THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Driving Behavior and Safety Targets: A Naturalistic Perspective

PIERLUIGI OLLEJA

Department of Mechanics and Maritime Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2026

Driving Behavior and Safety Targets: A Naturalistic Perspective

PIERLUIGI OLLEJA

ISBN 978-91-8103-380-9

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5837
ISSN 0346-718X
https://doi.org/10.63959/chalmers.dt/5837

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

Cover image:
Two vehicles ride over the lines of a simplified graph usually found in counterfactual simulation.

Printed by Chalmers digitaltryck
Gothenburg, Sweden 2026

To my family

Abstract

Road crashes are a major cause of deaths and serious injuries worldwide. New technologies can reduce their number and severity by supporting drivers with advanced driver assistance systems (ADASs), and by taking over the entire driving task—at least under certain conditions—with automated driving systems (ADSs). Virtual simulation is one of the methods for predicting how safe these systems will be once they are released on public roads. However, ensuring that this method provides meaningful and representative results remains challenging. Safety targets are required to ensure that ADAS and ADS assessments are effective, relevant, and fair and that the systems have a positive impact on safety. This thesis focuses on the foundations for formulating and assessing these safety targets, with the aim of supporting the development of ADSs and ADASs, and ultimately improving traffic safety.

The work addresses four main aspects of safety targets. First, the thesis investigates the impact of data selection on the outcomes of virtual safety assessment. The findings indicate that crashes artificially generated from these data can differ substantially from real-world crashes, leading to lower severity outcomes, reduced criticality, and inaccurate benefit estimations. Second, the thesis evaluates the safety performance of the reference models described in UN Regulation No. 157 to determine whether they represent adequate safety targets for ADSs. A comparison of the models' responses to those of real drivers in safety-critical scenarios reveals that the models do not perform like the competent and careful drivers they are intended to represent. Third, the thesis analyzes lane-changing behavior and its relation to the surrounding driving context. The results describe characteristics of lane-changing behavior that can be used in modeling, including a modified definition of lane-change initiation that incorporates a lateral speed threshold (in addition to the lateral position threshold used to define lane changes in current reference driver models). Finally, the thesis investigates the link between off-road glance behavior and crash risk for car-following scenarios. In line with previous literature, our results suggest that drivers adapt only modestly to time gap-related crash risk, yet they reduce both the frequency and duration of off-road glances as time to collision gets shorter.

The findings highlight issues in current regulatory behavior model-based safety targets and the challenges that current reference models face, both in their formulation and in the data and methods used to assess safety target validity. Moreover, the findings also suggest that using observed behavior as the sole basis for safety targets for ADASs is problematic, although including components such as urgency and glance behavior may improve their performance and relevance. Overall, the results highlight the importance of ensuring relevant and valid virtual safety assessments through a well-considered choice of data sources and the robust and accurate representation of safety targets through driver models.

Keywords: safety targets, virtual safety assessment, reference driver model, counterfactual simulations, glance behavior

Acknowledgments

First and foremost, I would like to express my gratitude to my supervisor Jonas Bärgman, for his continuous guidance and encouragement. I also wish to thank my co-supervisors Gustav Markkula (first part of the work), and Malin Svärd (last part of the work) for their support, and Mikael Ljung Aust as a project lead for the QUADRIS project. Thank you to my examiner Marco Dozza for many insightful and thought-provoking conversations. I am also thankful to Nils Lübbe, whom I met before beginning my PhD journey and who supervised my early work.

I am also grateful to the entire Division of Vehicle Safety, where everyone has always been helpful, kind and supportive. I will miss our lunch breaks and fikas. And thank you to the colleagues at Volvo Cars Safety Centre, who warmly welcomed me during the last part of my PhD. I also want to thank Carol Flannagan for her valuable insights into my research, and Kristina Mayberry for the language reviews.

Thank you to my family, who never feel as distant as geography might suggest. And to my friends Christopher, Amr, Alessio, Jonas, Erik, Eliott, Valeria, Lena, Maria, Chiara, Camilla, Morgan, and Noemi that witnessed my journey towards this achievement, and that supported me from near and from far away. Thank you to Ron, Thomas, and Alberto: colleagues who became friends, and who have always encouraged and supported me. A big thank you to Chiara, who patiently listened to my endless chats at work, and to Rahul, who was always there with invaluable help and who made work so fun. Finally, a special thank you to Clarissa. Your support knows no boundaries, I am lucky to have you.

Funding acknowledgments

This work was primarily funded by the FFI projects QUADRIS (grant number 2020-05156) and QUADARUM (grant number 2025-00834), but the initial part of the work (much of Paper I) was funded by Autoliv Research.

Declaration on the use of generative AI in the writing process

During the preparation of this work, the author used a generative large language model (LLM; Microsoft Copilot) to improve grammar and clarity in selected parts of the manuscript. The tool was applied only to individual sentences and paragraphs. After using the tool, the author reviewed and edited the text as needed and takes full responsibility for the content of the thesis. No “content” was generated using LLMs.

List of publications

This thesis is based on the following publications:

Paper I

Olleja, P., Bärgman, J., & Lubbe, N. (2022). Can non-crash naturalistic driving data be an alternative to crash data for use in virtual assessment of the safety performance of automated emergency braking systems? *Journal of Safety Research*, *83*, 139–151. https://doi.org/10.1016/j.jsr.2022.08.011

Author's contribution: Building on the provided concept, defined the research direction. Managed the setup and execution of the simulations and performed data analysis. Drafted the initial version of the paper and worked with the co-authors to finalize the manuscript.

Paper II

Olleja, P., Markkula, G., & Bärgman, J. (2025). Validation of human benchmark models for Automated Driving System approval: How competent and careful are they really? *Accident Analysis & Prevention, 213,* Article 107922. https://doi.org/10.1016/j.aap.2025.107922

Author's contribution: Part of establishing the conceptual framework and research plan. Curated the data and was responsible for the simulation setup and execution. Performed the analysis. Wrote the first draft of the paper and collaborated with the co-authors to finalize the manuscript.

Paper III

Jokhio, S., Olleja, P., Bärgman, J., Yan, F., & Baumann, M. (2024). Analysis of time-to-lane-change-initiation using realistic driving data. *IEEE Transactions on Intelligent Transportation Systems*, *25*(5), 4620–4633. https://doi.org/10.1109/tits.2023.3329690

Author's contribution: Collaborated with the main author to curate the data. Conducted the lane-change identification analysis, while supporting other analytical tasks. Contributed to revising the manuscript (the first draft was prepared by the main author).

Paper IV

Olleja, P., Svärd, M., Zhao, M., & Bärgman, J. (*manuscript for journal submission*). Analysis of the safety impact of off-road glances in car-following driving scenarios.

Author's contribution: Established the conceptual framework and research plan. Curated the dataset and managed the simulation setup and execution (excluding the glance risk calculation). Performed the analysis. Drafted the initial version of the paper and collaborated with co-authors to finalize the manuscript.

Paper V

Svärd, M., Olleja, P., & Bärgman, J. (*manuscript for journal submission*). Drivers' glance-behavior adaptation in response to looming during car-following.

Author's contribution: Was part of establishing the conceptual framework and the research plan. Collaborated with the main author to curate the data and supported analysis. Contributed to revising the manuscript (initial draft prepared by the main author).

List of additional work

Conference presentations (with extended abstracts) and publications not included in this thesis:

Jokhio, S., Olleja, P., Bärgman, J., Yan, F., & Baumann, M. (2024). Exploring turn signal usage patterns in lane changes: A Bayesian hierarchical modelling analysis of realistic driving data. *IET Intelligent Transport Systems*, *18*(2), 393–408. https://doi.org/10.1049/itr2.12457

Author's contribution: Curated the data. Was part of research direction discussions. Contributed to revising the manuscript (initial draft prepared by the main author).

Olleja, P., Bärgman, J., & Markkula, G. (2022, October 19–20). *Quantification of driver's side-glance frequency and duration in straight highway driving* [Conference presentation]. The 8th International Conference on Driver Distraction and Inattention, Gothenburg, Sweden. https://ddi2022.org/wp-content/uploads/2022/10/3.1_Pierluigi_Olleja_DDI.pdf

Author's contribution: Together with the supervisors conceptualized the work. Curated the data. Performed the simulations and analysis. Wrote the first draft of the extended abstract and the presentation. Contributed to the finalization of the abstract and the presentation. Presented the work.

Olleja, P., Rasch, A., & Bärgman, J. (2023, October 26–27). *A reference driver model for overtaking a cyclist* [Conference presentation]. International Cooperation on Theories and Concepts in Traffic safety (ICTCT), Catania, Italy. https://www.ictct.net/wp-content/uploads/35-Catania-2023/69-Olleja.pdf

Author's contribution: Together with the supervisors, conceptualized the work. Curated the data. Performed the simulations and analysis. Wrote the first draft of the extended abstract and the presentation. Contributed to the finalization of the abstract and the presentation. Presented the work.

Bärgman, J., Jokhio, S., Svärd, M., Fries, A., Denk, F., Paliotto, A., de Oliveira, L. F. A., Beckmann, J. N. B., Adjenughwure, K., Davidse, R. J., de Zwart, R., Fahrenkrog, F., Hammouda, M., & Olleja, P. (2025). *On human behavior models for virtual safety impact assessment: A scoping review*. Zenodo. https://doi.org/10.5281/zenodo.17253866

Author's contribution: Contributed to the literature review on reference behavior models and related topics. Participated in finalization of the manuscript.

List of abbreviations

ACC	–	Adaptive Cruise Control
ADAS	–	Advanced Driver Assistance System
ADS	–	Automated Driving System
AEB	–	Automated Emergency Braking
ALKS	–	Automated Lane-Keeping System
BTN	–	Brake Threat Number
CAN	–	Controller Area Network
CBM	–	Computational Human Driver Behavior Models
CCDM	–	Competent and Careful Driver Model
CFS	–	Critical Fuzzy Surrogate
CIDAS	–	China In-Depth Accident Study
CISS	–	Crash Investigation Sampling System
CZB	–	Comfort Zone Boundary
DMS	–	Driver Monitoring System
EVT	–	Extreme Value Theory
FCW	–	Forward Collision Warning
FSM	–	Fuzzy Safety Model
GIDAS	–	German In-Depth Accident Study
IGLAD	–	International Harmonized In-Depth Accident Data
IIS	–	Inattention Information System
LLM	–	Large Language Model
MCR	–	Model Estimate Crash Risk
NDS	–	Naturalistic Driving Study
NIEON	–	Non-Impaired Road User with their Eyes on the Conflict
ODD	–	Operational Design Domain
PCM	–	Pre-Crash Matrix
PDO	–	Property Damage Only
PET	–	Post-Encroachment Time
PFS	–	Predictive Fuzzy Surrogate
POV	–	Principal Other Vehicle
PRB	–	Positive Risk Balance
R157	–	UN Regulation No. 157
RAIDS	–	Road Accident In Depth Studies
RASSI	–	Road Accident Sampling System India
RSS	–	Responsibility-Sensitive Safety
SHRP2	–	Strategic Highway Research Program 2

SOTIF	–	Safety of the Intended Functionality
THW	–	Time Headway
TTC	–	Time to Collision
TTLCI	–	Time to Lane Change Initiation
UNECE	–	United Nations Economic Commission for Europe
VCC	–	Volvo Car Corporation
VRU	–	Vulnerable Road User

Contents

Part I

Introductory Chapters

1 Introduction

Traffic safety plays a major role in developing the mobility of the future. Road crashes are among the most common causes of death; it is estimated that in 2021 1.19 million people died in traffic crashes worldwide (WHO, 2023). The risk of road users incurring serious injuries or death in a crash drives the development of safer mobility solutions; a variety of stakeholders (e.g., governments, academia, institutes, and private companies) are steadily working on improving vehicles and infrastructure.

A central philosophy in modern traffic safety is Vision Zero. This strategy originated in Sweden and has since been adopted in several other countries. It is based on the ethical principle that no loss of life in the transport system is acceptable. It shifts the responsibility for safety from individual road users to system designers, emphasizing that while human error is inevitable, fatalities and serious injuries are not. This philosophy is put into practice with the Safe System Approach, a framework that recognizes human error as inevitable and aims to build a road system that is forgiving. It focuses on five key pillars: safe roads and roadsides, safe speeds, safe vehicles, safe road users, and effective post-crash response. The goal is to ensure that when crashes do occur, they do not result in death or serious injury.

While the Safe System Approach tolerates human error, it still expects drivers to follow traffic rules and operate their vehicles safely, raising the question: How do we define "safe driving”? Because of the high level of complexity of the driving task, safety is achieved not only by observing traffic rules, but also through the human ability to predict and react to safety-relevant driving scenarios.

Safety-relevant driving scenarios, in which the kinematics of the vehicles involved are such that there is a risk of crashing, cover a wide range of safety criticalities, from low to high risk (imminent crashes). It is common practice to divide drivers’ interventions into two categories: crash avoidance and conflict avoidance (Scanlon et al., 2022). Crash avoidance is the ability of the ego vehicle to perform an immediate action (e.g., braking or steering) to avoid an imminent crash when the vehicle is on a collision path with another vehicle, obstacle, or vulnerable road user (VRU). Conflict avoidance is the combination of actions intended to avoid situations that, without timely intervention, could increase the probability of a crash (Favaro et al., 2026).

Humans, however, cannot be attentive to everything in traffic at all times—they are prone to distractions and mistakes while driving, and may consequently fail to perceive potential hazards. It is for this reason that in the last few decades modern vehicle technologies have been developed with the aim of improving the crash and conflict avoidance capabilities of drivers. Specifically, advanced driver assistance systems (ADASs) and automated driving systems (ADSs) are designed to help mitigate crashes or avoid them altogether, by supporting or replacing the driver, respectively (at least under certain conditions). Automated emergency braking (AEB) is an example of an ADAS aimed at avoiding crashes that acts when safety criticality is high and a crash is imminent. As their development continues, ADASs are increasingly able to handle driving tasks even in low-criticality scenarios (Antony & Whenish, 2021; Nidamanuri et al., 2021). Unlike ADASs, ADSs take responsibility for the whole driving task at all levels of safety criticality, at least within a specific operational design domain (ODD; ISO, 2022a). ADSs consequently include both conflict and crash avoidance systems.

The development of ADASs and ADSs presents considerable challenges, including compliance with standards and regulations to ensure their safe operation (see Section 1.1) and the assessment of their safety benefits (see Section 1.2). A key aspect of the latter is establishing a fair and meaningful reference for the level of safety the system provides, answering the question, “What is ‘safe enough’?” This thesis is based on the premise that human behavior can be used as a reference for driving that is safe enough (see Section 1.3). This premise is based on the fact that ADASs and ADSs need to be compared and validated against the entity which they support/replace: the human driver. In this case, the validation for a given safety-critical scenario consists of showing that the system’s behavior leads to a safer outcome than the behavior expected from a human. These performance comparisons are typically based on outcome measures (e.g., crash or injury risk), so assessments can be made even when humans and systems differ fundamentally (as in, for example, in perception or prediction capability). It is a generally accepted fact that all models (including human-behavior, system, and vehicle and environmental models) inevitably contain inaccuracies. It should be noted that comparing systems to humans is just one of the components needed to assess whether a system is safe or not. Other components, such as assessing the compliance with functional safety standards and ensuring the acceptance of the systems by other road users, are just as important.

1.1 Standards and regulations

The safe operation of ADASs and ADSs is regulated by well-defined procedures for testing hardware and implementing software. This section outlines some of the most relevant standards and regulations. ISO 26262 is the main international standard defining functional safety for electrical and electronic systems installed in vehicles (ISO, 2018). It covers, for example, random system failures, and defines “functional safety” as the “absence of unreasonable risk due to hazards caused by malfunctioning behaviour” (p. 14; ISO, 2018). The standard ISO 21448 defines a complementary concept, the safety of the intended functionality (SOTIF), which is “the absence of unreasonable risk due to a hazard caused by functional insufficiencies” (ISO, 2022a). The standard ISO 34502 leverages on the SOTIF concept to define a scenario-based safety evaluation process for ADSs (ISO, 2022b). This standard codifies the infinite possible traffic interactions that ADSs can encounter with a finite number of scenario categories, and divides the driving task along three aspects: perception, judgement, and control. The categories can be used to obtain quantitative results from ADS testing: an example is described in the UN Regulation No. 157 (R157; UNECE: United Nations Economic Commission for Europe, 2023). This regulation contains provisions for the approval of automated lane-keeping systems (ALKSs) using scenario-based testing.

The standards and regulations mentioned are only a subset of the requirements that ADASs and ADSs must satisfy to be released for public use. In addition, companies have their own internal safety assessment processes. Complying with all these requirements necessitates a range of assessment tools, including virtual simulations, introduced in the next section.

1.2 Virtual safety assessment methods

As mentioned, the performance of systems such as ADASs and ADSs must be tested to ensure their safe operation, both as part of standards compliance testing and as part of vehicle manufacturers internal processes. However, it typically takes years for these systems to reach enough market penetration for long enough to enable valid retrospective safety evaluations in real vehicles on the road (Gulino et al., 2022; Smit et al., 2019; Wimmer et al., 2019), which make it possible to compare, for example, the number of crashes or average injury risk with and without the system. Consequently, the assessment methods used in their development and regulation (e.g., defining when a vehicle/system is safe enough to release on the roads) must be carried out prospectively, prior to the system's release on the market. Since these methods, often performed in virtual environments, are used to predict the impact of a system if it were available in traffic (Alvarez et al., 2017), they play a fundamental role in the development of ADASs and ADSs. Moreover, prospective methods are also becoming an important component in ADS approval processes (UNECE, 2023), and the inclusion of pre-crash virtual simulations in consumer rating programs of ADASs (Euro NCAP, 2023) is currently being considered.

Reaching an exhaustive assessment of the safety of a system based on virtual simulations is, however, not straightforward (Wimmer et al., 2019). The safety impact of a system is quantified by comparing the scenario with and without the system (treatment and baseline, respectively). The terms "baseline" and "treatment" were borrowed from the medical field; systems applied to reduce or avoid crashes are considered analogous to medicines treating a disease. The baseline represents, at least in theory, the real-world traffic scenarios considered important for safety (and that the systems aim to address) as accurately as possible. As such, the choice of baseline must be based on the evaluation scope: what is to be assessed and what the purpose is. Fahrenkrog et al. (2025) define the evaluation scope as a combination of evaluation questions which should "take current scientific knowledge and state-of-the-art in road traffic safety" into account; the scope should also "point out the gap that is addressed by the evaluation" (p. 30; Fahrenkrog et al., 2025).

Wimmer et al. (2023) and Fahrenkrog et al. (2025) define three fundamentally different approaches to baseline generation for prospective safety assessments, described in Fig. 1. Approaches A and B start with individual original scenarios, while approach C uses distributions of parameters describing the scenarios. The processing for approach A consists of digitally representing the unmodified original scenarios. Approach B is similar to A, but the processing can include parameter variations or the inclusion of behavior models for part of the scenario, resulting in more scenarios. For approach C, the sub-approach C1 aggregates parameter sets to generate a few representative scenarios; C2 typically uses sampling techniques and behavior models of road users to obtain large datasets of interactions.

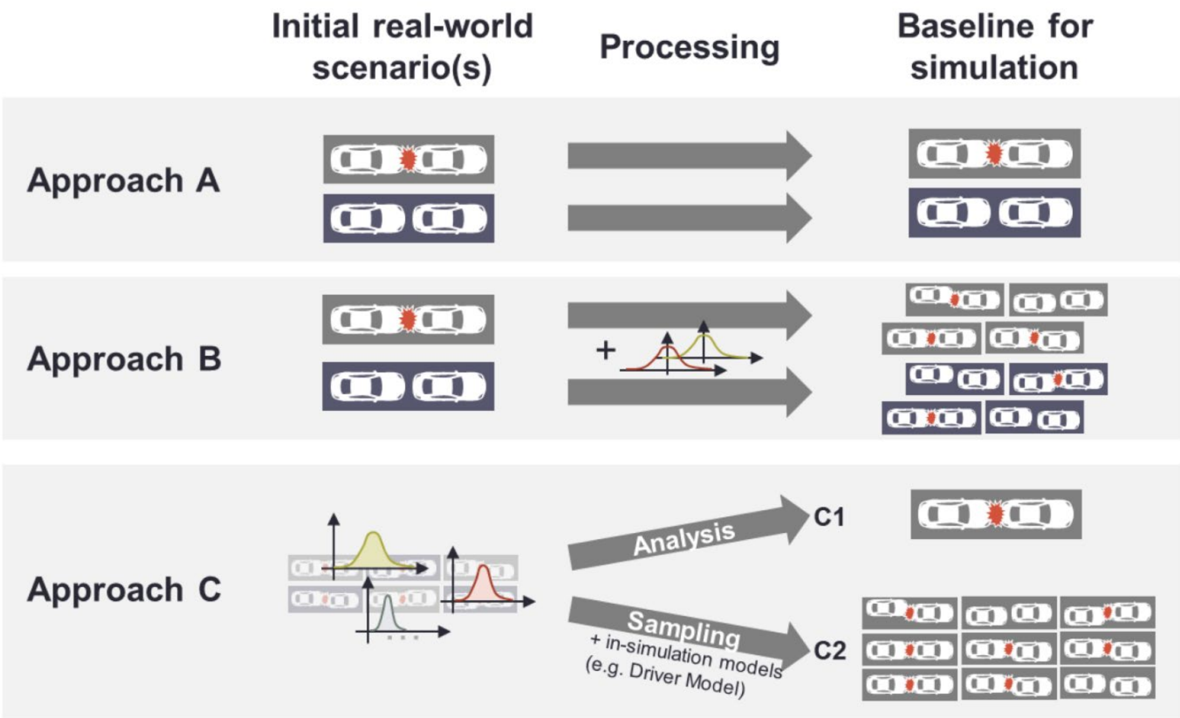

Fig. 1 Approaches for creating baseline data for virtual safety assessments from Wimmer et al. (2023; reproduced with permission from the authors). The first approach (A) uses unmodified real-world events as the baseline. The second approach (B) consists of increasing the number of available scenarios by modifying individual real-world scenarios, typically through simulation. The third approach (C) has two sub-approaches: in C1, statistical data are aggregated to a few representative scenarios, and C2 uses stochastic models (with or without computational behavior models in simulation) to create more scenarios for the baseline and the treatment, increasing the coverage.

As these approaches illustrate, baselines are all derived from the real world in some way, either directly via statistical models or indirectly via behavior models. However, the data they use as input can differ greatly. Complex and often expensive data collection is typically required to capture relevant interactions (James et al., 2015; Liers, 2018; Pelella et al., 2023). For example, in-depth crash databases are commonly used (Chen & Dai, 2018; Cuerden & McCarthy, 2016; Otte et al., 2003; Rameshkrishnan et al., 2013; Zhang et al., 2019). Their relatively high-fidelity data can provide baselines when assessing crash avoidance systems (Erbsmehl, 2009). Some of these databases capture detailed characteristics of real-world crashes, including pre-crash time-series data of individual events (Schubert et al., 2013). They can include trajectories of the agents (e.g., vehicles and VRUs) from (typically) five seconds prior to the crash up to the moment of the crash, as well as a relatively detailed description of the event itself (Schubert et al., 2013). However, these databases can only describe a relatively small part of the complexity of the driving task; moreover, the time-series reconstructions typically include several assumptions which limit how they can be used (James et al., 2015; Kiuchi, 2020; Liers, 2018).

Naturalistic driving studies (NDSs), which collect data during drivers' everyday driving, are another data source. They can complement in-depth crash databases by providing much more complete coverage of a larger range of safety criticalities (Bärgman, 2016; Hankey et al., 2016), albeit at the expense of capturing few (if any) crashes, and very few (if any) high-severity crashes (Ehsani et al., 2021). As crashes are rare, near-crashes are commonly used as surrogates for crashes (Guo et al., 2010; Hydén, 1987; Laureshyn & Várhelyi, 2018; Victor et al., 2014; Wu & Jovanis, 2012). A near-crash is a driving instance in which a crash was imminent but the involved road users managed to avoid it. Near-crashes have been used to generate crashes by modifying the real-world driving kinematics which successfully avoided the crash (Bärgman et al., 2015; Victor et al., 2014). However, this method is not always straightforward, and this area is therefore an active research field (see, e.g., Fahrenkrog et al., 2025; Wimmer et al., 2023). Actually, very few studies have been able to demonstrate that they can produce baseline crashes that are representative of the real world (Bärgman, Jokhio, et al., 2025).

Counterfactual simulations represent one safety assessment method that uses modifications of safety-critical scenarios from real-world driving (Davis et al., 2011). These simulations typically use a set of crashes (baseline scenarios) in which the virtual representation of the system under assessment is applied, generating treatment scenarios. That is, ADASs and ADSs are used in the treatment scenarios to complement or replace the road users in the baseline scenarios, to answer the question: What if the vehicle(s) had been equipped with the system? If the simulated outcome is better than what happened in the real world (the baseline), the system is considered to improve safety for that specific driving scenario. Of course, the terms "outcome" and "better" do not have global definitions, and are often defined differently for different safety assessments. Typical analyses of counterfactual simulations include assessing the outcomes in terms of crash avoidance and mitigation. If the application of a system turns a crash into a non-crash, the benefit of the system with respect to avoiding crashes is evident. If the system only mitigates the crash, the safety benefits of the system must be quantified by some other metrics. For example, delta-v can be obtained from the impact speed (Kullgren et al., 2003), and the risk of injuries can be predicted (Gennarelli & Wodzin, 2006; Kullgren, 2008): baseline and treatment outcomes can be compared with either metric.

1.3 The role of driver models

Some virtual safety assessment methods also need computational human driver behavior models (CBMs; ISO, 2021), defined here as mathematical and/or statistical representations of driver behavior. These models are used in the assessments to represent the behavior of drivers in traffic. They may include everyday driving (e.g., car-following), models of crash causation (e.g., off-road glance behavior), or driver responses to critical events (e.g., hard braking in response to a lead vehicle braking hard).

Over the last few decades, many CBMs have been developed (Boda et al., 2020; Engström et al., 2018; Kiefer et al., 2005; Lee & Jang, 2019; Maddox & Kiefer, 2012; Markkula, 2014; Markkula et al., 2016; Svärd, Markkula, et al., 2021; Svärd et al., 2017; Xue et al., 2018; Zgonnikov et al., 2024). See Bärgman, Jokhio, et al. (2025) for a recent literature review of CBMs used in safety impact assessments. CBMs may be intended either to capture driver variability (i.e., represent some group of drivers; Svärd, Bärgman, et al., 2021; Webb et al., 2020), or to represent the behavior characteristics of a single driver (UNECE, 2023). In the former, the variability is often represented by distributions of model parameters (Markkula et al., 2016; Rasch & Dozza, 2022), while the latter has fixed parameter values.

A combination of CBMs can be used to virtually generate crashes, although care must be taken to make sure that the generated crashes represent what they are supposed to. That is, they should align with the evaluation scope (Fahrenkrog et al., 2025) and they should be validated (Bärgman et al., 2024; Wu, Sander, et al., 2025). For treatment simulations, there are models that describe how drivers respond to a system, such as forward collision warning (FCW). Further, in counterfactual simulations part of the trajectory can be replaced by CBMs to generate alternatives to what actually happened. Note that driver models can be used in two of Wimmer et al.'s baseline approaches (B and C2; see Fig. 1) as part of the baseline generation process (Wimmer et al., 2023).

1.3.1 Safe human driving behavior

Since safety assessment methods are used to evaluate the safety impact of systems, it is important that any CBM used accurately represent the driver behaviors that are relevant for safety. This section describes some of the principal components of safe driving. While the overall driving task is complex, it can be broken down into components, which can be modeled individually or in combination with each other.

A primary component of safe driving is the ability to react to threats quickly to avoid, or at least mitigate, an imminent crash. Two important aspects of this reactive behavior are the type of reaction (e.g., braking and steering) and its timing (Green, 2000). For example, the reactions are affected by how surprising and urgent the event is (Markkula et al., 2016). The implications of this behavior for safety are clear: a quick and appropriate reaction that slows the vehicle or changes its trajectory can reduce or eliminate the risk of crashing and seriously injuring anyone (inside or outside the car). This reactive capability, however, is dependent on the driver's level of attention and their readiness to notice unpredictable events, as well as by how surprising and urgent the event is. In other words, a human driver's reaction to a threat can only happen if the driver is aware of the threat. The direction of the driver's glances is often used as an indicator of this awareness. This approach, however, has limitations, as it only captures the physical act of "looking at" a threat, not the cognitive process of actually perceiving it (Ahlström et al., 2021; Mack, 2003; Simons, 2000). Nonetheless, glance behavior has a direct impact on the timing of the reaction, as the threat cannot be assessed until drivers direct their gaze towards it (with the caveat that peripheral vision can, to some extent, support threat assessment; Svärd, Bärgman, et al., 2021). Deciding how to react can be based on considerations like the urgency of the situation (Markkula et al., 2016). Bärgman et al. (2015) first proposed the use of driver models and counterfactual simulations to estimate crash and injury risks associated with specific glance behaviors; both have subsequently been used by others. For example, Lee et al. (2018) described radio tuning as a normative, societally accepted distracting task.

In addition to the type and timing of the reaction, another component of driving behavior that impacts safety is the position of the vehicle with respect to surrounding traffic, for example during car-following (following a leading vehicle in the same lane) or when performing a lane change. While the safety implications of proactive conflict-avoidance-related behaviors like maintaining reasonable distances or using turn signals appropriately are less obvious than immediate reactions to threats (crash avoidance), they are still highly relevant for safe driving. In car-following scenarios, for example, the longitudinal distance to the leading vehicle impacts the time available to react to an unexpected event. Maintaining a safe distance can reduce the likelihood of a conflict, and consequently of a crash. In fact, a minimum safety distance is mandated by law in several countries (Conference of European Directors of Roads, 2009). Similarly, during lane changes, conflicts can be avoided with the proper use of turn signals, how lane changes are executed, and an awareness of the position of vehicles in the target lane (the lane the driver is moving into). Turn signal use differs substantially between drivers and countries; in this case also, there are legal requirements in several countries (Bundesministerium der Justiz und für Verbraucherschutz, 2024, § 7 para. 5; California Department of Motor Vehicles, 2025, Section 5).

1.3.2 Definitions of safety targets

If ADASs and ADSs should be compared to or based on safe human driving behavior, the concept of safe driving needs to be operationalized so it can be used in safety assessment or as part of the system. One practical solution is to formulate safety targets.

ADSs take responsibility for the entire driving task within the system's ODD, i.e., the conditions under which the ADS is intended to operate; thus their safety targets should define an agent or a set of rules that can safely manage all driving scenarios. While there are several paradigms for defining ADS safety targets, this work focuses on those based on human driving that consider driver behavior, capabilities, and constraints, as well as the associated crash risks. Achieving these targets ensures that the ADS performs at least as well as a human driver (or a population). This requirement is typically operationalized by developing CBMs that capture reference human behavior and comparing them with ADS performance in virtual simulations. However, the reference behavior might not reflect an important aspect of ADS evaluation: the system's ability to navigate traffic competently without being overly cautious or causing inconvenience to other road users. Fraade-Blanar et al. (2026) introduced the construct drivership, defined as driving actions aligned with societal norms and expectations for road conduct. In combination, safety and concepts such as drivership are at the core of safe and accepted ADSs.

Unlike ADSs, ADASs share the driving tasks with drivers, who retain full responsibility for the operation of the vehicle. Therefore, ADAS safety targets must account for the combined performance of the driver and the ADAS, including their respective capabilities and constraints. The capabilities and constraints can be obtained by quantifying the safety implications of certain driver behaviors and, for example, setting thresholds for related metrics which can be used by the ADAS to trigger warnings or interventions. Another difference between ADSs and ADASs is that the nuisance factor is particularly important for the safety of the latter: an ADAS that issues warnings that the driver perceives as excessive or unwarranted may be disabled by the driver (if possible), or the warnings may simply be ignored altogether. This "cry-wolf" effect can negatively impact safety, since drivers do not act on warnings when they should. Although drivers may also feel annoyed with an ADS, this annoyance should not compromise safety because the ADS will continue driving.

Another difference between ADS and ADAS safety targets, as defined in this work, is that the former are typically used offline to assess the safety performance of the ADS. For ADASs, on the other hand, safety targets are typically an integrated part of the system: the ADAS acts with a warning or an intervention when it identifies deviations from safe driving (as defined by the safety target).

1.3.3 Reference driver models

Comparing the performance of reference CBMs with that of ADSs supports one of the key concepts of ADS safety assessment: achieving a positive risk balance (PRB; Di Fabio et al., 2017; ISO, 2025). A PRB is achieved by ensuring that ADSs "decrease, or at least do not increase, the amount of harm" compared to a benchmark (p. 25; European Commission, 2020). The average driver's performance may, however, not be considered good enough for ADS safety assurance (Wood et al., 2019). Therefore, it has been proposed that human benchmark models should be representative of "attentive, skilled [sic] experienced" (p. 6; Rothoff et al., 2019) or "competent and careful" (p. 8; UNECE, 2023) drivers. These models, which describe the behavior of one driver with a specific skill set, can be used to quantify the safety impact of ADSs through pass/fail assessments. In addition to one-driver models, reference models may represent a population of drivers with a spectrum of driving styles, from conservative and safe to risk-taking and aggressive. These models can be used to go beyond the pass/fail assessment of one-driver models, quantifying the safety impact in terms of crash and injury reduction (Fahrenkrog et al., 2024; ISO, 2024).

In summary, the term "reference driver model" refers to a mathematical and/or statistical representation of one human driver or a population (distribution) of human drivers which is considered a benchmark (or a reference) for safe driving. Since reference driver models represent the level of safety performance that a human can reasonably achieve (as decided by, for example, a regulatory body), they can serve as safety targets for a set of systems.

At the time of writing this thesis, although the use of reference driver models in regulations is still being debated, models are already being incorporated into regulations. The one-driver reference models in R157, stated to be mathematical representations of a competent and careful human driver, are used to provide guidance in the defining of avoidable and unavoidable crashes in three types of scenarios (UNECE, 2023). (See Section 2.2.1 for a more detailed description of the models.) The first scenario is a deceleration (rear-end) scenario: the ego vehicle (the vehicle downstream in traffic, controlled by the driver model) and the principal other vehicle (POV; the vehicle that the ego vehicle is conflicting with) are traveling in the same lane and the ego vehicle needs to brake to avoid crashing into the POV. The second is a cut-in scenario: the POV moves into the ego vehicle's lane from an adjacent lane. The third is a cut-out scenario: the POV moves out of the ego vehicle's lane, revealing a slower vehicle ahead of the ego vehicle. Whether these models represent the behavior of a competent and careful driver in these scenarios has not, however, been confirmed. Mattas et al. (2022) applied these models to the highD (site-based naturalistic driving data from Germany; Krajewski et al., 2018) dataset scenarios, finding that one of the models often reacted to the lateral perturbations of surrounding vehicles, even when they were not changing lanes.

1.4 Aims and objectives

The aim of this work is to support the development and assessment of ADASs and ADSs by assessing reference driver models and characterizing driving through a safety-oriented lens, using real-world data and counterfactual simulations. The findings can guide the development of realistic and effective safety targets for both ADASs and ADSs.

To achieve this aim, the work addresses the following four objectives:

- To quantify the influence of data selection in pre-crash virtual safety assessments by comparing simulation outcomes across crash, near-crash, and everyday driving data.
- To evaluate the "competent and careful" UNECE reference driver models (i.e., ALKS) by applying them to near-crash scenarios from the Strategic Highway Research Program 2 (SHRP2) NDS.
- To analyze the timing of lane changes in naturalistic highway driving after the activation of the turn signal.
- To investigate to what extent drivers change their glance behavior to compensate for the increased crash risk associated with shorter time gaps and time to collision (TTC) in naturalistic car-following scenarios.

2 Methods

This chapter is divided into three parts. First, Section 2.1 describes the different types of data sources used in safety assessment methods, with a focus on the datasets used in this work. Second, Section 2.2 describes various aspects of analyzing and modeling driving behavior, paying particular attention to the analysis performed and the driver models used in this work. Third, Section 2.3 provides a more thorough description of the fundamental role that counterfactual simulations play in simulation-based safety assessments, and explains how these simulations can be used to generate crashes and assess driving behavior.

2.1 Data

The type of data used in virtual safety assessments plays a major role in the evaluation of systems and driver models. Four different data types were used in this thesis: in-depth crash databases, NDSs, synthetic crash databases, and realistic on-road data collections. This section presents the key elements of the data types, emphasizing the elements that determine their suitability for use in virtual safety assessments.

2.1.1 In-depth crash databases

In-depth crash databases consist of collections of detailed crashes, meticulously analyzed and reconstructed to generate a digital version of the events leading to the crash, and of the crash itself (Bakker et al., 2017). Examples include GIDAS in Germany (German In-Depth Accident Study; Otte et al., 2003), RAIDS in the UK (Road Accident In-Depth Studies; Cuerden & McCarthy, 2016), RASSI in India (Road Accident Sampling System – India; Rameshkrishnan et al., 2013), CIDAS in China (China In-Depth Accident Study; Chen & Dai, 2018), and CISS in the USA (Crash Investigation Sampling System; Zhang et al., 2019). Another example, IGLAD (Initiative for the Global Harmonisation of (in-depth) Accident Data; Bakker et al., 2017), harmonizes crash databases from various countries.

Data from these databases are commonly used for virtual safety assessment, both as input data (Bjorvatn et al., 2021) and as validation data (Bärgman et al., 2024). Moreover, the data are often used to virtually assess the crash-avoidance capabilities of ADASs and ADSs.

The in-depth reconstructed crashes used in this study come from GIDAS; the crash kinematics and impact speed are reconstructed for each crash. For a subset of the GIDAS reconstructed crashes, a pre-crash matrix (PCM) was created which includes detailed pre-crash kinematics for up to five seconds prior to each crash (Schubert et al., 2013). A specific subset of the GIDAS data consisting of rear-end crashes was used. The crashes were divided into two subsets: the first consisted of 134 crashes for which PCM data were available and the second consisted of 46 crashes without PCM data. In the latter, the ego vehicle did not show signs of an evasive braking maneuver—which is important for Paper I.

2.1.2 Naturalistic driving studies

NDSs are another common data source for virtual safety assessment. They can be either on-site or in-vehicle. On-site NDSs consist of location-based data collections, typically using a camera at a fixed location or on a drone, which records vehicles' trajectories (Bock et al., 2020; Krajewski et al., 2018; Krajewski et al., 2020; Laureshyn, 2010; Smith et al., 2009). On the other hand, in-vehicle NDSs are collected from vehicles equipped with dedicated sensors (e.g. cameras, radars, additional accelerometers; Blatt et al., 2015; Dingus et al., 2006) and driven by volunteers. Data are continuously recorded during the driving task, so all the driving scenarios that occur can be captured, regardless of the physical location. In many NDSs, the volunteers are not professional drivers: their only driving training is that required to obtain a driving license. Recently, however, the development of ADASs and ADSs has pushed automakers to collect proprietary driving datasets that reflect many of the characteristics of typical NDSs, such as the naturalistic setting of the driving conditions, with the difference that typically the drivers collecting the data are company employees (see Section 2.1.4 for a more detailed description).

NDSs can contain crashes, albeit rarely (Van Nes et al., 2013). Even when crashes do occur, there is typically less information for some crash aspects (e.g., injury outcomes and forces exchanged during the crash) than is available from in-depth reconstructed crash databases. However, the latter contain descriptions of road user pre-crash trajectories that are mostly based on assumptions about vehicle motion, while the NDS recordings can be used to extract (pre-crash) time-series data of the conflict kinematics, including trajectories, of the vehicles and other road users (Hankey et al., 2016; Krajewski et al., 2018).

NDSs also include other safety-critical driving scenarios, such as near-crashes, which need to be identified and separated from the many hours of uneventful driving typically collected. This work is usually done by automatic kinematic triggers (Hankey et al., 2016), which activate under certain conditions, such as when harsh braking is detected or when surrounding vehicles get unusually close to the instrumented vehicle. Additionally, drivers can often manually flag events they consider relevant with a button press. Expert annotators also play a role in identifying and classifying safety-critical events by reviewing the recordings, potentially finding events that the kinematic triggers missed. Because the majority of driving hours recorded by NDSs are non-safety-critical, the recordings can be used to study and model driver behavior (e.g., Das et al., 2023; Hankey et al., 2016; Markkula et al., 2016; Seppelt et al., 2017) and to study false system activations (but see, e.g., Svenson et al., 2017), which is beyond the scope of this work.

This work uses data from two different NDSs. The first is SHRP2, a large NDS collected in the USA (Blatt et al., 2015; VTTI, n.d.). The dataset includes data from more than 3000 volunteers, whose vehicles were equipped with cameras, radars, and other sensors (e.g., accelerometers and gyroscopes). Additionally, data from the GPS and the vehicle-internal controller area network (CAN) bus were also included. The data, collected between 2010 and 2013, have been widely used in research (e.g., Chen et al., 2022; Das et al., 2023; Hozhabr Pour et al., 2022; Markkula et al., 2016). The study for Paper I used a subset of

SHRP2 consisting of 211 rear-end near-crash events. For the study in Paper II, 38 cut-in near-crash events from SHRP2 were used.

The second NDS used in this work is the highD dataset (Krajewski et al., 2018). This dataset, collected between 2017 and 2018, is an on-site dataset consisting of recordings made by drones over sections of German highways. The data consist of 60 recordings with an average duration of approximately 17 minutes each. The trajectories of 110,000 vehicles were extracted from the recordings.

2.1.3 Synthetic crash databases

Another source of crashes for use in virtual safety assessment is synthetic crash databases. These databases include crashes whose kinematics have been synthetically generated from real-world scenarios (approach C2 in Wimmer et al. (2023); see also Fig. 1). Synthetic crashes are primarily obtained by one of two methods: the CBM-based method (Feng et al., 2021; Li et al., 2019) and the purely statistical method (Appendix 6.2 in Menzel et al., 2025). The former may be simulations based entirely on CBMs (e.g., traffic simulations; Van Lint & Calvert, 2018), or counterfactual simulations that may or may not include CBMs. Traffic simulations typically require lengthy, resource-intensive simulations to generate enough crashes across the full range of relevant severities, from crashes incurring property damage only (PDO) to fatal crashes. As a result, safety surrogate measures are often used instead of crashes when assessing safety using traffic simulation (Gettman & Head, 2003; Tarko, 2018). One example of such a measure is the number of events with TTC lower than a specific value (Westhofen et al., 2023; Åsljung et al., 2021). Scenario generation of non-crashes is however beyond the scope of this work. When a CBM is applied counterfactually in scenarios from crash databases, it typically only replaces one of the involved road users, and trajectory data from reconstructed in-depth crash data are used to represent the other road user(s).

The purely statistical method uses statistical variations of parameters describing pre-crash kinematics from one or more in-depth crash databases. This approach does not include CBMs; in fact, the baseline generation may not even include simulations (i.e., the baselines may be generated using only parameterized time-series data without any models needing runtime simulations). Both the counterfactual and the purely statistical methods can generate many more crashes than existed in the initial data: that is a reason for using these methods. It is important to note that if NDS data are used as a starting point for counterfactual scenario generation, there may not be any crashes in the original data. Paper I shows that using non-crash data as a starting point for crash generation is problematic, since the generated crashes are biased towards lower severity. (A discussion on the validity of NDSs for crash generation can be found in Section 4.1.2.) On the other hand, crashes generated from in-depth crash databases are biased towards higher severities (Elvik & Mysen, 1999; Yamamoto et al., 2008), due mainly to selection biases in the data collection (e.g., crashes in GIDAS are only reconstructed if at least one person was injured in the crash; Liers, 2018). These biases can affect the validity of a safety assessment (or other type of traffic safety analysis) that uses these data (Bärgman et al., 2024; Leledakis et al., 2021; Wang et al., 2022).

To address this limitation in the in-depth crash database approach, Wu (2024) proposed a method that combines data from NDSs with pre-crash data from crash databases, creating a set of models that generates crashes representative of real-world crashes across all levels of severity (from PDOs to fatalities). Paper IV in this thesis uses crashes generated by Wu, Flannagan, et al. (2025). Specifically, a subset of 174 generated rear-end crashes in car-following scenarios were used.

2.1.4 Realistic on-road data collections

Also used in this thesis are "realistic on-road data collections", which consist of recordings of driving sessions typically led by organizations with the aim of developing new vehicle technologies. In these collections, drivers are instructed to drive in a natural way without being given a specific driving task; they do not test a particular function which could affect their manual driving. However, unlike in NDSs, the data collected are not purely naturalistic. First, the drivers are usually employees of the company which collects the data; they may have been required to attend special driver training, so they may not be a completely realistic representation of the driving population. Second, the driving sessions often do not represent purely naturalistic driving. For example, the driving session may start and end at company sites, not at the drivers' homes. They may even be driving on prescribed routes, but without instructions to perform a specific task or do anything other than "just drive". Although not completely naturalistic, the data are often used to investigate driver behavior (Fries et al., 2023). Data from two of these realistic on-road data collections were used in this work.

The first was collected by Volvo Car Corporation (VCC) as part of the L3Pilot project (Penttinen et al., 2019). The drivers were exclusively Volvo employees. The collection consists of 6000 hours of driving, mostly on a highway ring-road around Gothenburg, Sweden. The vehicles were Volvo XC90s equipped with cameras facing the surrounding traffic and the drivers, as well as radars, accelerometers, angular rate sensors, and GPS. The data also include signals from the CAN bus as well as information about the longitudinal and lateral position and speed of surrounding vehicles.

The second realistic data collection was also done by VCC, using vehicles equipped with the latest generation of technologies available in the company at the time of data collection (2021–2022). The data were collected for use in the development of their ADASs and ADSs. The vehicles were equipped with, for example, cameras and lidar facing the surrounding traffic, and a driver monitoring system (DMS) providing glance data. For this work, a total of 53 hours of at least 20 s segments of steady-state car-following was used, from approximately 50 drivers employed specifically to collect data in this way.

2.2 Analysis and modeling of driving behavior

This thesis is part of a project which in part aims to contribute to the development of safety targets for ADASs and ADSs. Section 2.2.1 describes reference driver models for ADS assessments in depth—in particular the two existing computational reference driving models available in R157, which were evaluated in Paper II. Section 2.2.2 describes the driving behavior characterization performed in Paper III, which studied the characteristics of lane-change initiation. This work is relevant to the development of ADASs and ADSs and, in part, the formulation of safety targets. Finally, Section 2.2.3 describes the driving behavior characterization from Papers IV and V, which studied driver off-road glances as a function of driving context; this work, too, is relevant to the formulation of safety targets for both ADASs and ADSs.

2.2.1 Reference driver models for ADS safety assessment

The two reference driver models used in Paper II are referred to as “Performance model 1” and “Performance model 2” in R157. In this work they are referred to as the “competent and careful driver model” (CCDM) and the “fuzzy safety model” (FSM), respectively.

The CCDM is a threshold-based model that reacts to three traffic scenarios: decelerations (rear-end), cut-ins, and cut-outs (JAMA, 2022). The model observes the kinematics of the surrounding vehicles, using the TTC and lateral position of the POV as metrics for the assessment of possible threats, and (if needed) it reacts by braking. For the cut-in scenario (the only scenario considered in Paper II), the CCDM defines a “wandering zone”, centered in the POV’s lane. Its width is the width of the POV plus an additional 0.375 m on each side of the POV. The extra width allows for drivers’ small lateral corrections, which are to be expected in normal lane-keeping since drivers do not stay exactly in the middle of the lane. If the POV remains inside this zone, the driver of the ego vehicle is assumed to ignore the POV’s actions (i.e., no cut-in is detected). The CCDM detects an imminent cut-in only when the POV exits the wandering zone and the longitudinal TTC is less than 2 s. The total reaction time of the CCDM is made up of perception time (0.4 s), which starts when the cut-in is detected, and a subsequent braking delay of 0.75 s. At the end of the reaction time, the deceleration increases gradually with a jerk of 12.65 m/s^3 until a deceleration of 7.6 m/s^2 is reached.

The FSM, on the other hand, predicts the possibility of a collision based on lateral and longitudinal safety checks (Mattas et al., 2020; UNECE, 2023). When it is predicted that the trajectories will overlap in some future, the FSM applies the brakes. This model is capable of braking with any deceleration value between 0 and the maximum reachable deceleration (set to 6 m/s^2). It should be noted that the braking reaction of the FSM does not always reach the maximum braking capability of the driver. The actual deceleration value is determined using two metrics computed during the safety checks: the predictive fuzzy surrogate metric (PFS) and the critical fuzzy surrogate metric (CFS). The default value of both metrics is 0. For low-criticality scenarios, only the PFS changes value, increasing based on the criticality. The required deceleration increases gradually with the PFS, from 0 m/s^2 when PFS = 0 to 6 m/s^2 when PFS = 1 (the maximum value). However, if the scenario is critical and a collision is imminent, the CFS also increases, and the maximum required deceleration is reached with a jerk of 12.65 m/s^3.

There are also ADS safety-assessment approaches that, unlike the CCDM and FSM, do not use human behavior as a reference, instead relying on criteria based on objective safety. One such model is the responsibility-sensitive safety (RSS) model (Shalev-Shwartz et al., 2017). The RSS defines a set of rules (described by mathematical equations) to ensure that the ego vehicle is always in an objectively safe position with respect to the other road users; if the ego vehicle is in a safe position (e.g., far enough away from other road users), the ego vehicle can always avoid causing a collision. If all the vehicles respected these rules, there would be zero crashes, as all vehicles would have time to avoid a crash. However, it has been considered to be overly cautious (Mattas et al., 2022). In any case, ADS reference behavior models not based on driver behavior are beyond the scope of this work.

In summary, the safety assessment of ADSs can include many components, one of which is the comparison of the ADSs' safety performance against human reference models—rather than against what is objectively safe (e.g., the RSS). A discussion of safety target definitions, particularly comparing the use of observed human driving behavior to the use of objective safety, can be found in Section 4.4.1.

2.2.2 Understanding surrounding traffic in the lane-change scenario

Lane changes are common scenarios in everyday traffic in which drivers move to the left or right from their current lane to an adjacent lane (defined here as the target lane) traveling in the same direction. Understanding the driving behavior mechanisms that characterize lane changes is fundamental to building accurate CBMs of lane changes (and others' reactions to them). Every lane change comes with the risk of creating a conflict with vehicles in the target lane; they may be traveling at speeds higher than the ego vehicle, or the drivers may be distracted and not notice the vehicle changing lanes immediately. Understanding how lane changes are made and how they are identified by other road users can help developers of safety systems and safety targets for those systems.

The reaction time of a driver in the target lane depends in part on the driver's ability to detect the lane-change maneuver. It is, however, difficult to find consensus on the definition of the initiation of a lane change (Erdmann, 2015; Hidas, 2005; Moridpour et al., 2010). For example, it may be defined as the moment in which the lane-changing vehicle crosses the lane marking and entering the target lane. Another definition, given in Section 2.2.1 as part of the Performance model 1 (JAMA, 2022; UNECE, 2023), describes the initiation of a lane-change as the moment when a lane-changing driver leaves the wandering zone. This definition is more conservative than the first, as the lane change is detected before the lane-changing vehicle reaches the lane marking. A third definition, used by the Performance model 2 in (UNECE, 2023), assesses the possibility of conflicts between the lane-changing vehicle and the vehicle in the target lane by projecting their positions into the future to see if they may collide. In the absence of additional constraints, this definition will, in many cases, yield conservative—and potentially erroneous—estimates of lane-change initiation: a projected trajectory leading to a crash may result merely from minor lateral movements within the lane or from road curvature. Paper III introduces a definition of lane-change initiation based on a combination of lateral speed and position thresholds. Using this definition, a lane change is marked as initiated when the vehicle reaches a lateral velocity of 0.15 m/s. An additional check on the lateral position is made, so that even if the vehicle never reaches the lateral velocity threshold, the lane change is marked as initiated when the vehicle is moving laterally towards the target lane and positioned 5 cm or less from the lane marking. The lateral speed threshold was selected by first analyzing data of uneventful free driving (without lane changes). In 90% of the free driving, vehicles did not move laterally with speeds higher than 0.17 m/s. Lateral speeds from 0.10 m/s to 0.30 m/s (with 0.05 m/s increments) were then tested on real-world lane changes to assess which was most accurate for detecting lane changes. It was found that threshold values greater than or equal to 0.20 m/s would cause lane changes to be determined purely by the lateral distance threshold. A threshold of 0.15 m/s was most accurate in detecting lane changes in the available data, even though it did result in some false positives.

Paper III goes beyond the proposed method for identifying the start of a lane change to provide a more detailed categorization of the timing of lane-change initiation. It examines factors that influence when and how drivers initiate lane changes. While this information may not be directly applicable to developing reference driver models and safety targets, it can inform the creation of routine driving models that represent "average" drivers. These models can then complement the near-crash data used in Paper II to assess the performance of reference driver models, which should naturally be more conservative than average driving behavior. An even more important use of this information may be to identify which factors should be included in reference driver models.

2.2.3 Glance behavior analysis

As mentioned, the analysis of driver glance behavior is a critical part of understanding the impact of driver behavior on safety. Drivers are responsible for being aware of their surroundings, achieved mostly by looking at the forward road, along the sides of the vehicle, and at the side and rear-view mirrors. Additionally, demanding traffic conditions require the drivers to adapt their glances to the context to maintain safe driving (Tijerina, 1999). Typically, glance behavior studies define areas within the driver's field of view to categorize glance directions (see Kircher & Ahlström, 2024 for a discussion on glance coding approaches). A common categorization defines glances as on-road when they are directed towards the forward road and off-road when they are directed elsewhere. These definitions are relatively straightforward; during off-road glances, the driver may not be aware of changes in traffic conditions in front of the vehicle, and this lack of awareness may lead to a conflict or a crash. Another categorization divides the field of view into multiple areas, such as the front windshield, the dashboard, the infotainment system, and other passengers (ISO, 2020). In this work, the on/off-road glance dichotomy has been used.

In Paper IV of this thesis, car-following driving segments were extracted from a large, realistic on-road data collection and the glances in each segment were categorized as off-road or on-road. The glance direction was determined using DMS sensors and logic. The glance behavior was first analyzed by evaluating the total percentage of time spent looking at the road. This metric is widely used in studies that focus on analyzing general attention levels on the road (Seppelt et al., 2017; Victor et al., 2014). Second, the duration of each off-road glance was computed, resulting in a distribution of glance durations in the driving scenarios. Third, for each steady-state driving segment (e.g., a segment with the ego vehicle following a lead vehicle at a relatively constant relative speed for at least 20 s), the single longest off-road glance was extracted. The duration of this glance was used to represent the off-road glance duration limit, beyond which drivers do not feel comfortable looking away from the road, given a specific context (in Paper IV, considered to be the time gap to the lead vehicle). See Section 4.3 for a discussion of this approach.

In addition to the analysis of glance behavior as a function of time gap in Paper IV, in Paper V we analyzed the glance behavior as a function of TTC. The analysis focused on the initiation of off-road glances in relation to the inverse TTC, as well as the duration of off-road glances as a function of inverse TTC. A key challenge (also seen in Paper IV) was the very limited exposure in the distribution tail. Even with relatively large datasets, the uncertainty becomes substantial—for the shortest time gaps in Paper IV, and for the largest inverse TTCs in Paper V. (The latter is the greater problem). Another challenge for both papers was that the analyses could not be performed on a per-driver basis, as it was unknown which driver drove when. Generalizability is thus limited, as the tail of the distribution may be dominated by data from only a few drivers. (See Section 4.5.1 for more details).

2.3 Simulation-based safety assessment

2.3.1 Counterfactual simulations for ADAS and ADS assessment

This work categorizes the methods for virtual safety assessment following the work by Wimmer et al. (2023), as briefly outlined in the Introduction. Even though the three approaches have their individual properties, they also share similarities.

As mentioned, approaches A and B (Fig. 1) both use counterfactual simulations to prospectively evaluate system safety by modifying the kinematics in baseline scenarios and comparing outcomes with and without the system. Unlike approach A, however, approach B also uses counterfactual simulations to generate the baseline by modifying the original scenarios before applying the system. Fig. 2 illustrates this for rear-end crash scenarios; the system or model replaces the ego-vehicle behavior and alters its kinematics while the POV kinematics remain unchanged.

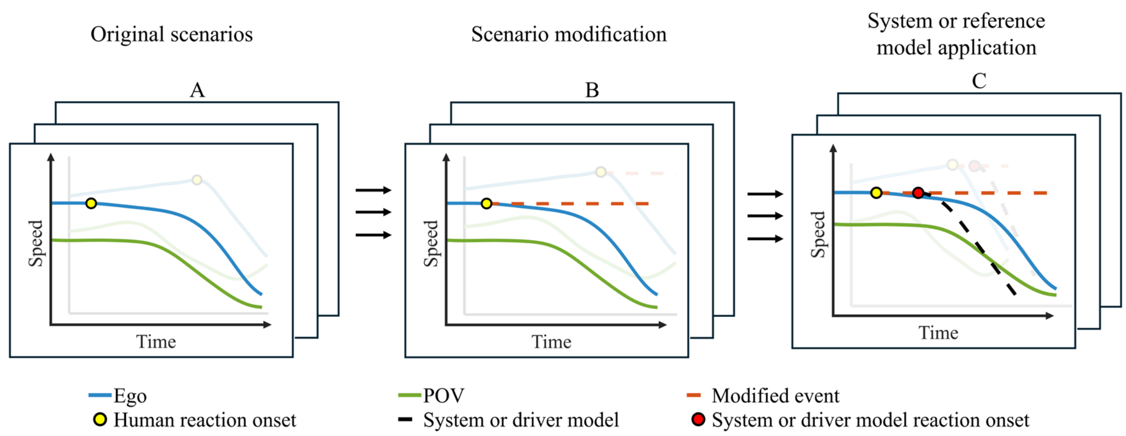

Fig. 2 Counterfactually modifying rear-end traffic scenarios by removing the original evasive maneuver of the ego vehicle and replacing it with a model (of a system or reference model), keeping the kinematics of the POV.

As mentioned, counterfactual simulations have been used to assess ADASs (Erbsmehl, 2009) and ADSs (Bjorvatn et al., 2021; Scanlon et al., 2021), as well as to evaluate driver behaviors (Bärgman et al., 2017; Bärgman et al., 2015; Lee et al., 2018); see

Section 2.3.3 for more details. The studies that are part of this thesis used counterfactual simulations with baseline approach B, using data from crashes, near-crashes, and everyday driving; systems and driver models were then applied to the baselines (i.e., the modified original scenarios).

Fig. 2C illustrates the application of a system or a driver model that initiates a braking maneuver, creating counterfactual versions of the original traffic scenarios. In these modified scenarios, the system or driver model can brake earlier or later than a human driver in the real world. (In the figure, only the case in which the reaction happens later is shown.) This change in timing of the evasive maneuver exposes one of the main challenges of counterfactual simulations: if a system is applied without giving careful thought to the timing, the validity of the simulations (and consequently, the safety assessment) can be undermined. It is important to choose a moment to apply the system under assessment (treatment) that does not fundamentally change the configuration of the safety-relevant event. As an example, imagine that a system including adaptive cruise control (ACC) is applied to the pre-crash kinematics of rear-end crashes—which typically include data for five seconds before the crash. The ACC function is designed to keep the desired speed in traffic unless the vehicles ahead are slower or stopped. Applying the ACC function to those few seconds of available pre-crash kinematics may not give the ACC enough time to reach its steady-state condition (the desired speed), compromising the validity of the simulation's representation of a conflict situation. In other words, a vehicle equipped with ACC would never have ended up in a traffic scenario described by the few seconds of pre-crash kinematics typically available for analysis, because it would have started slowing down earlier, perhaps avoiding the conflict completely, without the need for any emergency action (e.g., by an AEB system). This example illustrates the importance of accurately accounting for the timing of the application when performing prospective safety assessment of systems and driver models.

As explained below, counterfactual simulations were used throughout this thesis, first to generate crashes from non-crash data, and then to evaluate the behavior of human drivers and driver models.

2.3.2 Generation of crashes from non-crash data

Paper I investigates the challenges associated with increasing the number of baseline cases by using counterfactual simulations to generate rear-end crashes using two sources: SHRP2 near-crashes and highD car-following scenarios when the lead vehicle is braking. Generating baseline crashes using CBMs through counterfactual simulations requires that any evasive maneuver by the original driver in the ego vehicle be removed, to have a clean slate for the driver model (see Fig. 2B). Using the same CBM(s) for the baseline generation and the treatment ensures that any positive or negative effects of the system are due exclusively to the system. At the same time, however, the conditions that lead to the critical event must be maintained, so that the criticality of the scenario is preserved. (A similar problem with the counterfactual application of an ACC system was described in Section 2.3.1.) The problem is deciding at what point the original driver's evasive actions are to be replaced by those of the model. If they are replaced too soon the modified event may differ greatly from the original, and as a result the link to a real-world case gets weaker. If they are replaced too late, the model may not start intervening until after the original evasive maneuver, so the baseline is not a “clean slate” anymore. This problem is usually addressed by identifying the onset of the original driver’s evasive maneuver and removing the vehicle’s trajectory after that point (Bärgman et al., 2017; Bärgman et al., 2015; Scanlon et al., 2021). A new trajectory, generated from assumptions dictated by design choices (e.g., maintaining constant speed or constant acceleration), replaces what was removed (Fig. 2B). The CBM is then acting on this clean slate. For Paper I, the simplest behavior possible was investigated: crashes were generated by a following driver not responding at all to the actions of the lead vehicle.

2.3.3 Counterfactual simulations for driver behavior assessment

In addition to evaluating reference driver models, counterfactual simulations can be used to assess the safety implications of specific driver behaviors to quantify their impact on crash risk. This section focuses on the use of counterfactual simulations to assess a) reference driver models (from UNECE, 2023) in Paper II and b) the impact of drivers’ glance behavior on safety in Paper IV.

Paper IV examines how drivers’ glance behavior during car-following scenarios affects the ability to avoid rear-end crashes, using an approach originally proposed by Bärgman et al. (2015). The approach, as applied in this work, involves combining glance behavior distributions extracted from a large realistic on-road dataset with synthetic crash scenarios from Wu, Flannagan, et al. (2025). The synthetic crashes used in Paper IV include detailed pre-crash kinematics of car-following scenarios (as described in Section 2.2.3), in which the lead vehicle brake and the ego vehicle does not, resulting in a rear-end crash. The kinematics were modified to obtain crashes in which the initial car-following conditions match the range of time gaps for which glance behavior data were available. Introducing glance behavior in the simulation creates the possibility that the driver is looking away from the road (and therefore away from the threat) at the moment the scenario is no longer benign, and a crash becomes imminent. Counterfactual off-road glances can, therefore, delay the reaction time of the driver, causing a crash which can be less or more severe than the original.

The crash-risk estimation procedure starts from a set of seed (original) crashes; the original evasive maneuver of the following vehicle is removed, to create a standardized clean-slate baseline for counterfactual simulation. For a given driving context (here, a specific time gap), an off-road glance duration distribution is binned, and each glance duration bin is combined with the driver response model (the reaction time followed by a braking maneuver). The resulting counterfactual simulation is run and the outcome (crash or no crash) is recorded. This sampling-simulation-outcome loop is repeated until the full glance distribution is represented (i.e., all off-road glance duration bins are simulated). Each simulation then has a probability associated with it (based on the glance bin probabilities), that is used to weight the results. The same procedure is carried out across all seed crashes to cover the variability in crash kinematics. The results from all combinations of seed crashes and sampled glances (with weights) are aggregated into the model-estimated crash risk (MCR) for that context by combining the empirical crash-risk function with the glance-duration distribution. Uncertainty in the estimated risk is quantified via bootstrapping, resampling both crashes and glance observations to obtain a distribution (and confidence bounds) of the MCR estimate. Finally, the entire pipeline is repeated for each available time-gap bin, yielding context-dependent crash-risk estimates across the studied range of car-following conditions. This process is represented in Fig. 3.

The results of Paper IV show the increasing crash risk associated with the glance behaviors as the time gap decreases. A major implication is relevant to the aim of this thesis: the results can be used as a first step towards defining safety targets for ADASs. For example, a warning system based on glance behavior can use these results to factor in the safety implications of off-road glances in their warning logic.

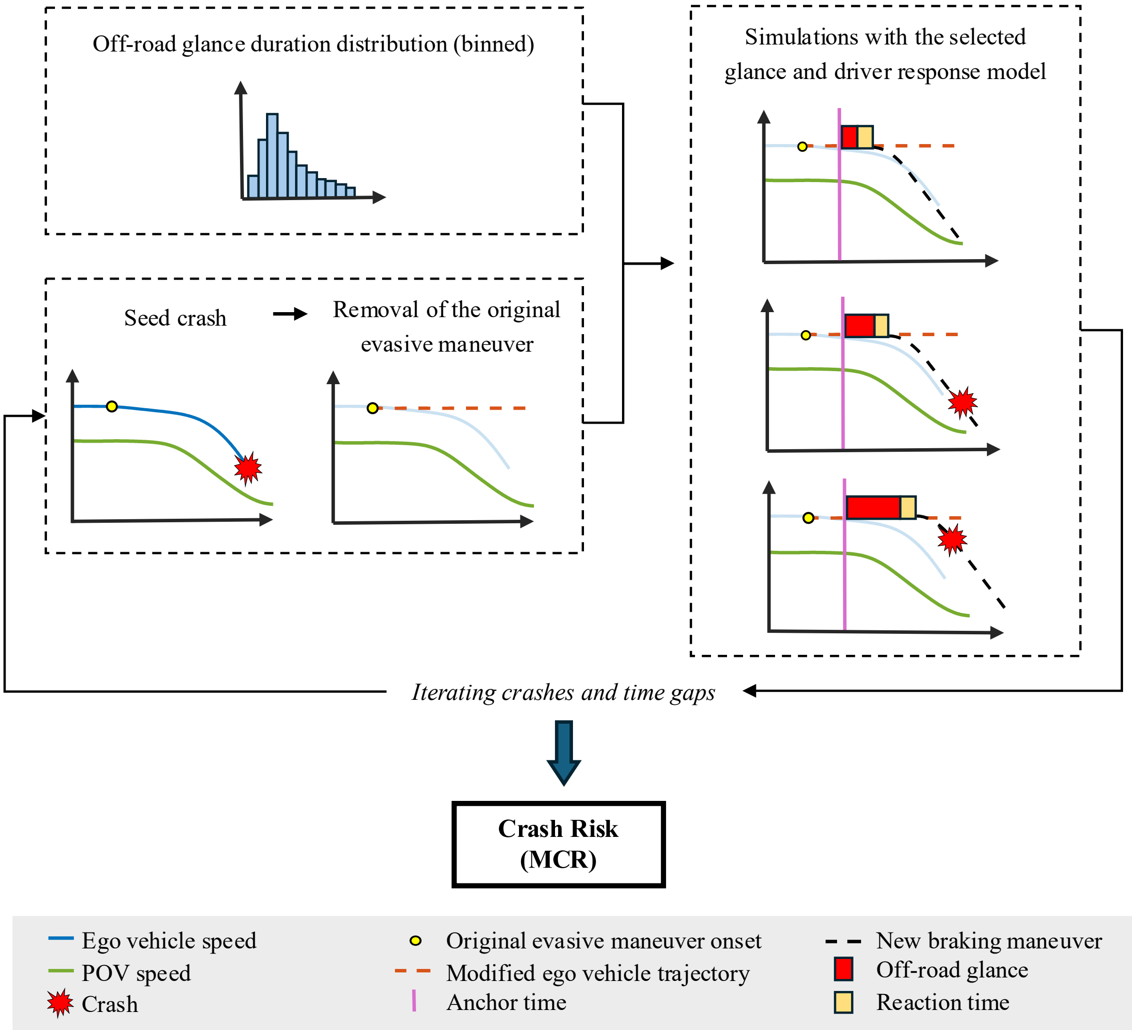

Fig. 3 Schematic of the counterfactual simulation procedure used to compute MCR. The figure shows three examples (right side) of the combination of seed crashes modified by removing the original evasive maneuver with the off-road glance distribution (left side). The three examples use three different durations of the off-road glance, therefore showing different outcomes (crash or no crash).

3 Summary of papers

3.1 Paper I

Can non-crash naturalistic driving data be an alternative to crash data for use in virtual assessment of the safety performance of automated emergency braking systems?

Olleja, P., Bärgman, J., & Lubbe, N. (2022)

Introduction

Assessing the safety benefit of AEB systems before their release on the market requires realistic crash data. In-depth crash databases provide detailed pre-crash kinematics, but are scarce. NDSs offer broader coverage but rarely include crashes. This study investigates whether crashes generated from non-crash NDSs can serve as substitutes for real crashes in the safety benefit assessment of AEB, focusing on differences in scenario criticality and severity.

Method

Data from two NDSs were used: highD everyday driving and SHRP2 near-crashes. Rear-end crashes were artificially generated by replacing the original evasive maneuver with a model of an unresponsive driver. These generated crashes were compared to in-depth rear-end crash reconstructions from GIDAS. Two AEB algorithms were applied to the generated scenarios: a basic AEB using longitudinal braking only, and a more advanced AEB that also considers comfortable steering (to reduce false positives).

Results

Crashes generated from SHRP2 near-crashes showed levels of criticality similar to GIDAS crashes, but lower severity. The highD-based crashes were both less critical and less severe than GIDAS crashes. Both AEB algorithms avoided more crashes in the NDS-generated datasets than in the GIDAS reconstructions, reflecting the lower severity of GIDAS crashes.

Discussion and conclusions

This work studied the validity of crashes generated from non-crash NDS datasets, to determine whether the generated crashes can be used in the virtual safety assessment of ADASs. The results show that the crashes generated from highD were both less severe and less critical than the real crashes, so applying this process to highD data is not a viable option for increasing the amount of data for use in virtual safety assessments. On the other hand, crashes generated from SHRP2 showed levels of criticality similar to that of crash-database crashes, meaning that SHRP2 is more suitable for generating crashes than highD. However, more research is still needed in order to make crashes generated from non-crash data more realistic—and to determine whether simulations based on behavior models built solely on non-crash data can produce realistic crash outcomes (e.g., relative impact speed and injury risk).

3.2 Paper II

Validation of human benchmark models for Automated Driving System approval: How competent and careful are they really?

Olleja, P., Markkula, G., & Bärgman, J. (2025)

Introduction

One of the requirements for ADS type approval in Europe is that the system must be at least as safe as a competent and careful human. This work investigated the validity of the two reference CBMs assumed to be representing competent and careful drivers in the UNECE Regulation No. 157. The safety performance of the original human drivers in a set of near-crash cut-in scenarios was compared to that of the UNECE models, which counterfactually replaced the original drivers' behaviors.

Method

First, videos of 38 near-crash cut-ins from SHRP2 were manually annotated to extract the vehicles' trajectories, and the original braking evasive maneuvers performed by the drivers were removed. The UNECE models were then counterfactually applied to the modified SHRP2 events. The timings of the evasive braking for the original humans and the models were compared.

Results

The first model reacted 0.5 s later, on average, than the SHRP2 drivers. This delay resulted in a crash for three of the cut-ins. In contrast, the second model's reaction preceded the original onset of the SHRP2 drivers' evasive maneuver by 0.7 s on average. Other noticeable differences between the models and the SHRP2 drivers were found by analyzing the lateral positions of the POV in the lane: the first model reacted well after the start of the POV's lateral lane-changing motion; the second model reacted earlier, often when the POV had not moved much laterally (when it was still in what the first model would define as the wandering zone).

Discussion and conclusions

One of the UNECE models was not careful enough, as it caused crashes that had not happened in the original real-world scenarios. The other model, on the other hand, was found to be overly careful at times, and consequently not very competent. As the models' performance differed substantially from that of humans, their use in ADS virtual safety assessment may be problematic. The main takeaways from this work are that more work is needed in the development of reference driver models for ADS assessment, and that, after these models are developed, they need to be validated on data that ranges across all levels of criticality, from everyday driving to high-severity crashes.

3.3 Paper III

Analysis of time-to-lane-change-initiation using realistic driving data
Jokhio, S., Olleja, P., Bärgman, J., Yan, F., & Baumann, M. (2023)

Introduction
Lane changing is a frequent and relatively complex driving maneuver. Understanding how drivers perform this maneuver can support the development of driver models, ADASs, and ADSs. One component of a lane change is the time between activating the turn signal and initiating the maneuver (the time to lane change initiation: TTLCI). This study investigates TTLCI in real-world driving and identifies factors influencing it, aiming to inform traffic safety policies as well as the developers of ADASs and ADSs.

Method
The data were collected within the L3Pilot project on Gothenburg's urban motorway, using vehicles equipped with camera and radar sensors and data from the CAN bus. The study analyzed 1,073 lane changes which included drivers signaling before initiating the maneuver. TTLCI was examined using survival analysis and a regression model with driver-level effects to account for repeated observations per driver. Predictor variables included vehicle speed, lane-change direction and type, presence of lead/lag vehicles, and gap sizes. In addition, a practical method for identifying lane-change initiation from vehicle trajectories was developed and evaluated.

Results
Drivers typically initiated lane changes shortly after activating the turn signal. TTLCI varied systematically with driving context, depending on maneuver characteristics (type and direction), vehicle speed, and surrounding traffic (including lead/lag presence and available gaps). The lane-change initiation identification method provided a way to separate lane-change execution from normal within-lane lateral wandering, supporting more consistent measurement of TTLCI.

Discussion and conclusions
The findings indicate that, at least in the traffic culture studied here, drivers rarely provide sufficient advance warning of an imminent lane change and TTLCI is influenced by maneuver complexity, surrounding traffic, and speed. Incorporating TTLCI into traffic models and ADAS/ADS algorithms could improve the prediction of lane-change timing and enhance safety in mixed traffic. The study also shows that TTLCI depends on how lane-change initiation is identified in trajectory data. The proposed method—using observed lateral motion during normal lane-keeping to distinguish purposeful lane-change onset from within-lane wandering—supports a better identification of TTLCI. Future research should explore TTLCI for heavy vehicles, different road types, and driver demographics.

3.4 Paper IV

Analysis of the safety impact of off-road glances in car-following driving scenarios
Olleja, P., Svärd, M., Zhao, M., & Bärgman, J. (*manuscript for journal submission*)

Introduction
ADASs and ADSs aim to improve safety but must balance safety benefits against drivers' acceptance issues and annoyance. For ADSs, a key challenge is defining reference driver models that represent safe driving without being overly conservative. For ADASs, user-system interactions matter. For example, inattention information systems (IISs) based on DMSs must avoid excessive "cry-wolf" alerts that reduce acceptance. Current off-road glance guidelines are largely based on normative tasks and are not context-sensitive, yet glance behavior may vary with driving context. This study seeks to support the design of context-aware IIS design and more realistic ADS reference models by testing whether drivers adapt off-road glances to objective rear-end crash risk in steady-state car-following.

Method
We used >6,000 hours of realistic DMS-based on-road data with the gaze direction automatically classified into predefined zones. Steady-state car-following segments with stable time gaps were extracted. Off-road glance distributions were analyzed by time gap, including both the overall distribution and the single longest glance per segment. To assess drivers' risk compensation, we estimated rear-end crash risk using counterfactual simulations that combine time-gap-specific glance distributions with an empirical crash-risk function, yielding a model-estimated crash risk (with uncertainty estimated via bootstrapping). We also estimated what the objective crash risk would be if the drivers had not adjusted their glance behavior as the time gap decreased.

Results
Applying long time-gap glance behavior (3.1 s) to short-gap crashes resulted in even higher risk than the original (context-adapted) glance behavior, indicating drivers adapt somewhat to context. However, their adaptation was not sufficient to counteract the elevated risk at short time gaps; at 0.7 s gaps, drivers risk compensation reduced the crash risk by about 25% but the risk remained substantially higher than at longer gaps. In summary, the adaptation that drivers showed was not nearly large enough to counteract the increased risk of shorter time gaps.

Discussion and conclusions
At short gaps, the results show only minor reductions in glance duration, and these were mainly for the longest glances; drivers do not adequately regulate glance behavior to compensate for increased scenario criticality. Further, the crash risk rises sharply at short gaps, and while drivers adapt somewhat to the increased risk, the compensation is again limited. These findings contribute to the creation of context-aware IIS alerts, even though they do not provide specific values and recommendations. The study also highlights the value of large-scale DMS and external sensor data for understanding driver behavior in context and informing IIS design and ADS reference models.

3.5 Paper V

Drivers' glance-behavior adaptation in response to looming during car-following
Olleja, P., Svärd, M., & Bärgman, J. (*manuscript for journal submission*)

Introduction
IISs must balance safety benefits against nuisance alerts and drivers' acceptance issues. Existing glance-related guidelines and IIS implementations are not typically sensitive to driving context, despite the fact that crash risk changes markedly with traffic dynamics. This paper investigates whether drivers adapt their off-road glance behavior to objective crash risk in car-following when risk is expressed through looming (inverse time-to-collision; inv-TTC).

Method
This study used a large realistic on-road dataset, which included DMS-based glance classification and surrounding-traffic sensing. Highway car-following sequences in which the ego vehicle approached a lead vehicle at moderate-to-high speeds were extracted and off-road glances were analyzed as a function of looming at glance initiation. Off-road glance frequency was estimated across looming bins and described using distribution fitting and a risk-ratio analysis comparing "low" vs. "high" looming conditions. Off-road glance durations within each looming bin were modeled using lognormal fits, and changes across looming were quantified by analyzing duration percentiles.

Results
Drivers initiated far fewer off-road glances as looming increased. In fact, off-road glance initiation became rare at higher looming levels. In contrast with the strong effect on glance initiation, the mean off-road glance duration remained relatively stable across looming levels. However, longer glances became increasingly uncommon as looming increased. This pattern was most visible in the upper tail of the duration distribution, indicating that drivers constrain long off-road glances under higher objective risk.

Discussion and conclusions
The findings provide evidence that drivers modulate both off-road glance frequency and the tail of the off-road glance-duration distribution in response to looming. Thus looming appears to be a meaningful operationalization of driver risk perception for understanding attention allocation in car-following, aligning with earlier evidence that inv-TTC thresholds are important for driver response timing in rear-end conflicts. The results have implications for context-aware IIS strategies; warning logic incorporating looming could better reflect momentary criticality to reduce unnecessary alerts while maintaining safety. In addition, the results indicate the advantages of considering context-dependent glance behavior when developing behavior-based driver models and reference assumptions for ADS evaluation.

4 Discussion

This PhD project aims to characterize and assess safe driving in order to inform and improve the development of the safety targets used to assess ADASs and ADSs. Section 4.1 discusses the challenges presented by different data sources in creating a virtual assessment baseline, based on the findings of Paper I. Section 4.2 discusses the performance of existing reference driver models as analyzed in Paper II, and reflects on possible improvements to the models currently used in ADS safety assessments. Section 4.3 discusses the characteristics of driving behavior that are analyzed in Papers IV and V, with a focus on glance behavior context dependency and crash risk. The section also includes a reflection on the potential integration of driving behavior components such as glance behavior into reference driver models and safety targets for ADASs and ADSs. Section 4.4 gathers the main conclusions of the previous three sections to discuss their role in the formulation and development of safety targets. Finally, Section 4.5 describes the limitations and future directions of this research.

4.1 The importance of data choice in virtual safety assessment

Virtual safety assessment, a key part of ADAS and ADS development, is performed before their release on the market. The assessments need to be valid and based on data from real-world traffic, since these systems must be able to operate safely on public roads. Therefore, the real-world traffic data used must be relevant for the type and scope of the specific assessment to be performed. One of the aims of this work is quantifying the impact of the choice of data on the safety impact assessment of pre-crash safety systems. In practice, this means studying the impact of data collected in situations with different levels of criticality (i.e., everyday driving, near-crashes, and crashes) on baseline scenario generation and the subsequent virtual safety assessment of an ADAS. This section starts with a discussion of the approach to generating crashes used in Paper I (4.1.1). Then, the validity of crash-generation approaches in general is discussed (4.1.2).

4.1.1 Crash generation in Paper I

One of the objectives of this work is to quantify the influence of using non-crash data in baseline generation. Specifically, data from everyday driving and near-crashes were used to generate crashes, and the characteristics of the resulting crashes were analyzed. Near-crashes have often been proposed as surrogates for crashes (Guo et al., 2010; Hydén, 1987; Laureshyn & Várhelyi, 2018; Victor et al., 2014; Wu & Jovanis, 2012). However, they can be used as surrogates only if appropriate sampling and validation methods are employed, and then only within a given scenario type (Bärgman, Zhao, et al., 2025; Knipling, 2015). Everyday driving data, on the other hand, captures a much broader set of traffic scenarios. However, most of them are less relevant for safety than near-crashes and crashes, particularly when assessing an ADAS.

Naturally, an ADS should also be able to perform everyday driving safely, so less critical scenarios must be included in the assessment. However, that consideration is beyond the scope of this work, which is confined to safety assessments based on crashes.

Paper I used the assumption of an unresponsive driver to generate worst-case scenario crashes, using rear-end near-crash data from SHRP2 and (more or less) normal driving data from highD. The unresponsive driver replaced the driver of the ego vehicle in the original scenario, who had reacted and avoided a collision by braking. An unresponsive driver who fails to react can be considered the worst-case scenario; although theoretically the resulting crashes would be even more severe if the driver were to accelerate, the probability of that scenario occurring in the real world is low, as shown by Wu, Flannagan, et al. (2025). The generated crashes were then compared with real-world crashes. The results of the comparison indicate that they are profoundly different in their outcome severity. In the real-world crashes, the vehicles involved had much larger speed differences, and the POV performed a stronger braking maneuver than in the artificially generated crashes. Crashes generated from highD data, which at most contained very minor conflicts, were substantially different from the real-world crashes not only in terms of severity, but also in terms of criticality. Specifically, the time that it would have taken for an unresponsive driver to crash into the POV after it initiated braking was much higher in the highD-generated crashes than in the SHRP2-generated crashes. One reason for this difference is likely to be found in the deceleration values reached by the POVs in the highD data, which were much lower (due to less harsh braking) than the ones reported in the SHRP2 near-crashes.

However, it is not only the lower deceleration of the POV that distinguishes the real crashes from the crashes generated from routine driving data. The study in Paper I used only rear-end scenarios in which the POV had braked with more than 2 m/s^2 of deceleration. That is, only *car-following* crashes were considered, not *catch-up* crashes. In a catch-up crash, also called a *closure-from-long-range* crash (Woodrooffe et al., 2012), the POV is driving substantially more slowly than the ego vehicle (or even standing still), and the ego vehicle does not react early enough to the presence of the POV. The late reaction may be due to sleepiness/drowsiness, sudden sickness, or very long off-road glances/distractions. According to the GIDAS POV deceleration data in Paper I (Fig. 5a), more than 50% of the crashes were catch-up crashes, since the POVs did not decelerate at all.

It is clearly an intrinsic limitation of the method in Paper I that catch-up crashes could not be created, as only scenarios with the POV decelerating harder than 2 m/s^2 were used. However, it is not obvious how to generate catch-up crashes from everyday driving data such as highD, especially if the data consist mainly of dense traffic, since almost all vehicles are just following one another. The vehicles tend to travel at similar speeds and with relatively short car-following time headways (THWs); therefore, the severity of any generated catch-up crashes using these data tends to be low. An exception is the “end of traffic jam” scenario, when a vehicle encounters a traffic jam after having been driving at free-flow speeds: in this case, catch-up crashes generated from everyday data can be quite severe. This scenario is, however, hard to capture for drone-based datasets such as highD, as it would require that the drone be flying exactly over the end of a traffic jam. Actually, if available, these data would be valuable for future studies.

In Paper I, the validity of the generated crashes was assessed through the application of an AEB system. Its performance showed substantial differences in crash avoidance and mitigation across different crash datasets. Specifically, the proportion of crashes generated from highD which the AEB system avoided was substantially larger than the proportion of GIDAS crashes which were avoided. This discrepancy suggests that the generated crashes were not representative of real crashes. The AEB also avoided a larger percentage of highD-generated crashes compared to SHRP2-generated crashes, although the difference was less pronounced than that between highD and GIDAS. These differences support the argument that the set of crashes generated directly from the available non-crash data are not suitable for crash-outcome safety assessments. Note that Paper I focused solely on generating crashes directly from non-crash data. However, research suggests that it is possible to identify critical events (including varying crash severity levels) by combining non-crash data with machine learning and extreme value theory (EVT; Jiao et al., 2025). This approach may also support the generation of valid crash kinematics, although such applications remain unexplored. The scope of the research field validating scenario generation is certainly larger than that of Paper I, which only considered the generation of crashes of unresponsive drivers. Accordingly, Section 4.1.2 considers the results of Paper I from a broader perspective.

4.1.2 Implications for virtual safety assessment: Are the generated crashes valid?

Because crashes generated from near-crashes (and even lower-severity conflicts) in virtual safety assessments do not capture the criticality of real crashes, systems like AEB in Paper I have more time to assess the threat and react to it, avoiding the generated crashes more easily. These results raise the questions: *Are counterfactually generated crashes valid for the virtual safety assessment of crash avoidance systems?* and *What makes the generated crashes valid (or not)?* Bärgman et al. (2024) address scenario-generation validation for rear-end crashes, including a more realistic approach for crash generation than the one used in Paper I. Their method combines sleepiness with crashes generated from distributions of off-road glances, as well as crashes arising when ego drivers brake less forcefully than physically possible in the given situation.

Crashes modeled with responsive drivers will be inherently less severe than those modeled with unresponsive drivers. For example, crashes generated using a distribution of off-road glances (or simple distributions of reaction times; Kusano & Gabler, 2012) would include drivers who react more quickly to a critical scenario than crashes generated using a model of an unresponsive driver. Thus, the crashes generated using off-road glances (as in Bärgman et al., 2024) are inherently less frequent and less severe than unresponsive-driver-generated crashes. Fig. 4 in Bärgman et al. (2024) shows that the crashes generated with the more complex crash-causation model have about half the mean delta-v of the crashes generated with the unresponsive drivers. If, instead of an unresponsive driver, a more realistic crash generation model had been used in Paper I, the generated crashes would have been even more different from the real-world crashes than they were. It should be noted that

Victor et al. (2014) and Kusano and Victor (2022) leveraged the concept of an unresponsive driver in counterfactual simulations to define the maximum injury potential of a traffic scenario. This metric, in its simplest form, is derived by assuming an unresponsive driver. It enables comparisons of scenarios' potential severity, so that the relevance and appropriateness of various data sources to virtual safety assessment can be demonstrated.

The overall results from Paper I suggest that crashes generated from everyday driving, even in the worst case (counterfactual simulations with an unresponsive driver), lack the severity and criticality of real crashes—at least if no catch-up crashes are generated. It is important to note that a recent preprint of a scoping literature review on CBM-based scenario generation found almost no studies assessing the crash-outcome measure (e.g., relative speed at impact or injury risk) validity of generated crashes (Bärgman, Jokhio, et al., 2025). The findings from Paper I indicate that the type of data used have a substantial impact on safety benefit results, underscoring the need for future research to validate virtual safety assessment in general and crash generation in particular.

So far, this discussion has focused on the lack of realism of counterfactually generated crashes based directly on everyday driving data. While other methods can be used to generate crashes based on everyday driving data (for example, using traffic simulations with CBMs based on everyday driving; Feng et al., 2021; Li et al., 2019), the question remains: Can these other methods generate realistic crashes? Although this work does not address that question, the results of Paper I provide a basis for speculation. Consider Fig. 1 in Paper I, showing the distribution of the minimum acceleration of the POV and THW, when the ego vehicle is behind the POV in the same lane. The figure shows that in the great majority of cases, the POV barely applies the brakes as the two vehicles drive undisturbed along the highway. The number of harsh deceleration maneuvers detected in the highD dataset was quite low, even though it included trajectories from as many as 110,000 vehicles. If CBMs for traffic simulations are based on deceleration distributions such as those in highD, it is hard to see how the simulations could generate realistic crashes. Recall, however, that some literature suggests that EVT could help detect critical events in the tail of the crash distribution (Jiao et al., 2025).

Regardless of the method used to generate crashes, what is truly important is determining which metrics the generated crashes are valid for (and validated on; Wu, Sander, et al., 2025). In Paper I we compared the generated and real-world crashes in terms of criticality (maximum deceleration of the POV and duration of the distraction needed to generate a crash) and severity (relative speed at impact). There do not appear to be any studies using the traffic-simulation approach that validated their results with respect to crash outcome metrics such as delta-v or injury risk (Bärgman, Zhao, et al., 2025). Before traffic simulations are used to assess the crash and crash severity-reduction potential of ADASs and ADSs, they need to be validated on scenarios ranging from everyday driving, via near-crashes, to low- and high-severity crashes—not least because of the issues of using everyday driving and near-crash data for scenario generation identified in Paper I.

4.2 Safety targets in ADS development and deployment

Section 4.2.1 discusses the performance of the existing reference driver models that were analyzed in Paper II and, more generally, the role of these models in the development and safe deployment of ADSs. In Section 4.4.2, the conclusions of Paper II are used as a basis for a discussion of components which may be relevant to the development of safety targets.

4.2.1 Assessing the performance of reference driver models

When discussing reference driver models, questions that naturally arise are: *What should a reference driver model represent?* and *How should the representativity be assessed?* In this work, a reference driver model is defined as a benchmark for safe driving obtained by quantifying human driver behavior. However, this definition is far from a practical formulation of a model, which requires parameters and conditions. A computational model that represents human drivers needs to be described with mathematical equations, or, at least, statistical representations of behavior-related variables (e.g., reaction time).

Paper II analyzed the two driver models described in the R157, which is part of the approval process for ADSs in Europe. At the time of writing, these models are the only reference driver models cited in a regulation, at least in Europe. The models, CCDM and FSM, are to be used as guidance for defining preventable and unpreventable crashes in specified traffic scenarios. Because it is not clear what "guidance" means, there is a danger that implementers of ADSs will use these models as the actual target: "If our systems perform better than the models, they are good enough." This approach, as we argue in Paper II, is highly problematic.

In Paper II, the models were counterfactually applied to near-crash cut-ins from SHRP2 to evaluate their performance compared to the SHRP2 drivers. The CCDM did not manage to avoid a crash in all the near-crashes; it even generated new crashes, and generally reacted later than the SHRP2 drivers. The FSM, on the other hand, avoided crashes in all the scenarios, generally reacting considerably sooner than the SHRP2 driver, to the point that in some of the scenarios the timing was considered unrealistic and arguably unhuman-like. These results raise questions about the capability of the models to accurately capture human behavior. First, a reference driver model should (at the least) avoid all the crashes that were avoided by the type of human that the model is supposed to represent, such as a competent and careful driver. Second, the driver models should also behave in a predictable manner. A driver model that reacts to any risk of collision, even far in the future, may not realistically describe a human driver—even a competent and careful one. Actually, there is a risk that such a model would avoid crashes normally unavoidable by humans by reacting early in the scenario to some cue that a human would ignore. The model could, for example, be acting on some benign, in-lane lateral perturbations by the POV, missing the cues that actually cause reactions in human drivers (as it is already avoiding the crash). In fact, when Mattas et al. (2022) applied the FSM to highD data, they found that the model often reacted to the lateral motion of surrounding vehicles, even when they were not changing lanes. While this behavior on the one hand easily avoids most potential conflicts, it is, on the other hand, not very human-like. Because this model is part of the regulations, it could influence the way that ADSs are developed. Specifically, the FSM could set unrealistic requirements for ADSs and perhaps even impact ADS deployment; a system that is actually good enough might not be released if it is not as cautious as the model.

4.2.2 Urgency and surprise

The responses of crash avoidance driver models, such as the UNECE models, to specific traffic scenarios can be characterized by two main components. One is scenario criticality, referred to in this work as "urgency", which dictates the type (e.g., braking or steering) and intensity of the response to an unexpected event. The other component defines the start of the unexpected event, the moment in time when the model notices that something is not right, and is referred to in this work as "surprise". It is the first point in time when there is a substantial deviation of the expectations of how an event unfolds, such as the initiation of hard braking by a lead vehicle, or of a lane change that is too close. If the surprise makes the situation urgent, the driver will, if attentive, initiate an avoidance maneuver. Note, however, that just because a situation is surprising does not mean that it is urgent.

As previously mentioned, there are numerous examples of easily computed urgency metrics that have historically been used to define scenario criticality: TTC (Hayward, 1972; Hydén, 1987; Sayed et al., 1994), brake threat number (BTN; Brännström et al., 2008), deceleration rate (Allen et al., 1978; Darzentas et al., 1980), and post-encroachment time (PET; Allen et al., 1978). These metrics can be calculated using the kinematics of the involved road users. In the CCDM model, urgency is operationalized as a simple threshold for longitudinal TTC, while in the FSM there is no urgency operationalization (other than a modulation of the braking level; see description below). Defining such thresholds is not straightforward. It should be noted that although simple thresholds can be applied to the basic metrics (as in the CCDM), there are more elaborate modeling approaches that better computationally describe human behaviors such as responses to urgency. One such approach is evidence accumulation. In this approach, models integrate information over time, using a visual cue metric or the difference between an actual observation and an expectation (Xue et al., 2018; Zgonnikov et al., 2024). When enough evidence has accumulated, the reaction is initiated (Markkula, 2014; Markkula et al., 2018). A threshold is used in those models too, but it is based on accumulated evidence rather than just a single visual cue. These models have been shown to be more human-like than simple threshold-based models (Markkula, 2014; Svärd et al., 2017). It would be natural to consider this approach, or alternative criticality identification approaches (Siebinga et al., 2024), in reference driver models.

In Paper II, while urgency is identified as key in the cut-in scenario assessments of both UNECE models, they implement the concept differently. The results indicate that they are not capturing urgency in the way that human drivers experience (and act on) it. The CCDM uses TTC in the longitudinal direction to define the urgency of a scenario, while urgency in the lateral direction is not considered. That is, the model does not assess urgency-based metrics such as lateral TTC or lateral speed of the cut-in vehicle. Instead, the model assumes that a scenario is critical as soon as a vehicle leaves the wandering zone (see Section 2.2.1), which equates to the ego driver feeling comfortable as long as the POV does not leave the zone. This use of the wandering zone can be interpreted as a form of driver satisficing. Satisficing, a concept introduced by Simon (1955) and further developed by Summala (2007), refers to a decision-making strategy in naturalistic contexts: individuals aim for outcomes that are considered “good enough” rather than strictly optimal. In the case of CCDM, the ego driver does not need to optimize the vehicle’s longitudinal positioning as long as the POV stays in the wandering zone.

The FSM, on the other hand, assesses the need for a braking reaction based on the projected paths of the vehicles. If a collision is predicted, even far in the future, the FSM triggers a braking reaction. The analysis in Paper II revealed that this design approach does not adequately account for non-urgent scenarios (when a collision is not immediately predicted), since the FSM does not consider urgency in the timing. However, it uses urgency in the braking maneuver itself, since the level of deceleration is modulated based on the criticality of the scenario: slight decelerations are performed when the scenario is not very critical. This response is clearly not how drivers react to other drivers’ satisficing. That is, a driver would not perceive a car in the adjacent lane moving slightly within its wandering zone as a threat and would therefore not react to it.

The second component that characterizes the response onset of driver models in critical scenarios, surprise, also leads to an avoidance maneuver—if the scenario is urgent enough. The two UNECE models consider this component differently as well. The CCDM perceives a potentially dangerous cut-in scenario when the cut-in vehicle has left the wandering zone: it only assesses the vehicle’s lateral position using that fixed threshold and disregards other metrics, such as the POV’s lateral speed. This model’s implementation of surprise could be the cause of some of its delayed reactions (relative to the SHRP2 drivers). In contrast, the FSM bases its determination on whether the cut-in vehicle is moving laterally towards the ego vehicle on a trajectory that causes their future paths to overlap. Note that the scenario is not necessarily surprising; it does not need to be critical or unexpected for the model to start braking. For example, it would not be surprising for most drivers if a POV two lanes over moves into the adjacent lane, but the FSM may still activate and start braking the ego vehicle. The FSM is very sensitive to the lateral movements of the POV, reacting to any possible encroachment onto the future path of the ego vehicle (regardless of how far in the future it is). The result is a model that reacts rather early (albeit applying the brake pedal gradually when the scenario is not very critical). It can be argued that the FSM does not implement surprise as defined in this work.

Another human-based reference model, introduced by Engström et al. (2024), is the “non-impaired road user with their eyes on the conflict” (NIEON) model, based entirely on the concept of surprise. “Surprise” is, in their work, based on information theory (Shannon, 1948) and Bayesian surprise (Itti & Baldi, 2009), and defined as an observation that is not explained or predicted by the prior belief about how a driving scenario can unfold. This model is fundamentally different from the UNECE models, as “the prior belief is both context-dependent and determined by the road user’s prior experience” (p. 4; Engström et al., 2024). The dependency on context makes NIEON particularly promising for the issue of "when to start the clock" of the reaction time. However, this model only captures surprise, not urgency.

Paper III offers a different definition of lane-change onset which can, at least partly, address some of the shortcomings of the CCDM and FSM. Paper III uses a combination of lateral speed and lateral position to define the moment the lane changes begin. (Recall that the CCDM bases its assessment of lane change purely on the lateral position of the lane-changing vehicle, and the FSM considers only lateral motion without a minimum threshold for the lateral speed.) Paper III also studies the lane-changing process in greater detail, focusing on the factors that influence when drivers initiate lane changes. While not all results from Paper III directly relate to urgency, the information it provides about lane-change initiation can still support the development of realistic CBMs. Like the results of Paper IV, these results can support the development of models of routine driving behavior that can be used to simulate normal lane-change behavior, rather than careful and competent driving. These models can then be used as benchmarks for the assessment of careful and competent CBMs.

4.3 Glance behavior in the context of safety targets

As mentioned, drivers' glance behavior can greatly affect their perception of the surrounding traffic as well as their reaction to new and imminent threats. Clearly, glance behavior has a strong impact on safety (Horrey & Wickens, 2007; Klauer et al., 2014; Victor et al., 2014). However, the role of glance behaviors in the formulation of safety targets for ADASs and ADSs is not immediately obvious. Notably, there is a difference between safety targets for ADASs and for ADSs. For ADSs, as mentioned, one type of safety target is the comprehensive representation of a safe human driver for a specific scenario, encompassing both conflict and crash avoidance. For ADASs, the safety target can be defined as the combined performance of the driver and the system, as the driver remains responsible for the driving task when the ADAS is active. Currently, off-road glance behavior is considered to a limited extent in such safety targets. EuroNCAP, for example, has started to differentiate their scoring based on DMSs (cars should give warnings depending on whether driver is attentive or not)(Euro NCAP, 2025). However, this effort is still at an early stage, and therefore it is likely to be refined in the next iterations of the EuroNCAP protocols. This section discusses whether glance behaviors may have a role in the definition of future safety targets, and the implications of their inclusion.

The reference driver models for the safety assessment of ADSs analyzed in Paper II assume that the driver's perception and reaction times to a threat are constant. See Section 2.2.1, for details about the models (JAMA, 2022; UNECE, 2023). Glance behavior is not described in the model specifications, although perception time may implicitly include some glance-related delay. However, the delay does not clearly relate to the models' representation of a competent and careful driver. Other approaches to reference driver models have different modeling strategies regarding glance behavior. For example, the NIEON model mentioned above assumes eyes on the threat at all times, a design choice which does not leave room for reaction times that result from context-dependent glance behaviors described earlier. However, even assuming that a driver model is extremely attentive, it still has to look around to see everything, and not all types of threats can be easily detected. For example, going beyond the eyes-on-threat paradigm, in the case of sideswipes (i.e., crashes in which two vehicles crash side-by-side) the threat might come from the side or from behind the ego vehicle; it is impossible for a driver to always be aware of the risk. Olleja et al. (2022) derived the drivers' side glance frequency from an NDS to study the role of side glances in the avoidance of sideswipe crashes. The results indicate the effect that is very small, mainly due to the low frequency of side glances.

However, in line with other publications (Han et al., 2023; Seppelt et al., 2018), Papers IV and V show that glance behavior, and consequently reaction time, is not constant across all driving scenarios and criticalities. In fact, it has been shown that drivers adjust their behavior towards lower scenario criticalities if they engage in secondary tasks (Morgenstern et al., 2020; Tivesten & Dozza, 2014). Including realistic glance behavior in these models may therefore contribute to a more realistic representation of human drivers. It is important to note, however, that the glance behavior data analyzed in Papers IV and V came from drivers employed specifically to collect those data; although they were driving in a naturalistic setting, their glance behavior does not necessarily represent that of a reference "safe" driver. For this reason, glance behaviors may need to be carefully selected in order to be suitable for inclusion in reference driver models. One option is to select only the best behavior performance, based on a defined percentile threshold; short glances by the most attentive drivers could approximate the profile of a safe driver. Another option is to include only safety-related glances (such as those directed toward mirrors or blind spots) which contribute to situational awareness and hazard detection. This approach would allow reference models to include realistic glance behavior while maintaining a conservative representation of human performance.

ADASs and ADSs have different definitions for safety targets, as only the ADASs' safety performance is directly linked to and dependent on the driver's behavior. Actually, there are also several types of safety targets for ADASs. As an example, Paper IV uses a method that links glance behavior to the risk of crashing as a function of time gap. The reference behavior can thus be defined by the threshold of crash risk in terms of glance behavior (in this work, defined as the MCR): if the glance behavior of the driver exceeds the MCR threshold, the system triggers. This method should also account for context (e.g., time gap), and it can include the risk associated with the glance behavior over some time window. There may be more elaborate ways of considering the risk of individual glances and glances over some time window, but they are beyond the scope of this paper.

An additional factor that possibly should be considered when defining safety targets for ADASs and ADSs is the influence of cognitive load on driver responses. Extensive research has demonstrated that cognitive load influences glance behavior (Engström et al., 2005; Lee et al., 2007). However, the literature on the effects of cognitive load on real-world traffic safety is less consistent, as analyses of NDSs have repeatedly shown that cognitive load has only a limited impact on actual crash risk (Klauer et al., 2014; Victor et al., 2014). For ADAS safety targets, it may still be appropriate to consider including a cognitive-load component. However, cognitive load is notoriously difficult to measure in naturalistic driving, which makes adding this component to ADAS functionality challenging in practice.

For ADS reference-driver models, on the other hand, it is not clear that cognitive load should play a role—at least not if the model is intended to represent a careful and competent driver. In this case, assuming a negligible impact of cognitive load is likely reasonable. Still, future research could assess and quantify whether cognitive load should be considered part of careful and competent driving, and whether its influence is substantial enough also for a careful and competent driver to justify the effort required to model and integrate cognitive load into reference driver models.

Paper IV also studies the single longest off-road glances (see Section 2.2.3 for a more detailed description), which can represent the riskiest behavior observed in each driving segment. Setting limits at the tail of glance duration distributions may be another way to define reference behavior. Using a very similar approach, current guidelines such as the NHTSA (2012) Distraction Guidelines put simple constraints on single off-road glance durations. Linking the behavior of the driver (here the glance behavior) to its objective impact on safety is a concept that may translate into other crash scenarios and other ADASs.

Paper V advances the existing body of knowledge by showing that drivers' off-road glance behavior is dependent on context (specifically looming). The study observed that off-road glances from the dataset were seldom initiated at inv-TTC values above approximately 0.2–0.3 s^{-1}. The results therefore motivate the inclusion of context dependent metrics in safety targets based on human behavior.

4.4 Other considerations when developing safety targets

As mentioned, not only do safety targets for ADSs and ADASs differ, but they can be formulated in various ways. Section 4.4.1 discusses various strategies used to define safety targets. Section 4.4.2 further discusses the inclusion of comfort zone boundaries (CZBs) in driver behavior analyses as a way to operationalize safety targets both for ADASs and ADSs.

4.4.1 What should safety targets be based on?

The following discusses some of the ways safety targets can be defined, and the implications of the different definitions for safety assessment, traffic safety, and system acceptance (by drivers and other road users). The discussion is separated into ADS and ADAS safety targets.

ADS safety targets

The category of ADS safety targets can be categorized in many ways. For this discussion, they are defined based on:

a) driver behavior
b) societal expectations of what drivers should do
c) driver comfort-zone boundaries
d) objective safety

a) ADS safety targets based on driver behavior rely on developers acquiring data from some population of drivers and using the data to operationalize the safety targets. To develop a target that is genuinely linked to safety, developers must determine how to select and combine behavioral components in a way that can be justified as safety-related and that represents the level of safety the modeled behavior is intended to convey. However, there is no consensus on what approach is more appropriate; it may be that a one-size-fits-all solution does not exist. One of the current UNECE performance models, the CCDM, is an example of a behavior-based reference driver model which is intended to represent a "competent and careful" driver (UNECE, 2023). However, it is unclear how the measured data were turned into a safety target, as they appear to be based on a small sample of drivers in a driving simulator in Japan, with no direct link to drivers' competence and carefulness. The implications of this approach on the ADS safety evaluation have been shown in Paper II: the CCDM did not perform as carefully as drivers from SHRP2, and therefore its validity is in question. One option for developing reference models which more closely reflect the characteristics of a competent and careful driver is to instruct the participants in the study to drive carefully. Of course, this approach would have to be assessed before it could actually be used for model development. Another option is to simply collect driver behavior data, and define what constitutes an appropriate representation of the safety target based on those data. In Olleja et al. (2023) we proposed selecting a percentile of the behavior (e.g., 80th percentile) as a reasonably safe driver. However, any such choice becomes problematic if the safety targets include multiple safety-related parameters (e.g., TTC at overtaking and lateral clearance, as in Rasch & Dozza, 2022). The combination of criteria may produce overly strict safety targets; that is, it may be unrealistically hard to comply with them all. Alternatively, all individual safety targets would need to be relaxed enough to allow for an easier "pass". See Wu, Sander, et al. (2025) for a description of this issue in the context of scenario generation validation.

Another approach to defining safety targets in terms of driver behavior involves selecting safety targets that approximate the limits of what drivers can reasonably achieve in best-case conditions. This may include *how threats are identified* (e.g., visual cues combined with urgency and surprise), *when* threats are identified (e.g., assuming eyes-on-threat or incorporating glance behavior), *how quickly* drivers can react (e.g., reaction time), and *how* they react (e.g., brake amplitude). Depending on how these elements are operationalized, safety targets defined in this manner may include superhuman components—such as assuming eyes-on-threat—which are not feasible in real life. One example of such a superhuman model is the NIEON model which, as noted, assumes eyes-on-threat at all times. A potential issue with these safety targets is that they may delay the introduction of systems that do not reach the superhuman benchmark, even if they achieve a PRB and would otherwise be considered safe enough for deployment.

In summary, this type of safety target can be defined by selecting a driver (or a distribution of drivers) from anywhere along the spectrum of driver behavior, from the "average driver" to a "superhuman". The implications for ADS assessment span a similarly wide range, from enabling the introduction of ADSs that may not yet be sufficiently safe to impeding the deployment of capable ADSs due to overly stringent or unrealistic targets.

b) ADS safety targets based on societal expectations of what drivers should do attempt to convert the vague concept of societal expectations into safety targets (Fraade-Blanar et al., 2026). One possible, although to my knowledge unrealized, operationalization would be for driving-school teachers to perform driving maneuvers the way optimal drivers do them (i.e., what they teach), and then base the safety targets on those behaviors.

c) ADS safety targets based on comfort-zone boundaries use the limits of what drivers find acceptable (e.g., comfortable braking/steering) to define urgency/thresholds. See Section 4.4.1 for a discussion of this approach.

d) ADS safety targets based on objective safety already exist. The most prominent is the RSS, introduced in Section 2.2.1. An issue that has repeatedly been acknowledged in the literature is that these safety targets tend to be very cautious, even when compared to careful and competent human drivers. For example, Liu et al. (2021) describe the original parameter selection of the RSS as overly conservative.

There are many situations in traffic where it is necessary to drive very slowly if a no-crash paradigm is to be enforced. Consider driving on a narrow street with parked cars on each side: the vehicle would have to drive very slowly to be able to stop in time if someone were to come out between the cars suddenly (e.g., a child). This dilemma has been acknowledged extensively in the literature (Almaskati et al., 2024; Sun et al., 2024). This issue will only be resolved when it is agreed what is to be considered safe enough—with respect to speed and other driving parameters, as well as to the response times and amplitudes used for safety targets.

Vision Zero focuses on the avoidance of severe injuries and fatalities and pushes for a resilient safe system approach (see also the Introduction). Within this strategy it will not be possible to define realistic safety targets that allow for reasonable mobility if the targets are defined solely in terms of crash avoidance; these restricted definitions may cause the safety targets to become overly conservative, effectively pushing for near-zero crash probability in all situations. In many everyday traffic environments, thresholds would become so strict that they would constrain normal driving (e.g., calling for very low speeds or very early interventions), thereby compromising mobility and usability, and potentially undermining acceptance and real-world safety impact of the systems.

ADAS safety targets

ADAS safety targets must balance risk reduction and acceptance, maybe even more than those of ADSs do; if ADAS targets lead to frequent/early (cry-wolf) warnings, the driver may disable/ignore them. In this work, the safety targets are defined based on:

a) driver behavior
b) objective safety

Note that unlike ADS safety target definition approaches, the approaches for ADASs do not include societal expectations for routine driving (related to the concept of drivership), as the focus of this work is on crash avoidance, rather than conflict avoidance. However, societal expectations are still relevant for those ADAS safety targets related to conflict avoidance (e.g., ACC).

a) ADAS safety targets based on driver behavior can use behavioral limits similar to those used by ADSs, while incorporating additional considerations for driver acceptance and nuisance. One example is a model of reference driving behavior which represents only the "safest" driving behavior from a distribution (e.g., choosing values above a certain percentile on some safety-relevant metric) and sets thresholds that balance safety with other aspects of the driving experience, such as annoyance.

Given the current push toward DMS-based IISs, defining safety target strategies for regulations and consumer testing (e.g., Euro NCAP) is becoming increasingly urgent. Papers IV and V highlight the importance of considering context in IISs, since they show that drivers do adjust their behavior based on contextual factors (e.g., in relation to TTC, though far less with respect to time gap). These findings highlight the need for context-aware safety-target definitions, rather than uniform thresholds that ignore situational variation.

b) ADAS safety targets based on objective safety link behavior to objective outcomes (e.g., crash or injury risk). Defining objective safety targets for crash avoidance in ADASs is in many respects similar to defining them for ADSs: an objective determination must be made regarding which outcomes are acceptable when a crash-relevant situation is developing. One strategy is to set targets aimed at avoiding crashes altogether. However, for IIS-based ADASs this could lead to very strict glance-behavior limits, particularly if they are defined without context. Current guidelines, such as the NHTSA distraction guidelines (NHTSA, 2012) are not based on avoiding crashes at all costs, but rather on acceptable risk (e.g., secondary tasks such as radio tuning). The current guidelines and regulations are not context-dependent, although the context dependency demonstrated for TTC in Paper V indicates that they should be (as should ADAS safety targets). Further, the lack of (adequate) driver adaptation to decreasing time gaps (corroborating findings from Tijerina, 1999) illustrates why formulating purely objective, safety-based IIS thresholds is problematic. Drivers apparently do consider it safe enough to look away even with relatively short time headway, and may not accept warnings purely based on objective safety. More research is needed to understand why, and under which circumstances they would accept IIS warnings, thus avoiding the "cry wolf" effect which could lead to attempts to disable the system.

Further, safety-target definitions for ADASs may need to move beyond focusing solely on crash avoidance. A more viable path may be to define targets based on a combination of crash avoidance and injury mitigation, aligning the framework more closely with the principles of Vision Zero, which emphasizes injury avoidance rather than absolute crash elimination.

4.4.2 *Comfort zone boundaries vs. satisficing*

The development of safety targets from measured human behavior is, as mentioned in the previous section, an approach currently used in regulations for the approval of ADSs. However, the operationalization of these safety targets can suffer from the effects of satisficing behavior by the human drivers whose behavior is being measured (e.g., in the form of reaction times, lane keeping behavior, or glance behavior). The recorded behavior may not be optimal, but instead represents what was “good enough” at the time when it was recorded. The effect of collecting data that may include drivers satisficing is that it is not possible to define the limit of the urgency of the driving scenario which drivers can comfortably withstand, unless experiments and data collections are specifically aimed at minimizing satisficing behavior.

One promising approach to defining thresholds for urgency metrics uses the concept of CZBs. The concept was introduced by Näätänen and Summala (1974), whose study was based on the theory of “field of safe travel” proposed by Gibson and Crooks (1938). The boundaries define maneuvers that drivers feel comfortable doing: for example, braking to avoid a collision with an acceptable level of deceleration, or steering to avoid a collision with an acceptable level of lateral acceleration (Brännström et al., 2010, 2014; Sander, 2017; Yang et al., 2024). The CZBs, limits of what the driver considers acceptable without feeling discomfort, can be used to define thresholds in engineering-based metrics. For example, if a collision in a rear-end scenario can be avoided with a slight deceleration—below the limit of comfortable braking—the scenario is not considered urgent. In both Paper I and the work by Yang et al. (2024), CZBs were used to tune and assess AEB systems. In Paper I, CZBs in the AEB algorithm determined the urgency of a rear-end scenario by assessing the possibility that the driver could avoid the collision with comfortable steering. Yang et al. (2024) dug deeper into how CZBs can be used in AEB algorithms. The CZBs in both papers were partially derived from Brännström et al. (2014), who define 5 m/s^2 as the comfortable limit for deceleration in their CZB-based AEB algorithm, although it is not stated how the threshold was selected. It is worth noting that this level of deceleration is an unusually harsh braking maneuver: Fitch et al. (2010) found that the mean deceleration reached by nonprofessional drivers during emergency braking maneuvers was close to 5 m/s^2—and those are clearly not comfortable scenarios. Fig. 5a in Paper I shows that in about 50% of the SHRP2 events (near-crashes) the POV exceeded this rate, indicating that as the ego vehicle was approaching at a high speed and with a relatively low THW, its driver probably had to exceed 5 m/s^2 of deceleration to avoid a collision in many of the events. Although the deceleration values for the ego vehicle are not reported in the results of Paper I, the harsh braking maneuvers by the POV and the relatively low THW highlight the urgency of the SHRP2 near-crashes. On the other hand, the scenarios extracted from the highD database, which mostly contains everyday driving data, were generally not urgent at all: the POV rarely reached a deceleration of 5 m/s^2, and the THW (Fig. 1 of Paper I) was rarely short enough to be considered unsafe. In previous research, in which I explored defining safety targets based on CZBs for cars overtaking cyclists (Olleja et al., 2023), I proposed using the 80th percentile of metrics such as lateral distance and TTC in observed overtaking behavior, represented through a parametric model (Rasch & Dozza, 2022). However, the 80th percentile of that behavior is not equivalent to the 80th percentile of the CZB, which may impact the acceptance of safety targets based solely on measured behavior. Further research is needed to examine the relationship between satisficing, CZBs, and acceptance. It should also be noted that Jiao et al. (2025) describe a survival analysis-based approach for identifying the extreme edge of proximities in road-user interactions. This method warrants deeper investigation, particularly regarding how these edges relate to drivers' perceived safety and comfort zone boundaries.

Urgency, in Paper I, can be discussed not only in terms of AEB tuning, but also from the perspective of crash generation. The clear relation between the nature of the datasets analyzed and their intrinsic urgency, from both closeness-to-crash (i.e., engineering), and human (i.e., CZBs) perspectives, has already been demonstrated. All crashes are urgent by definition: at some point in time the driving scenario becomes critical, and the driver is not able to avoid a collision even by braking well beyond the levels of comfortable crash avoidance.

4.5 Limitations and future work

This section explains the main limitations of the studies in this work as well as related suggestions for future work. The first focus is on the limitations intrinsic to the data used. Then, limitations about the methods used are described.

4.5.1 Data limitations and suggestions for future work

A main limitation of the analysis in Paper I consisted of difficulties in obtaining scenarios whose characteristics could be matched across the datasets. For example, highD and GIDAS both originate from German highways, while SHRP2 data were collected in the USA. Moreover, unlike the near-crashes in highD and GIDAS, those in SHRP2 rarely occurred on highways, resulting in substantially lower vehicle speeds. Future work could focus on analyzing datasets with scenarios that more closely match each other. Additionally, the analysis in Paper I was limited to the rear-end scenario, primarily because the AEB functions applied to the crashes were designed to work in those conditions. Future studies should focus on other types of scenarios, such as cut-ins and cut-outs, to extend (or overturn) the results obtained in Paper I for rear-end scenarios.

Paper II also focused only on one type of scenario: cut-ins. The choice was made to assess the UNECE models for that particular scenario, as it is much more demanding from a modeling perspective than pure rear-end crashes. Further, poor data quality and issues with the annotations limited the number of usable cases extracted. Specifically, the poor quality of some SHRP2 videos (related to resolution, light conditions, blooming, and compression) at times made it difficult to determine the exact position of the POV. When that was the case, additional annotation points were added to smooth inaccuracies. Further, the signal from the radar was of low quality; its use was limited to the cases in which it was reliable enough to refine the longitudinal position of the POV. In fact, the radar was unable to capture the lateral movement needed for the assessment (the field of view was too narrow). Future work could further improve the video annotation tools (e.g., increase the use of automatic image processing) to increase the quantity and quality of data or use new, better datasets. It is unlikely that another study with the sheer scale of SHRP2 will be conducted again soon, but maybe data from event data recorders with video can be used in the future (Piccinini et al., 2017). Further, as the reference driver models described in R157 are intended to work not only for cut-in scenarios, but also for deceleration (rear-end) and cut-out scenarios, future studies could apply methods similar to the one in Paper II to those traffic scenarios.

Papers III, IV, and V used data from realistic on-road data collections; the differences between these data and naturalistic data need to be accounted for when interpreting the results, since they might result in a driving style that is not purely naturalistic. The main differences usually found in these collections are that the population of drivers consists of employees of a company and that they have been instructed to drive on potentially designated routes. First, the limited population of drivers affect the generalizability of the results. For example, there are probably few young or elderly drivers among the people explicitly hired to collect data. Second, the style of driving is not entirely naturalistic, which may also limit the generalizability of the results (and the validity of the conclusions drawn in the studies that use them).

4.5.2 Methodological limitations and suggestions for future work

The studies in this thesis also have limitations due to the methods used. The counterfactual crash-generation process in Paper I relies on the assumption that the ego driver is completely unresponsive to the braking maneuver of the POV (basically, the driver is sleeping). However, it is not realistic to assume that all drivers are sleeping, unless the object is to study the worst-case scenario of human driving behavior (Kusano & Victor, 2022). Adding more crash-causation components, such as distributions of glance behavior and limited brake responses, to the driver model (Bärgman et al., 2017; Bärgman et al., 2015; Lee et al., 2018; Morando et al., 2019) would improve the validity of the crashes generated (Bärgman, Jokhio, et al., 2025). However, Paper I nonetheless demonstrates that, *even in the worst-case scenario* of unresponsive drivers, the lower-severity datasets do not show the high level of criticality and severity seen in the crash datasets. This difference is partially due to the more benign scenario, but it is probably due to a large extent to the censoring of low-severity crashes in the crash datasets: recall that in GIDAS, PDO crashes are completely missing and low-severity injuries are underreported. Future studies should investigate the contribution of benign scenarios (e.g., a slight lateral movement within a lane, without the intent to change lanes) and of the censoring (e.g., the effect of censoring the data for both modeling and validation) to the difference in urgency and outcome severity (Morando, 2025). One option would be to apply a transformation to the generated crashes that accounts for GIDAS' data censoring—making the crashes comparable with the other datasets (Kusano et al., 2024).

In Paper II, the assumption that the POV's behavior was the only factor influencing the ego vehicle's braking evasive maneuver was a main methodological limitation. Any possible effect due to other road users or external factors was disregarded. However, manual verification indicates that this assumption is well founded and the ego vehicle's reaction was probably caused by the POV's behavior; usually there were few/no vehicles in the immediate vicinity. However, it was still not possible, from the data available, to be certain that there were no other factors which influenced the ego driver behavior.

In Paper III, the selection of the lateral speed thresholds for the identification of a lane-change maneuver disregarded the possible impact of the driver's motivation for the lane change. That is, the lateral speed of a vehicle performing a lane change may differ for mandatory and non-mandatory lane changes (Vechione et al., 2018). Mandatory lane changes may consist of quicker lateral movements, which in turn could affect the selection process for the lateral speed threshold to detect cut-ins.

In summary, future work should focus on continuing the effort towards relevant and valid safety benefit assessments of ADSs and ADASs. In particular, the work towards the formulation of more widely accepted forms of valid safety targets should continue.

4.6 Ethics and sustainability

This work has clear implications for both ethics and sustainability. First, the overall goal of improving traffic safety by reducing deaths and serious injuries is the main ethical motivation. The way safety is assessed (and what data and models are used) can influence decisions about development, approval, and deployment of safety systems. This is an ethical issue, because it can lead to decisions that either delay beneficial safety technology or implement technology that is not beneficial. In particular, if safety targets or simulation baselines are not representative (for example, due to data selection effects or model assumptions), then the assessment may systematically overestimate or underestimate safety benefits.

Ensuring safety is an ethical duty for system designers, regulators, and industry, especially when it involves making choices about what scenarios and what outcomes to prioritize. When resources are limited, safety decisions can end up benefiting some situations more than others, for example with assessments focusing mainly on certain crash types, severities, or driving contexts. Therefore, transparency about scope, limitations, and biases is important. Furthermore, as mobility is a fundamental part of people's lives, safety targets should not require unrealistic driving that would severely reduce usability and access to mobility.

From a sustainability perspective, improved traffic safety reduces short- and long-term impacts on people as well as reducing societal and resource costs related to injuries requiring healthcare, including recovery. In the longer term, ADSs may also support more efficient transport (for example through improved traffic flow and shared mobility services), which could reduce the need for private car ownership and encourage alternative, more sustainable transportation.

5 Conclusions

The research presented in this thesis contributes to the field of traffic safety by supporting the development of safety targets for ADSs and ADASs. For example, this work reveals how different safety-target design choices can impact the validity and development of virtual safety assessments. After identifying limitations in current safety targets and their evaluation methods, the thesis has proposed improvements on three fronts. First, given the importance of using relevant, valid practices, the thesis highlights the challenges and risks of using only naturalistic driving data to assess reference driver models intended as safety targets. Second, the thesis demonstrates the importance of including urgency in lane-change detection by reference driver models. Third, the thesis studies context-dependent glance behavior and discusses its use in the formulation of safety targets.

This work identifies the importance of the choice of data for virtual safety assessments, highlighting the substantial risks of generating baseline crash scenarios solely from non-crash naturalistic driving data. The findings indicate that crashes artificially generated from everyday driving data (and to a lesser extent from near-crashes) differ substantially from real-world crashes, resulting in reduced criticality and lower severity outcomes. Even when worst-case scenarios are generated, using near-crashes and assuming an unresponsive driver, the resulting baselines often lack the high level of crash severity found in in-depth crash databases. Because these generated scenarios are inherently less critical, safety systems (such as AEB) avoid them more easily than they would real crashes, which can lead to over-estimation of a system's safety benefits. Similarly, this thesis concludes that reference driver models need to be assessed on data which reflect the conditions which they are designed to operate in. If the data do not capture the criticality of the real-world conflicts that the models are expected to handle, the validity of the reference models cannot be confidently assured. Consequently, the safety benefit assessment of an ADS may be compromised if it relies on reference behavior models that have not been adequately validated.

The evaluation of two current models in the regulations indicates that they do not accurately represent competent and careful drivers. When tested against real-world near-crash cut-in scenarios, one model reacted too late, leading to crashes that the human drivers had avoided, while the other was overly cautious; neither reacted like the human drivers they were compared to. Consequently, this work highlights the importance of validating reference driver models on a variety of data types, including critical events. The thesis also proposes the incorporation of definitions of lane-change initiation that account for lateral speed (or some other lateral urgency metric) to improve the realism of the models' responses to cut-ins.

Finally, this thesis clearly demonstrates the importance of context in defining safety targets for ADSs and ADASs, assessing to what extent drivers adapt their glance behavior as a function of context. For ADSs, this work argues that reference driver models may become more realistic and human-like if their responsiveness to safety-critical scenarios is influenced by glance behavior (for example, by shifting from a fixed response time to one influenced by the probability of the driver glancing away from the forward road). For ADASs, this work advocates not only for the use of direct glance-behavior safety targets, as is common today, but also for the incorporation of safety targets based on crash risk derived from off-road glances. With this information, the safety target can be represented, at least in part, by a context-dependent threshold on glance-related crash risk derived from reference glance behavior. Objective crash risk derived from observed glance behavior grounds the safety targets in terms of behavior and objective safety, thus supporting the development of IISs that minimize nuisance while ensuring safety.

Overall, while many challenges in virtual safety assessment methods and in the development of safety targets remain, this work lays the groundwork for more precise and robust safety targets and safety assessment methods. This will likely lead to more safe, robust, and accepted ADASs and ADSs, which in turn will save more lives on our roads.

References

Ahlström, C., Kircher, K., Nyström, M., & Wolfe, B. (2021). Eye tracking in driver attention research—How gaze data interpretations influence what we learn. *Frontiers in Neuroergonomics*, *2*, Article 778043. https://doi.org/10.3389/fnrgo.2021.778043

Allen, B. L., Shin, B. T., & Cooper, P. J. (1978). Analysis of traffic conflicts and collisions. *Transportation Research Record*, *667*, 67–74. http://onlinepubs.trb.org/Onlinepubs/trr/1978/667/667-009.pdf

Almaskati, D., Kermanshachi, S., & Pamidimukkala, A. (2024). Investigating the impacts of autonomous vehicles on crash severity and traffic safety. *Frontiers in Built Environment*, *10*, Article 1383144. https://doi.org/10.3389/fbuil.2024.1383144

Alvarez, S., Page, Y., Sander, U., Fahrenkrog, F., Helmer, T., Jung, O., Hermitte, T., Düering, M., Döering, S., & Op den Camp, O. (2017). Prospective effectiveness assessment of ADAS and active safety systems via virtual simulation: A review of the current practices. In *Proceedings of the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)* (Paper Number 17-0346). National Highway Traffic Safety Administration.

Antony, M. M., & Whenish, R. (2021). Advanced driver assistance systems (ADAS). In M. Kathiresh & R. Neelaveni (Eds.), *Automotive embedded systems: Key technologies, innovations, and applications* (pp. 165–181). Springer. https://doi.org/10.1007/978-3-030-59897-6_9

Bakker, J., Jeppsson, H., Hannawald, L., Spitzhüttl, F., Longton, A., & Tomasch, E. (2017). IGLAD – International harmonized in-depth accident data. In *Proceedings of the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)* (Paper Number 17-0248). National Highway Traffic Safety Administration. https://www-nrd.nhtsa.dot.gov/pdf/ESV/Proceedings/25/25ESV-000248.pdf

Bjorvatn, A., Page, Y., Fahrenkrog, F., Weber, H., Aittoniemi, E., Heum, P., Lehtonen, E., Silla, A., Bärgman, J., Borrack, M., Innamaa, S., Itkonen, T., Malin, F., Pedersen, K., Schuldes, M., Sintonen, H., Streubel, T., Hagleitner, W., Hermitte, T., … Torrao, G. (2021). *Impact evaluation results* (Deliverable D7.4). https://l3pilot.eu/fileadmin/user_upload/Downloads/Deliverables/Update_14102021/L3Pilot-SP7-D7.4-Impact_Evaluation_Results-v1.0-for_website.pdf

Blatt, A., Pierowicz, J. A., Flanigan, M., Lin, P.-S., Kourtellis, A., Lee, C., Jovanis, P., Jenness, J., Wilaby, M., Campbell, J., Richard, C., Good, D., Czar, N., & Hoover, M. (2015). *Naturalistic driving study: Field data collection* (SHRP 2 Report S2-S07-RW-1). The National Academies Press. https://doi.org/10.17226/22367

Bock, J., Krajewski, R., Moers, T., Runde, S., Vater, L., & Eckstein, L. (2020). The inD dataset: A drone dataset of naturalistic road user trajectories at German intersections. In *2020 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1929–1934). IEEE. https://doi.org/10.1109/IV47402.2020.9304839

Boda, C.-N., Lehtonen, E., & Dozza, M. (2020). A computational driver model to predict driver control at unsignalised intersections. *IEEE Access*, *8*, 104619–104631. https://doi.org/10.1109/ACCESS.2020.2999851

Brännström, M., Coelingh, E., & Sjöberg, J. (2010). Model-based threat assessment for avoiding arbitrary vehicle collisions. *IEEE Transactions on Intelligent Transportation Systems*, *11*(3), 658–669. https://doi.org/10.1109/TITS.2010.2048314

References

Brännström, M., Coelingh, E., & Sjöberg, J. (2014). Decision-making on when to brake and when to steer to avoid a collision. *International Journal of Vehicle Safety*, *7*(1), 87–106. https://doi.org/10.1504/IJVS.2014.058243

Brännström, M., Sjöberg, J., & Coelingh, E. (2008). A situation and threat assessment algorithm for a rear-end collision avoidance system. In *2008 IEEE Intelligent Vehicles Symposium* (pp. 102–107). IEEE. https://doi.org/10.1109/IVS.2008.4621250

Bundesministerium der Justiz und für Verbraucherschutz. (2024). *Straßenverkehrs-Ordnung (StVO) [Road traffic regulations]*. https://www.gesetze-im-internet.de/stvo_2013/BJNR036710013.html

Bärgman, J. (2016). *Methods for analysis of naturalistic driving data in driver behavior research* [Doctoral dissertation, Chalmers University of Technology]. http://publications.lib.chalmers.se/records/fulltext/244575/244575.pdf

Bärgman, J., Boda, C. N., & Dozza, M. (2017). Counterfactual simulations applied to SHRP2 crashes: The effect of driver behavior models on safety benefit estimations of intelligent safety systems. *Accident Analysis & Prevention*, *102*, 165–180. https://doi.org/10.1016/j.aap.2017.03.003

Bärgman, J., Jokhio, S., Svärd, M., Fries, A., Denk, F., Paliotto, A., de Oliveira, L. F. A., Beckmann, J. N. B., Adjenughwure, K., Davidse, R. J., de Zwart, R., Fahrenkrog, F., Hammouda, M., & Olleja, P. (2025). *On human behavior models for virtual safety impact assessment: A scoping review*. Zenodo. https://doi.org/10.5281/zenodo.17253866

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. https://doi.org/10.1016/j.trf.2015.10.011

Bärgman, J., Svärd, M., Lundell, S., & Hartelius, E. (2024). Methodological challenges of scenario generation validation: A rear-end crash-causation model for virtual safety assessment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *104*, 374–410. https://doi.org/10.1016/j.trf.2024.04.007

Bärgman, J., Zhao, M., & Flannagan, C. (2025). Challenges and recommendations using surrogates for safety impact assessment: Correcting misestimations. In *Proceedings of the 2025 IEEE International Automated Vehicle Validation Conference (IAVVC)*. IEEE. https://doi.org/10.1109/IAVVC61942.2025.11219617

California Department of Motor Vehicles. (2025). *California driver's handbook*. https://www.dmv.ca.gov/portal/file/california-driver-handbook-pdf/

Chen, Q., & Dai, B. (2018). *Characteristics and casualty analysis of two-wheeler accidents in China, data source: The China in-depth accident study (CIDAS)* (SAE Technical Paper 2018-01-1052). SAE International. https://doi.org/10.4271/2018-01-1052

Chen, T., Wong, Y. D., Shi, X., & Wang, X. (2022). Optimized structure learning of Bayesian network for investigating causation of vehicles' on-road crashes. *Reliability Engineering & System Safety*, *224*, Article 108527. https://doi.org/10.1016/j.ress.2022.108527

Conference of European Directors of Roads. (2009). *Safe distance between vehicles*. https://www.cedr.eu/docs/view/60794fa6cf0c0-en

Cuerden, R., & McCarthy, M. (2016). *The methodology and initial findings for the Road Accident In Depth Studies (RAIDS) Programme* (Report No. PPR808). Transport Research Laboratory. https://trl.co.uk/uploads/trl/documents/PPR808%20-

%20The%20methodology%20and%20initital%20findings%20for%20the%20RAIDS%20programme.pdf

Darzentas, J., Cooper, D. F., Storr, P., & McDowell, M. (1980). Simulation of road traffic conflicts at T-junctions. *Simulation*, *34*(5), 155–164. https://doi.org/10.1177/003754978003400505

Das, A., Khan, M. N., & Ahmed, M. M. (2023). Deep learning approach for detecting lane change maneuvers using SHRP2 naturalistic driving data. *Transportation Research Record*, *2677*(1), 907–928. https://doi.org/10.1177/03611981221103229

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. https://doi.org/10.1016/j.aap.2011.05.001

Di Fabio, U., Broy, M., Brüngger, R. J., Eichhorn, U., Grunwald, A., Heckmann, D., Hilgendorf, E., Kagermann, H., Losinger, A., Lutz-Bachmann, M., Lütge, C., Markl, A., Müller, K., & Nehm, K. (2017). *Ethics commission: Automated and connected driving*. Federal Ministry of Transport and Digital Infrastructure. https://bmdv.bund.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., & Knipling, R. R. (2006). *The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car field experiment* (Report No. DOT HS 810 593). National Highway Traffic Safety Administration. https://rosap.ntl.bts.gov/view/dot/37370

Ehsani, J. P., Harbluk, J. L., Bärgman, J., Williamson, A., Michael, J. P., Grzebieta, R., Olivier, J., Eusebio, J., Charlton, J., & Koppel, S. (2021). Naturalistic driving studies: An overview and international perspective. In R. Vickerman (Ed.), *International encyclopedia of transportation* (Vol. 7, pp. 20–38). Elsevier. https://doi.org/10.1016/B978-0-08-102671-7.10651-7

Elvik, R., & Mysen, A. (1999). Incomplete accident reporting: Meta-analysis of studies made in 13 countries. *Transportation Research Record*, *1665*(1), 133–140. https://doi.org/10.3141/1665-18

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

Engström, J., Liu, S. Y., Dinparastdjadid, A., & Simoiu, C. (2024). Modeling road user response timing in naturalistic traffic conflicts: A surprise-based framework. *Accident Analysis & Prevention*, *198*, Article 107460. https://doi.org/10.1016/j.aap.2024.107460

Engström, J., Markkula, G., Xue, Q., & Merat, N. (2018). Simulating the effect of cognitive load on braking responses in lead vehicle braking scenarios. *IET Intelligent Transport Systems*, *12*(6), 427–433. https://doi.org/10.1049/iet-its.2017.0233

Erbsmehl, C. (2009). Simulation of real crashes as a method for estimating the potential benefits of advanced safety technologies. In *Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles (ESV)* (Paper Number 09-0162). National Highway Traffic Safety Administration. https://www-nrd.nhtsa.dot.gov/pdf/ESV/Proceedings/21/09-0162.pdf

Erdmann, J. (2015). SUMO's Lane-Changing Model. In M. Behrisch & M. Weber (Eds.), *Modeling Mobility with Open Data* (pp. 105–123). https://doi.org/10.1007/978-3-319-15024-6_7

Euro NCAP. (2023). *Euro NCAP rating review 2023*. https://www.euroncap.com/media/79890/euro-ncap-rating-review-2023.pdf

Euro NCAP. (2025). *Safe driving: Driver engagement* (Version 1.1). https://www.euroncap.com/media/91704/euro-ncap-protocol-safe-driving-driver-engagement-v11.pdf

European Commission. (2020). *Ethics of connected and automated vehicles: Recommendations on road safety, privacy, fairness, explainability and responsibility*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/035239

Fahrenkrog, F., Das, A., Sander, D., Bärgman, J., Urban, M., Pohle, M., Glasmacher, C., Menzel, T., Chajmowicz, H., Davidse, R., Denk, F., Op den Camp, O., Charoniti, E., Kalisvaart, S., Lorente Mallada, J., La Torre, F., Paliotto, A., Meocci, M., Wimmer, P., … Wijnen, W. (2024). *Prospective safety assessment framework – Instruction* (Deliverable D2.1). Horizon Europe project V4SAFETY. https://v4safetyproject.eu/media/pages/outputs/b52fa3f678-1730289836/deliverable_d2_1_framework_instruction_v10-r.pdf

Fahrenkrog, F., Op den Camp, O., Sander, D., Bärgman, J., Menzel, T., Glasmacher, C., Chajmowicz, H., Denk, F., Lorente Mallada, J., La Torre, F., Wimmer, P., Sander, U., Wijnen, W., & Metzger, J. (2025). *Prospective safety assessment framework – Method* (Deliverable D2.2). Horizon Europe project V4SAFETY. https://v4safetyproject.eu/media/pages/outputs/9de0a5cdea-1755592956/d2-2_prospective_safety_assessment_framework_method_v4safety.pdf

Favaro, F. M., Fraade-Blanar, L., Schnelle, S., Victor, T., Peña, M., Engström, J., Scanlon, J., Kusano, K., & Smith, D. (2026). Building a credible case for safety: Approach proposal for automated driving systems. *Journal of Safety Research*, *96*, 181–196. https://doi.org/10.1016/j.jsr.2025.10.019

Feng, S., Yan, X., Sun, H., Feng, Y., & Liu, H. X. (2021). Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment. *Nature Communications*, *12*(1), Article 748. https://doi.org/10.1038/s41467-021-21007-8

Fitch, G. M., Blanco, M., Morgan, J. F., & Wharton, A. E. (2010). Driver braking performance to surprise and expected events. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(24), 2075–2080. https://doi.org/10.1177/154193121005402412

Fraade-Blanar, L., Favaro, F., Engstrom, J., Cefkin, M., Best, R., Lee, J., & Victor, T. (2026). Being good (at driving): Characterizing behavioral expectations on automated and human driven vehicles. In G. Meyer & S. Beiker (Eds.), *Road Vehicle Automation 12* (pp. 81–105). Springer. https://doi.org/10.1007/978-3-032-12374-9_6

Fries, A., Lemberg, L., Fahrenkrog, F., Mai, M., & Das, A. (2023). Modeling driver behavior in critical traffic scenarios for the safety assessment of automated driving. *Traffic Injury Prevention*, *24*(Suppl. 1), S105–S110. https://doi.org/10.1080/15389588.2023.2211187

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

Gettman, D., & Head, L. (2003). Surrogate safety measures from traffic simulation models. *Transportation Research Record*, *1840*(1), 104–115. https://doi.org/10.3141/1840-12

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

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

Gulino, M. S., Fiorentino, A., & Vangi, D. (2022). Prospective and retrospective performance assessment of Advanced Driver Assistance Systems in imminent collision scenarios: The CMI-Vr approach. *European Transport Research Review*, *14*, Article 3. https://doi.org/10.1186/s12544-022-00527-4

Guo, F., Klauer, S. G., Hankey, J. M., & Dingus, T. A. (2010). Near crashes as crash surrogate for naturalistic driving studies. *Transportation Research Record*, *2147*(1), 66–74. https://doi.org/10.3141/2147-09

Han, S., Guo, F., & Klauer, C. (2023). *In-depth analysis of crash risk associated with eyes-off-road duration by road control type and intersection type* (Report No. 23-UM-122). Virginia Tech Transportation Institute. http://hdl.handle.net/10919/113544

Hankey, J. M., Perez, M. A., & McClafferty, J. A. (2016). *Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets*. Virginia Tech Transportation Institute. http://hdl.handle.net/10919/70850

Hayward, J. C. (1972). Near miss determination through use of a scale of danger. *Highway Research Record*, *384*, 24–34. https://trid.trb.org/View/115323

Hidas, P. (2005). Modelling vehicle interactions in microscopic simulation of merging and weaving. *Transportation Research Part C: Emerging Technologies*, *13*(1), 37–62. https://doi.org/10.1016/j.trc.2004.12.003

Horrey, W. J., & Wickens, C. D. (2007). In-vehicle glance duration: Distributions, tails, and model of crash risk. *Transportation Research Record*, *2018*(1), 22–28. https://doi.org/10.3141/2018-04

Hozhabr Pour, H., Li, F., Wegmeth, L., Trense, C., Doniec, R., Grzegorzek, M., & Wismüller, R. (2022). A machine learning framework for automated accident detection based on multimodal sensors in cars. *Sensors*, *22*(10), Article 3634. https://doi.org/10.3390/s22103634

Hydén, C. (1987). *The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique* [Doctoral dissertation, Lund University]. https://www.ictct.net/wp-content/uploads/SMoS_Library/LIB_Hyden_1987.pdf

ISO. (2018). *Road vehicles — Functional safety — Part 1: Vocabulary* (ISO 26262-1:2018). https://www.iso.org/standard/68383.html

ISO. (2020). *Road vehicles — Measurement and analysis of driver visual behaviour with respect to transport information and control systems* (ISO 15007:2020). https://www.iso.org/standard/63220.html

ISO. (2021). *Road vehicles — Prospective safety performance assessment of pre-crash technology by virtual simulation — Part 1: State-of-the-art and general method overview* (ISO/TR 21934-1:2021). https://www.iso.org/standard/76497.html

ISO. (2022a). *Road vehicles — Safety of the intended functionality* (ISO 21448:2022). https://www.iso.org/standard/77490.html

ISO. (2022b). *Road vehicles — Test scenarios for automated driving systems — Scenario based safety evaluation framework* (ISO 34502:2022). https://www.iso.org/standard/78951.html

ISO. (2024). *Road vehicles — Prospective safety performance assessment of pre-crash technology by virtual simulation — Part 2: Guidelines and requirements for application* (ISO/TS 21934-2:2024). https://www.iso.org/standard/81790.html

ISO. (2025). *Road vehicles — Safety for automated driving systems — Design, verification and validation* (ISO/TS 5083:2025). https://www.iso.org/standard/81920.html

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

JAMA. (2022). *Automated driving safety evaluation framework* (Ver. 3.0). Japan Automobile Manufacturers Association. https://www.jama.or.jp/english/reports/docs/Automated_Driving_Safety_Evaluation_Framework_Ver3.0.pdf

James, W., McKinzie, S., Benson, W., & Heise, C. (2015). *Crash investigation and reconstruction technologies and best practices* (Report No. FHWA-HOP-16-009). Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/50639

Jiao, Y., Calvert, S. C., Van Cranenburgh, S., & Van Lint, H. (2025). A unified probabilistic approach to traffic conflict detection. *Analytic Methods in Accident Research*, *45*, Article 100369. https://doi.org/10.1016/j.amar.2024.100369

Kiefer, R. J., LeBlanc, D. J., & Flannagan, C. A. (2005). Developing an inverse time-to-collision crash alert timing approach based on drivers' last-second braking and steering judgments. *Accident Analysis & Prevention*, *37*(2), 295–303. https://doi.org/10.1016/j.aap.2004.09.003

Kircher, K., & Ahlström, C. (2024). A comparison of glance coding approaches for driver attention assessment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *100*, 243–253. https://doi.org/10.1016/j.trf.2023.12.003

Kiuchi, T. (2020). *Using PCM on micro data for autonomous vehicles: A collaboration with VUFO in Germany*. Institute for Traffic Accident Research and Data Analysis. https://www.itarda.or.jp/contents/10235/23_06microPCM_e.pdf

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. https://doi.org/10.1056/NEJMsa1204142

Knipling, R. R. (2015). Naturalistic driving events: No harm, no foul, no validity. In *Proceedings of the 8th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 197–203). University of Iowa. https://doi.org/10.17077/drivingassessment.1571

Krajewski, R., Bock, J., Kloeker, L., & Eckstein, L. (2018). The highD dataset: A drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 2118–2125). IEEE. https://doi.org/10.1109/ITSC.2018.8569552

Krajewski, R., Moers, T., Bock, J., Vater, L., & Eckstein, L. (2020). The rounD dataset: A drone dataset of road user trajectories at roundabouts in Germany. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. https://doi.org/10.1109/ITSC45102.2020.9294728

Kullgren, A. (2008). Dose-response models and EDR data for assessment of injury risk and effectiveness of safety systems. In *Proceedings of the 2008 International IRCOBI Conference on the Biomechanics of Injury* (pp. 3–14). IRCOBI. https://www.ircobi.org/wordpress/downloads/irc0111/2008/BertilAldmanLecture/bal.pdf

Kullgren, A., Krafft, M., Tingvall, C., & Lie, A. (2003). Combining crash recorder and paired comparison technique: Injury risk functions in frontal and rear impacts with special reference to neck injuries. In *Proceedings of the 18th International Technical Conference on Enhanced Safety of Vehicles (ESV)* (Paper Number 404). National

Highway Traffic Safety Administration. https://www.nhtsa.gov/sites/nhtsa.gov/files/18esv-000404.pdf

Kusano, K., & Victor, T. (2022). Methodology for determining maximum injury potential for automated driving system evaluation. *Traffic Injury Prevention*, *23*(Suppl. 1), S224–S227. https://doi.org/10.1080/15389588.2022.2125231

Kusano, K. D., & Gabler, H. C. (2012). Safety benefits of forward collision warning, brake assist, and autonomous braking systems in rear-end collisions. *IEEE Transactions on Intelligent Transportation Systems*, *13*(4), 1546–1555. https://doi.org/10.1109/TITS.2012.2191542

Kusano, K. D., Scanlon, J. M., Chen, Y.-H., McMurry, T. L., Chen, R., Gode, T., & Victor, T. (2024). Comparison of Waymo rider-only crash data to human benchmarks at 7.1 million miles. *Traffic Injury Prevention*, *25*(Suppl. 1), S66–S77. https://doi.org/10.1080/15389588.2024.2380786

Laureshyn, A. (2010). *Application of automated video analysis to road user behaviour* [Doctoral dissertation, Lund University]. https://portal.research.lu.se/files/5337337/1522970.pdf

Laureshyn, A., & Várhelyi, A. (2018). *The Swedish Traffic Conflict Technique: Observer's manual*. Department of Technology and Society, Lund University. https://lucris.lub.lu.se/ws/files/51195704/TCT_Manual_2018.pdf

Lee, J., & Jang, K. (2019). A framework for evaluating aggressive driving behaviors based on in-vehicle driving records. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 610–619. https://doi.org/10.1016/j.trf.2017.11.021

Lee, J. Y., Lee, J. D., Bärgman, J., Lee, J., & Reimer, B. (2018). How safe is tuning a radio?: Using the radio tuning task as a benchmark for distracted driving. *Accident Analysis & Prevention*, *110*, 29–37. https://doi.org/10.1016/j.aap.2017.10.009

Lee, Y.-C., Lee, J. D., & Ng Boyle, L. (2007). Visual attention in driving: The effects of cognitive load and visual disruption. *Human Factors*, *49*(4), 721–733. https://doi.org/10.1518/001872007X215791

Leledakis, A., Lindman, M., Östh, J., Wågström, L., Davidsson, J., & Jakobsson, L. (2021). A method for predicting crash configurations using counterfactual simulations and real-world data. *Accident Analysis & Prevention*, *150*, Article 105932. https://doi.org/10.1016/j.aap.2020.105932

Li, W., Pan, C. W., Zhang, R., Ren, J. P., Ma, Y. X., Fang, J., Yan, F. L., Geng, Q. C., Huang, X. Y., Gong, H. J., Xu, W. W., Wang, G. P., Manocha, D., & Yang, R. G. (2019). AADS: Augmented autonomous driving simulation using data-driven algorithms. *Science Robotics*, *4*(28), Article eaaw0863. https://doi.org/10.1126/scirobotics.aaw0863

Liers, H. (2018, November 20). *Traffic accident research in Germany and the German In-Depth Accident Study (GIDAS)* [Conference presentation]. ACMA-SIAM-VDA Conference on Safer & Sustainable Road Transportation, New Delhi, India. https://www.gidas.org/pdf/181120_SIAM_ACMA_VDA_Conference_Liers_GIDAS-Presentation.pdf

Liu, S., Wang, X., Hassanin, O., Xu, X., Yang, M., Hurwitz, D., & Wu, X. (2021). Calibration and evaluation of responsibility-sensitive safety (RSS) in automated vehicle performance during cut-in scenarios. *Transportation Research Part C: Emerging Technologies*, *125*, Article 103037. https://doi.org/10.1016/j.trc.2021.103037

Mack, A. (2003). Inattentional blindness: Looking without seeing. *Current Directions in Psychological Science*, *12*(5), 180–184. https://doi.org/10.1111/1467-8721.01256

Maddox, M. E., & Kiefer, A. (2012). Looming threshold limits and their use in forensic practice. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 700–704. https://doi.org/10.1177/1071181312561146

Markkula, G. (2014). Modeling driver control behavior in both routine and near-accident driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 879–883. https://doi.org/10.1177/1541931214581185

Markkula, G., Boer, E., Romano, R., & Merat, N. (2018). Sustained sensorimotor control as intermittent decisions about prediction errors: Computational framework and application to ground vehicle steering. *Biological Cybernetics*, *112*(3), 181–207. https://doi.org/10.1007/s00422-017-0743-9

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*, 209–226. https://doi.org/10.1016/j.aap.2016.07.007

Mattas, K., Albano, G., Dona, R., Galassi, M. C., Suarez-Bertoa, R., Vass, S., & Ciuffo, B. (2022). Driver models for the definition of safety requirements of automated vehicles in international regulations. Application to motorway driving conditions. *Accident Analysis & Prevention*, *174*, Article 106743. https://doi.org/10.1016/j.aap.2022.106743

Mattas, K., Makridis, M., Botzoris, G., Kriston, A., Minarini, F., Papadopoulos, B., Re, F., Rognelund, G., & Ciuffo, B. (2020). Fuzzy surrogate safety metrics for real-time assessment of rear-end collision risk. A study based on empirical observations. *Accident Analysis & Prevention*, *148*, Article 105794. https://doi.org/10.1016/j.aap.2020.105794

Menzel, T., Chajmowicz, H., Fahrenkrog, F., Sander, U., Glasmacher, C., Wimmer, P., Lorente-Mallada, J., Schories, L. T., Kalisvaart, S., Bärgman, J., & Sander, D. (2025). *Performance assessment of safety solutions and countermeasures in chosen use cases* (Deliverable D6.1). Horizon Europe project V4SAFETY. https://v4safetyproject.eu/media/pages/outputs/7b62d750be-1761749900/d6.1-performance-assessment-of-safety-solutions-and-countermeasures-in-chosen-use-cases.pdf

Morando, A. (2025). *Method for recovering data on unreported low-severity crashes*. arXiv. https://doi.org/10.48550/arXiv.2503.04529

Morando, A., Victor, T., & Dozza, M. (2019). A reference model for driver attention in automation: Glance behavior changes during lateral and longitudinal assistance. *IEEE Transactions on Intelligent Transportation Systems*, *20*(8), 2999–3009. https://doi.org/10.1109/tits.2018.2870909

Morgenstern, T., Petzoldt, T., Krems, J. F., Naujoks, F., & Keinath, A. (2020). Using European naturalistic driving data to assess secondary task engagement when stopped at a red light. *Journal of Safety Research*, *73*, 235–243. https://doi.org/10.1016/j.jsr.2020.03.007

Moridpour, S., Sarvi, M., & Rose, G. (2010). Modeling the lane-changing execution of multiclass vehicles under heavy traffic conditions. *Transportation Research Record*, *2161*(1), 11–19. https://doi.org/10.3141/2161-02

NHTSA. (2012). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices.* (Docket No. NHTSA–2010–0053). http://www.gpo.gov/fdsys/pkg/FR-2012-02-24/pdf/2012-4017.pdf

Nidamanuri, J., Nibhanupudi, C., Assfalg, R., & Venkataraman, H. (2021). A progressive review: Emerging technologies for ADAS driven solutions. *IEEE Transactions on Intelligent Vehicles*, *7*(2), 326–341. https://doi.org/10.1109/TIV.2021.3122898

Näätänen, R., & Summala, H. (1974). A model for the role of motivational factors in drivers' decision-making. *Accident Analysis & Prevention*, *6*(3-4), 243–261. https://doi.org/10.1016/0001-4575(74)90003-7

Olleja, P., Bärgman, J., & Markkula, G. (2022, October 19–20). *Quantification of driver's side-glance frequency and duration in straight highway driving* [Conference presentation]. The 8th International Conference on Driver Distraction and Inattention, Gothenburg, Sweden. https://ddi2022.org/wp-content/uploads/2022/10/3.1_Pierluigi_Olleja_DDI.pdf

Olleja, P., Rasch, A., & Bärgman, J. (2023, October 26–27). *A reference driver model for overtaking a cyclist* [Conference presentation]. International Cooperation on Theories and Concepts in Traffic safety (ICTCT), Catania, Italy. https://www.ictct.net/wp-content/uploads/35-Catania-2023/69-Olleja.pdf

Otte, D., Krettek, C., Brunner, H., & Zwipp, H. (2003). Scientific approach and methodology of a new in-depth-investigation study in Germany so called GIDAS. In *Proceedings of the 18th International Technical Conference on Enhanced Safety of Vehicles (ESV)* (Paper Number 161). National Highway Traffic Safety Administration. https://www-nrd.nhtsa.dot.gov/pdf/ESV/Proceedings/18/18ESV-000161.pdf

Pelella, F., Zaccaria, G., Montanino, M., Punzo, V., Najafabadi, A. N., Weidl, G., Raiyn, J., & Casas, J. (2023). *Critical review of state-of-the-art techniques to model drivers' heterogeneity* (Deliverable 4.1). i4Driving. https://i4driving.eu/wp-content/uploads/2024/07/D4.1.pdf

Penttinen, M., Rämä, P., Dotzauer, M., Hibberd, D., Innamaa, S., Louw, T., Streubel, T., Metz, B., Woerle, J., Brouwer, R., Rösener, C., & Weber, H. (2019). *Experimental procedure* (Deliverable D3.2). L3Pilot. https://l3pilot.eu/fileadmin/user_upload/Downloads/Deliverables/Update_07102021/L3Pilot-SP3-D3.2-Experimental_procedure-v1.0_for_website.pdf

Piccinini, G. B., Engström, J., Bärgman, J., & Wang, X. (2017). Factors contributing to commercial vehicle rear-end conflicts in China: A study using on-board event data recorders. *Journal of Safety Research*, *62*, 143–153. https://doi.org/10.1016/j.jsr.2017.06.004

Rameshkrishnan, N., Sathyakumar, A., Balakumar, S., Hassan, A., Rajaraman, R., & Padmanaban, J. (2013). The new in-depth, at-the-scene, accident investigation database in India. In *Proceedings of the 2013 International IRCOBI Conference on the Biomechanics of Injury* (pp. 36–50). IRCOBI. https://www.ircobi.org/wordpress/downloads/irc13/pdf_files/13.pdf

Rasch, A., & Dozza, M. (2022). Modeling drivers' strategy when overtaking cyclists in the presence of oncoming traffic. *IEEE Transactions on Intelligent Transportation Systems*, *23*(3), 2180–2189. https://doi.org/10.1109/TITS.2020.3034679

Rothoff, M., Zarghampour, H., Tivesten, E., Victor, T., Ödblom, A., Bergh, T., Lindenberg, B., Lindman, M., Sangberg, F., Aust, M. L., Tayeh, R., Hartelius, E., Granum, F., Ringwall, E., Iancovici, S., Ivancic, A., Persson, C., Pourabdollah, M., Bjärkvik, E., … Palm, M. (2019). *ADEST autonomous driving effects on sustainable transportation & ADFE autonomous driving fuel economy: Public report*. FFI. https://www.vinnova.se/globalassets/mikrosajter/ffi/dokument/slutrapporter-ffi/trafiksakerhet-och-automatiserade-fordon-rapporter/2014-06012-driveme.pdf

Sander, U. (2017). Opportunities and limitations for intersection collision intervention—A study of real world 'left turn across path' accidents. *Accident Analysis & Prevention*, *99*(Part A), 342–355. https://doi.org/10.1016/j.aap.2016.12.011

Sayed, T., Brown, G., & Navin, F. (1994). Simulation of traffic conflicts at unsignalized intersections with TSC-Sim. *Accident Analysis & Prevention*, *26*(5), 593–607. https://doi.org/10.1016/0001-4575(94)90021-3

Scanlon, J., Kusano, K., Engström, J., & Victor, T. (2022). *Collision avoidance effectiveness of an automated driving system using a human driver behavior reference model in reconstructed fatal collisions*. Waymo. https://waymo.com/research/collision-avoidance-effectiveness-of-an-automated/

Scanlon, J. M., Kusano, K. D., Daniel, T., Alderson, C., Ogle, A., & Victor, T. (2021). Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain. *Accident Analysis & Prevention*, *163*, Article 106454. https://doi.org/10.1016/j.aap.2021.106454

Schubert, A., Erbsmehl, C., & Hannawald, L. (2013). Standardized pre-crash-scenarios in digital format on the basis of the VUFO simulation. In *Proceedings of the 5th International Conference on ESAR "Expert Symposium on Accident Research"*. https://bast.opus.hbz-nrw.de/frontdoor/deliver/index/docId/630/file/31_Schubert_Erbsmehl.pdf

Seppelt, B., Seaman, S., Angell, L., Mehler, B., Reimer, B., Victor, T., Bruyas, M., Regan, M., Brusque, C., & Fort, A. (2018). Assessing the effect of in-vehicle task interactions on attention management in safety-critical events. In *Proceedings of the 6th International Conference on Driver Distraction and Inattention*. http://ddi2018.org/wp-content/uploads/2018/10/S1.3-Seppelt.pdf

Seppelt, B. D., Seaman, S., Lee, J., Angell, L. S., Mehler, B., & Reimer, B. (2017). Glass half-full: On-road glance metrics differentiate crashes from near-crashes in the 100-Car data. *Accident Analysis & Prevention*, *107*, 48–62. https://doi.org/10.1016/j.aap.2017.07.021

Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2017). *On a formal model of safe and scalable self-driving cars*. arXiv. https://arxiv.org/abs/1708.06374

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Siebinga, O., Zgonnikov, A., & Abbink, D. A. (2024). A model of dyadic merging interactions explains human drivers' behavior from control inputs to decisions. *PNAS Nexus*, *3*(10), Article pgae420. https://doi.org/10.1093/pnasnexus/pgae420

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences*, *4*(4), 147–155. https://doi.org/10.1016/S1364-6613(00)01455-8

Smit, S., Tomasch, E., Kolk, H., Plank, M. A., Gugler, J., & Glaser, H. (2019). Evaluation of a momentum based impact model in frontal car collisions for the prospective assessment of ADAS. *European Transport Research Review*, *11*, Article 2. https://doi.org/10.1186/s12544-018-0343-3

Smith, K., Thome, A., Blåberg, C., & Bärgman, J. (2009). An invariant may drive the decision to encroach at unsignalized intersections. In *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 313–319). University of Iowa. https://doi.org/10.17077/drivingassessment.1337

Summala, H. (2007). Towards understanding motivational and emotional factors in driver behavior: Comfort through satisficing. In C. P. Cacciabue (Ed.), *Modelling driver behavior in automotive environments* (pp. 189–207). Springer. https://doi.org/10.1007/978-1-84628-618-6_11

Sun, Z., Lin, M., Chen, W., Dai, B., Ying, P., & Zhou, Q. (2024). A case study of unavoidable accidents of autonomous vehicles. *Traffic Injury Prevention*, *25*(1), 8–13. https://doi.org/10.1080/15389588.2023.2255333

Svenson, A. L., Grove, K., Atwood, J., & Blanco, M. (2017). Using naturalistic data to evaluate heavy vehicle crash avoidance systems performance. In *Proceedings of the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)* (Paper Number 17-0231). National Highway Traffic Safety Administration. https://www-nrd.nhtsa.dot.gov/pdf/ESV/Proceedings/25/25ESV-000231.pdf

Svärd, M., Bärgman, J., & Victor, T. (2021). Detection and response to critical lead vehicle deceleration events with peripheral vision: Glance response times are independent of visual eccentricity. *Accident Analysis & Prevention*, *150*, Article 105853. https://doi.org/10.1016/j.aap.2020.105853

Svärd, M., Markkula, G., Bärgman, J., & Victor, T. (2021). Computational modeling of driver pre-crash brake response, with and without off-road glances: Parameterization using real-world crashes and near-crashes. *Accident Analysis & Prevention*, *163*, Article 106433. https://doi.org/10.1016/j.aap.2021.106433

Svärd, M., Markkula, G., Engström, J., Granum, F., & Bärgman, J. (2017). A quantitative driver model of pre-crash brake onset and control. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 339–343. https://doi.org/10.1177/1541931213601565

Tarko, A. P. (2018). Surrogate measures of safety. In D. Lord & S. Washington (Eds.), *Safe mobility: Challenges, methodology and solutions* (pp. 383–405). Emerald Publishing Limited. https://doi.org/10.1108/S2044-994120180000011019

Tijerina, L. (1999). Driver eye glance behavior during car following on the road. *SAE transactions*, *180*, 2369–2383. https://doi.org/10.4271/1999-01-1300

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

UNECE. (2023). *UN Regulation No. 157. Uniform provisions concerning the approval of vehicles with regard to Automated Lane Keeping Systems* (E/ECE/TRANS/505/Rev.3/Add.156/Amend.4). https://unece.org/transport/documents/2023/03/standards/un-regulation-no-157-amend4

Van Lint, J., & Calvert, S. C. (2018). A generic multi-level framework for microscopic traffic simulation—Theory and an example case in modelling driver distraction. *Transportation Research Part B: Methodological*, *117*, 63–86. https://doi.org/10.1016/j.trb.2018.08.009

Van Nes, N., Christoph, M., Hoedemaeker, M., & Van der Horst, R. A. (2013). The value of site-based observations complementary to naturalistic driving observations: A pilot study on the right turn manoeuvre. *Accident Analysis & Prevention*, *58*, 318–329. https://doi.org/10.1016/j.aap.2013.06.026

Vechione, M., Balal, E., & Cheu, R. L. (2018). Comparisons of mandatory and discretionary lane changing behavior on freeways. *International Journal of Transportation Science and Technology*, *7*(2), 124–136. https://doi.org/10.1016/j.ijtst.2018.02.002

Victor, T., Dozza, M., Bärgman, J., Boda, C. N., Engström, J., Flannagan, C., Lee, J., & Markkula, G. (2014). *Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk* (SHRP 2 Report S2-S08A-RW-1). The National Academies Press. https://doi.org/10.17226/22297

VTTI. (n.d.). *InSight data access website: SHRP 2 naturalistic driving study.* Retrieved March 11, 2024, from https://insight.shrp2nds.us/

Wang, X., Peng, Y., Xu, T., Xu, Q., Wu, X., Xiang, G., Yi, S., & Wang, H. (2022). Autonomous driving testing scenario generation based on in-depth vehicle-to-powered two-wheeler crash data in China. *Accident Analysis & Prevention*, *176*, Article 106812. https://doi.org/10.1016/j.aap.2022.106812

Webb, N., Smith, D., Ludwick, C., Victor, T., Hommes, Q., Favaro, F., Ivanov, G., & Daniel, T. (2020). *Waymo's safety methodologies and safety readiness determinations*. arXiv. https://doi.org/10.48550/arXiv.2011.00054

Westhofen, L., Neurohr, C., Koopmann, T., Butz, M., Schütt, B., Utesch, F., Neurohr, B., Gutenkunst, C., & Böde, E. (2023). Criticality metrics for automated driving: A review and suitability analysis of the state of the art. *Archives of Computational Methods in Engineering*, *30*, 1–35. https://doi.org/10.1007/s11831-022-09788-7

WHO. (2023). *Global status report on road safety 2023*. https://iris.who.int/bitstream/handle/10665/375016/9789240086517-eng.pdf?sequence=1

Wimmer, P., Düring, M., Chajmowicz, H., Granum, F., King, J., Kolk, H., Op den Camp, O., Scognamiglio, P., & Wagner, M. (2019). Toward harmonizing prospective effectiveness assessment for road safety: Comparing tools in standard test case simulations. *Traffic Injury Prevention*, *20*(Suppl. 1), S139–S145. https://doi.org/10.1080/15389588.2019.1616086

Wimmer, P., Op den Camp, O., Weber, H., Chajmowicz, H., Wagner, M., Mallada, J. L., Fahrenkrog, F., & Denk, F. (2023). Harmonized approaches for baseline creation in prospective safety performance assessment of driving automation systems. In *Proceedings of the 27th International Technical Conference on the Enhanced Safety of Vehicles (ESV)* (Paper Number 23-0032). National Highway Traffic Safety Administration. https://www-nrd.nhtsa.dot.gov/pdf/ESV/Proceedings/27/27ESV-000032.pdf

Wood, M., Robbel, P., Maass, M., Tebbens, R. D., Meijs, M., Harb, M., Reach, J., Robinson, K., Wittmann, D., Srivastava, T., Bouzouraa, M. E., Liu, S. Y., Wang, Y., Knobel, C., Boymanns, D., Löhning, M., Dehlink, B., Kaule, D., Krüger, R., … Schlicht, P. (2019). *Safety first for automated driving* [White paper]. https://www.aptiv.com/docs/default-source/white-papers/safety-first-for-automated-driving-aptiv-white-paper.pdf

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* (Report No. UMTRI-2011-36). The University of Michigan Transportation Research Institute.

Wu, J. (2024). *Generation of representative pre-crash scenarios across the full severity range using real-world crash data: Towards more accurate virtual assessments of active safety technologies* [Licentiate thesis, Chalmers University of Technology]. https://research.chalmers.se/publication/542271

Wu, J., Flannagan, C., Sander, U., & Bärgman, J. (2025). Model-based generation of representative rear-end crash scenarios across the full severity range using pre-crash data. *IEEE Transactions on Intelligent Transportation Systems*, *26*(10), 15932–15950. https://doi.org/10.1109/TITS.2025.3573386

Wu, J., Sander, U., Flannagan, C., Zhao, M., & Bärgman, J. (2025). Practical equivalence testing and its application in synthetic pre-crash scenario validation. In *Proceedings of the 2025 IEEE International Automated Vehicle Validation Conference (IAVVC)*. IEEE. https://doi.org/10.1109/IAVVC61942.2025.11219586

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

Xue, Q., Markkula, G., Yan, X., & Merat, N. (2018). Using perceptual cues for brake response to a lead vehicle: Comparing threshold and accumulator models of visual looming. *Accident Analysis & Prevention*, *118*, 114–124. https://doi.org/10.1016/j.aap.2018.06.006

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. https://doi.org/10.1016/j.aap.2007.10.016

Yang, X., Lubbe, N., & Bärgman, J. (2024). Evaluation of comfort zone boundary based automated emergency braking algorithms for car-to-powered-two-wheeler crashes in China. *IET Intelligent Transport Systems*, *18*(9), 1599–1615. https://doi.org/10.1049/itr2.12532

Zgonnikov, A., Abbink, D., & Markkula, G. (2024). Should I stay or should I go? Cognitive modeling of left-turn gap acceptance decisions in human drivers. *Human Factors*, *66*(5), 1399–1413. https://doi.org/10.1177/00187208221144561

Zhang, F., Subramanian, R., Chen, C.-L., & Noh, E. Y. (2019). *Crash investigation sampling system: Design overview, analytic guidance, and FAQs* (Report No. DOT HS 812 801). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812801

Åsljung, D., Zandén, C., Fredriksson, J., & Vakilzadeh, M. K. (2021). On automated vehicle collision risk estimation using threat metrics in subset simulation. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)* (pp. 58–63). IEEE. https://doi.org/10.1109/ITSC48978.2021.9564695