lindman, M., . . . Wimmer, P. (2015). A comprehensive and harmonized method for assessing the effectiveness of advanced driver assistance systems by virtual simulation: The P.E.A.R.S. initiative*. Proceedings of the 24th International Technical Conference on the Enhanced Safety of Vehicles (ESV); Paper nr. 15-0370.* Göteborg, Sweden.
Park, H., & Kim, K.-H. (2015). Adaptive multimodal in-vehicle information system for safe driving. *ETRI Journal. 37*(3), 626-636. doi:10.4218/etrij.15.0114.1104
Parker, D., Reason, J. T., Manstead, A. S. R., & Stradling, S. G. (1995). Driving errors, driving violations and accident involvement. *Ergonomics. 38*(5), 1036-1048. doi:10.1080/00140139508925170
Paulsson, R. (2005). *Deliverable 5.2: In-depth accident causation data study methodology development report* Chalmers University of Technology. Göteborg, Sweden. Retrieved from http://erso.swov.nl/safetynet/fixed/WP5/SN_Deliverable_5%202%20v3_051123_Final%20submission.pdf
Peng, Y., Boyle, L. N., & Hallmark, S. L. (2013). Driver's lane keeping ability with eyes off road: Insights from a naturalistic study. *Accident Analysis & Prevention. 50*(0), 628-634. doi:10.1016/j.aap.2012.06.013
Pettersson, I., & Karlsson, M. (2015). Setting the stage for autonomous cars: A pilot study of future autonomous driving experiences. *IET Intelligent Transport Systems. 9*(7), 694-701. doi:10.1049/iet-its.2014.0168
Piccinini, G. F. B., Engström, J., Bärgman, J., & Wang, X. (2016). Factors contributing to commercial vehicle rear-end conflicts in China: A study using on-board event data recorders. Submitted for publication
Prasad, P., & Chou, C. C. (2002). A review of mathematical occupant simulation models. In A. M. Nahum & J. W. Melvin (Eds.), *Accidental Injury: Biomechanics and Prevention* (pp. 121-186). New York, NY: Springer New York.
Reason, J., Manstead, A., Stradling, S., Baxter, J., & Campbell, K. (1990). Errors and violations on the roads - a real distinction. *Ergonomics. 33*(10-11), 1315-1332. doi:10.1080/00140139008925335
Regan, M., Lee, J., & Young, K. L. (Eds.). (2008). *Driver distraction: Theory, effects, and mitigation*. Boca Raton, FL: CRC Press.
Reimer, B., & Mehler, B. (2013). *The effects of a production level “voice-command” interface on driver behavior: Summary findings on reported workload, physiology, visual attention, and driving performance*. (2013-18A). MIT Agelab. Cambridge, MA. Retrieved from http://agelab.mit.edu/files/MIT_AgeLab_White_Paper_2013-18A_(Voice_Interfaces).pdf

Rizzi, M., Strandroth, J., & Tingvall, C. (2009). The effectiveness of antilock brake systems on motorcycles in reducing real-life crashes and injuries. *Traffic Injury Prevention. 10*(5), 479-487. doi:10.1080/15389580903149292

Rizzo, M., McGehee, D. V., Dawson, J. D., & Anderson, S. N. (2001). Simulated car crashes at intersections in drivers with alzheimer disease. *Alzheimer Disease & Associated Disorders. 15*(1), 10-20.

Rose, N., Carter, N., Pentecost, D. B., W. , & Voitel, T. (2013). *Using data from a drivecam event recorder to reconstruct a vehicle-to-vehicle impact*. (2013-01-0778). SAE Technical Paper. doi:10.4271/2013-01-0778.

Rothman, K. J. (2012). *Epidemiology: An introduction. 2nd edition*. New York: Oxford University Press, Inc.

Rydström, A., Ljung Aust, M., Broström, R., & Victor, T. (2015). Repeatability of the NHTSA visual-manual guidelines assessment procedure*. Proceedings of the 24th International Technical Conference on the Enhanced Safety of Vehicles (ESV). Paper nr. 15-0240.* Göteborg, Sweden. (pp. 1-5).

Räsänen, M., & Summala, H. (1998). Attention and expectation problems in bicycle–car collisions: an in-depth study. *Accident Analysis & Prevention. 30*(5), 657-666. doi:10.1016/S0001-4575(98)00007-4

SAE. (2015). Operational definitions of driving performance measures and statistics *J2944_201506*: Society of automotive engineers.

Sagberg, F., Selpi, Bianchi Piccinini, G. F., & Engström, J. (2015). A review of research on driving styles and road safety. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 57*(7), 1248-1275. doi:10.1177/0018720815591313

Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 48*(2), 362-380. doi:10.1518/001872006777724417

Salvucci, D. D., & Gray, R. (2004). A two-point visual control model of steering. *Perception. 33*(10), 1233-1248. doi:10.1068/p53433

Sandin, J. (2009a). An analysis of common patterns in aggregated causation charts from intersection crashes. *Accident Analysis & Prevention. 41*(3), 624-632. doi:10.1016/j.aap.2009.02.015

Sandin, J. (2009b). *Analysis of intersection crash statistics to define pre-crash test scenarios for detection sensors*. SAFER at Chalmers University of Technology. Göteborg, Sweden. Retrieved from http://www.chalmers.se/safer/EN/publications/pre-crash/2009/analysis-intersection

Sandin, J., & Ljung, M. (2007). Understanding the causation of single-vehicle crashes: a methodology for in-depth on-scene multidisciplinary case studies. *International Journal of Vehicle Safety. 2*(3), 316-333. doi:10.1504/IJVS.2007.015546

Sayer, J., Bogard, S., Buonarosa, M. L., LeBlanc, D., Funkhouser, D., Bao, S., . . . Winkler, C. (2011). *Integrated vehicle-based safety systems light-vehicle field operational test key findings report*. The University of Michigan Transportation Research Institute. Ann Arbor, Michgan. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2011/811416.pdf

Sayer, J., Bogard, S., Funkhouser, D., LeBlanc, D., Bao, S., Blankespoor, A., . . . Winkler, C. (2010). *Integrated vehicle-based safety systems heavy-truck field operational test key findings report*. U.S. Department of Transportation. Washington, DC. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2010/811362.pdf

Sayer, J., Devonshire, J., & Flannagan, C. (2005a). *The effects of secondary tasks on naturalistic driving performance*. (UMTRI-2005-29). University of Michigan Transportation Research Institute. Ann Arbor, MI. Retrieved from https://deepblue.lib.umich.edu/bitstream/handle/2027.42/57450/99225.pdf;sequence=1

Sayer, J., Devonshire, J., & Flannagan, C. (2007). Naturalistic driving performance during secondary tasks*. Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design.* Stevenson, WA. (pp. 224-230). Retrieved from http://drivingassessment.uiowa.edu/DA2007/PDF/039_SayerDevonshire.pdf

Sayer, J., LeBlanc, D., Bogard, S., Hagan, M., Sardar, H., Buonarosa, M., & Barnes, M. (2008). *Integrated vehicle-based safety systems preliminary field operational test plan* (DOT HS 811 010). The University of Michigan Transportation Research Institute. Ann Arbor, MI. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/2008/811010.pdf

Sayer, J., Mefford, M. L., & Huang, R. (2003). The effects of lead-vehicle size on driver following behavior: is ignorance truly bliss? *Proceedings of the 2nd International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.* Park City, Utah. (pp. 221-225). Retrieved from http://drivingassessment.uiowa.edu/DA2003/pdf/52_Sayerformat.pdf

Sayer, J. R., Mefford, M. L., Shirkey, K., & Lantz, J. (2005b). Driver distraction: Naturalistic observation of secondary behaviors with the use of driver assistance systems*. Proceedings of the 3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design.* Rockport, ME. Retrieved from http://drivingassessment.uiowa.edu/DA2005/PDF/37_JimSayerformat.pdf

Scanlon, J. M., Sherony, R., & Gabler, H. (2016). *Using event data recorders from real-world crashes to investigate the earliest detection opportunity for an intersection advanced driver assistance system*. (2016-01-1457). SAE Technical Paper. doi:10.4271/2016-01-1457.

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy. 2*, 419-436. doi:10.1207/S15327078IN0204_02

Seeck, A., Gail, J., Sferco, R., Otte, D., Hannawald, L., Zwipp, H., & Bakker, J. (2009). Development of the accident investigation and data handling methodologiy in the GIDAS project*. Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles (ESV); Paper nr 282.* Stuttgart, Germany.

Selpi, Borgen, S., Bärgman, J., Svanberg, E., Dozza, M., Nisslert, R., . . . Metz, B. (2012). *Deliverable D3.3 - Data management in euroFOT*. EuroFOT Consortium and the European Commisson. Aachen, Germany. Retrieved from http://www.eurofot-ip.eu/download/library/deliverables/eurofotsp320110831v10dld33_data_management_in_eurofot.pdf

Seppelt, B. (2009). *Supporting operator reliance on automation through continuous feedback.* (Doctoral dissertation), University of Iowa, Iowa City. Retrieved from http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1617&context=etd

Seppelt, B. D., & Victor, T. W. (2016). Potential solutions to human factors challenges in road vehicle automation. In G. Meyer & S. Beiker (Eds.), *Road Vehicle Automation 3*. Berlin: Springer.

SHRP2. (2012). *Researcher dictionary for video reduction data, v2.1*. Transportation Research Board. Washington,DC.

SmartDrive. (2016). SmartDrive main webpage. Retrieved from http://smartdrive.net/

Smith, K., & Källhammer, J.-E. (2010). Driver acceptance of false alarms to simulated encroachment. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 52* 446-476. doi:10.1177/0018720810372218

Smith, K., Thome, A., Blåberg, C., & Bärgman, J. (2009). An invariant may drive the decision to encroach at unsignalized intersections*. Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.* Big Sky, MT. (pp. 313-319). Retrieved from http://drivingassessment.uiowa.edu/DA2009/043_SmithThorne.pdf

Soccolich, S. A., & Hickman, J. S. (2014). Potential reduction in large truck and bus traffic fatalities and injuries using Lytx's drivecam program. Retrieved from http://hdl.handle.net/10919/64308

Strandroth, J. (2015a). *Identifying the potential of combined road safety interventions - a method to evaluate future effects of integrated road and vehicle safety technologies.* (Doctoral dissertation), Chalmers University of Technology, Göteborg, Sweden. (ISBN:978-91-7597-143-8).

Strandroth, J. (2015b). Validation of a method to evaluate future impact of road safety interventions, a comparison between fatal passenger car crashes in Sweden 2000 and 2010. *Accident analysis and prevention. 76*, 133-140. doi:10.1016/j.aap.2015.01.001

Strayer, D., & Cooper, J. (2015). Driven to distraction. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 57*(8), 1343-1347. doi:10.1177/0018720815610668

Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., & Biondi, F. (2015). Assessing cognitive distraction in the automobile. *Human Factors: The Journal of the Human Factors and Ergonomics Society. 57*(8), 1300-1324. doi:10.1177/0018720815575149

Stutts, J., Feaganes, J., Reinfurt, D., Rodgman, E., Hamlett, C., Gish, K., & Staplin, L. (2005). Driver's exposure to distractions in their natural driving environment. *Accident Analysis & Prevention. 37*(6), 1093-1101. doi:10.1016/j.aap.2005.06.007

Summala, H., Lappi, O., Pekkanen, J., Lehtonen, E., & Hietamäki, J. (2012). Skin conductance and cognitive estimates while driving through narrow gaps: implications to driver behavior theories *Traffic and Transport behaviour: interaction between theory and practice*. Groningen, Netherlands.

Svensson, M. Y., & Lövsund, P. (1992). A dummy for rear-end collisions: Development and validation of a new dummy neck. *Proceedings of the 1992 International IRCOBI conference on the biomechanics of impacts.* Verona, Italy. (pp. 299-310). Retrieved from http://www.ircobi.org/wordpress/downloads/irc1992/pdf_files/1992_24.pdf

Svensson, Å. (1992). *Vidareutveckling och validering av den svenska konflikttekniken (Development and validation of the Swedish traffic conflicts technique)*. Lund University, Department of Traffic Planning and Engineering. Lund, Sweden.

Svensson, Å. (1998). *A method for analysing the traffic process in a safety perspective.* (Doctoral dissertion), Lund University of Technology, Lund, Sweden. (ISSN:1404-272X, No. 166).

Tarko, A., Boyle, L. N., & Montella, A. (2013). Emerging research methods and their application to road safety. *Accident Analysis & Prevention. 61*, 1-2. doi:10.1016/j.aap.2013.07.006

Theeuwes, J., & Godthelp, H. (1995). Self-explaining roads. *Safety Science. 19*(2–3), 217-225. doi:10.1016/0925-7535(94)00022-U

Tijerina, L., Chovan, J. D., Pierowicz, J. A., & Hendricks, D. L. (1994). *Examination of unsignalized intersection, straight crossing path crashes and potential IVHS countermeasures*. (DOT HS 808 1 5 2). US Department of Transportation. Washington. Retrieved from http://ntl.bts.gov/lib/jpodocs/repts_te/7824.pdf

Tillmann, W. A., & Hobbs, G. E. (1949). The accident-prone automobile driver. A study of the psychiatric and social background. *The American Journal of Psychiatry. 106*(5), 321-331.

Tivesten, E., & Dozza, M. (2014a). Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving. *Transportation Research Part F: Traffic Psychology and Behaviour. 26, Part A*(0), 258-272. doi:10.1016/j.trf.2014.08.004

Tivesten, E., & Dozza, M. (2014b). Driving context influences drivers' decision to engage in visual-manual phone tasks: evidence from a naturalistic driving study. *Transportation*

Research Part F: Traffic Psychology and Behaviour. 26, 258-272. doi:10.1016/j.trf.2014.08.004

Tivesten, E., & Dozza, M. (2015). Driving context influences drivers' decision to engage in visual–manual phone tasks: Evidence from a naturalistic driving study. *Journal of Safety Research. 53*, 87-96. doi:10.1016/j.jsr.2015.03.010

Tivesten, E., Jonsson, S., Jakobsson, L., & Norin, H. (2012). Nonresponse analysis and adjustment in a mail survey on car accidents. *Accident Analysis & Prevention. 48*, 401-415. doi:10.1016/j.aap.2012.02.017

Toth, G. R., Radja, G. A., Thiriez, K. K., & Carra, J. S. (2003). Large truck crash causation study in the united states*. Proceedings of the International Technical Conference on the Enhanced Safety of Vehicles (ESV); Paper nr. 252.* Nagoya, Japan.

TRB. (2014). *SHRP2 safety facts sheet: New safety data available to tackle tough safety issues, improve highway safety*. US Transportation Research Board. Washington, DC. Retrieved from http://www.fhwa.dot.gov/goshrp2/Content/Documents/Factsheets/SHRP2_Safety_Fact_Sheet.pdf

Treat, J. R., Tumbas, N. S., McDonald, S. T., Shinar, D., Hyume, R. D., Mayer, R. E., . . . Castellan, N. J. (1977). *Tri-level study of the causes of traffic accidents: Final report. Volume i: Causal factor tabulations and assessments;* . (DOT HS 805 085). National Highway Traffic Safety Administration ad the US Department of Transportation. Washington. Retrieved from http://ntl.bts.gov/lib/25000/25500/25515/DOT-HS-805-085.pdf

Uchida, N., Kawakoshi, M., Tagawa, T., & Mochida, T. (2010). An investigation of factors contributing to major crash types in Japan based on naturalistic driving data. *IATSS Research. 34*(1), 22-30. doi:10.1016/j.iatssr.2010.07.002

Utesch, F., Bärgman, J., Engström, J., Petzoldt, T., Winkelbauer, M., Christoph, M., . . . Saintpierre, G. (2014). *Preliminary analysis plan - UDRIVE D11.2. Manuscript submitted to European Commission for publication*. UDrive Consortium.

Vaa, T. (2007). Drivers' information processing, decision-making and the role of emotions: Predictions of the risk monitor model. In C. P. Cacciabue (Ed.), *Modelling driver behavior in automotive environments*. London: Springer.

Vadeby, A., Forsman, Å., Kecklund, G., Åkerstedt, T., Sandberg, D., & Anund, A. (2010). Sleepiness and prediction of driver impairment in simulator studies using a Cox proportional hazard approach. *Accident Analysis & Prevention. 42*(3), 835-841. doi:10.1016/j.aap.2009.09.023

Valldorf, J., & Gessner, W. (Eds.). (2003). *Advanced Microsystems for Automotive Applications 2003*. Berlin: Springer.

Van Elslande, P., & Fouques, K. (2007). *Analyzing 'human functional failures' in road accidents* TRACE_concortium. Retrieved from http://www.trace-project.org/publication/archives/trace-wp5-d5-1.pdf

Vaux, L. M., Ni, R., Rizzo, M., Uc, E. Y., & Andersen, G. J. (2010). Detection of imminent collisions by drivers with Alzheimer's disease and Parkinson's disease: A preliminary study. *Accident Analysis & Prevention. 42*(3), 852-858. doi:10.1016/j.aap.2009.07.009

Viano, D., & Ridella, S. (1996). Crash causation: A case study of fatal accident circumstances and configurations. *SAE Transactions. 105*, 458-469 doi:10.4271/960458

Victor, T., Bärgman, J., Hjälmdahl, M., Kircher, K., Svanberg, E., Hurtig, S., . . . Moeschlin, F. (2010). *Sweden-Michigan naturalistic field operational test (SeMiFOT) - phase 1: final report*. SAFER. Göteborg, Sweden. Retrieved from https://www.chalmers.se/safer/EN/publications/traffic-safety-analysis/2010/sweden-michigan8650

Victor, T., Dozza, M., Bärgman, J., Boda, C.-N., Engström, J., Flannagan, C., . . . Markkula, G. (2015). *Analysis of naturalistic driving study data: SAFER glances, driver inattention, and crash risk*. (SHRP 2 SO8A). Transportation Research Board of the

National Academy of Sciences. Washington, D.C. ISBN: 978-0-309-27423-4. Retrieved from http://onlinepubs.trb.org/onlinepubs/shrp2/SHRP2_S2-S08A-RW-1.pdf

Viti, F., Hoogendoorn, S. P., Alkim, T. P., & Bootsma, G. (2008, 4-6 June 2008). Driving behavior interaction with ACC: results from a Field Operational Test in the Netherlands. *Proceedings of the Intelligent Vehicles Symposium, 2008 IEEE.* (pp. 745-750). doi:10.1109/IVS.2008.4621199.

Warner, H. W., & Sandin, J. (2010). The intercoder agreement when using the Driving Reliability and Error Analysis Method in road traffic accident investigations. *Safety Science. 48*(5), 527-536. doi:10.1016/j.ssci.2009.12.022

White, C. B., & Caird, J. K. (2010). The blind date: The effects of change blindness, passenger conversation and gender on looked-but-failed-to-see (LBFTS) errors. *Accident Analysis & Prevention. 42*(6), 1822-1830. doi:10.1016/j.aap.2010.05.003

WHO. (2013). *Global status report on road safety 2013: supporting a decade of action*. World Health Organization: Department of violence and injury prevention and disability. Geneva, Switzerland. ISBN:978 92 4 156456 4.

Wickens, C., Hollands, J. G., Banbury, S., & Parasuaman, R. (2016). *Engineering psychology and human performance (4rd ed.)*. New York, NY: Routledge.

Wicksteed, P. H., & Cornford, F. M. (2 vols., 1929). *The Physics (Aristotle).* Cambridge, Massachusetts: Harvard University Press.

Wilde, G. J. S. (1982). The theory of risk homeostasis: Implications for safety and health. *Risk Analysis. 2*(4), 209-226. doi:10.1111/j.1539-6924.1982.tb01384.x

Woodrooffe, J., Blower, D., Bao, S., Bogard, S., Flannagan, C., Green, P. E., & LeBlanc, D. (2012). *Performance characterization and safety effectiveness estimates of forward collision avoidance and mitigation systems for medium/heavy commercial vehicles*. (UMTRI-2011-36). The University of Michigan Transportation Research Institute for US Department of Transportation.

Wortelen, B., Baumann, M., & Lüdtke, A. (2013). Dynamic simulation and prediction of drivers' attention distribution. *Transportation Research Part F: Traffic Psychology and Behaviour. 21*(0), 278-294. doi:10.1016/j.trf.2013.09.019

Wu, K.-F., & Jovanis, P. P. (2012). Crashes and crash-surrogate events: Exploratory modeling with naturalistic driving data. *Accident Analysis & Prevention. 45*(0), 507-516. doi:10.1016/j.aap.2011.09.002

Yamamoto, T., Hashiji, J., & Shankar, V. N. (2008). Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. *Accident Analysis & Prevention. 40*(4), 1320-1329. doi:10.1016/j.aap.2007.10.016

Yang, K. H., Hu, J., White, N., King, A., Chou, C., & Prasad, P. (2006). Development of numerical models for injury biomechanics research: A Review of 50 years of publications in the Stapp Car Crash conference. *Stapp Car Crash Journal. 50*, 429-490. PCMID:17311173

Yurtsever, E., Miyajima, C., Selpi, & Takeda, K. (2015). Driving Signature Extraction. *Proceedings of the 3rd International Symposium on Future Active Safety Technology Towards Zero Traffic Accidents* Göteborg, Sweden. (pp. 183-187).

Östh, J. (2014). *Muscle responses of car occupants: Numerical modeling and volunteer experiments under pre-crash braking conditions.* (Doctoral dissertation), Chalmers University of Technology, Göteborg, Sweden. (ISBN:978-91-7385-987-5).

Östlund, J., Nilsson, L., Carsten, O., Merat, N., Jamson, H., Jamson, S., . . . Brouwer, R. (2004). *HASTE Deliverable 2: HMI and Safety-Related Driver Performance*. (HASTE-D2). HASTE_consortium.