THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Towards More Reliable Pre-Crash Virtual Safety Assessment

The impact of the choice of data types and reference driver models on the assessment of vehicle automation

PIERLUIGI OLLEJA

Department of Mechanics and Maritime Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2024

Towards More Reliable Pre-Crash Virtual Safety Assessment
The impact of the choice of data types and reference driver models on the assessment of vehicle automation
PIERLUIGI OLLEJA

Technical report no 2024:08

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

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

Chalmers digitaltryck
Gothenburg, Sweden 2024

Abstract

Road crashes are a major cause of deaths and serious injuries worldwide. New technologies offer the opportunity to reduce road crashes 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). Methods are in place to assess how safe these systems are. One of these methods employs virtual simulations to predict the impact on safety that the systems would have once released on public roads. However, the process for ensuring that a virtual simulation provides an effective, relevant, and fair assessment of ADASs and ADSs is not always straightforward. This thesis contributes to the development of virtual safety assessment methods by investigating the impact of different data and models on the resulting simulations.

Specifically, the first objective of the thesis is to measure the impact of data selection on the outcomes of virtual safety assessment. Crashes were artificially generated from near-crashes and everyday driving data, using a model of an unresponsive driver. The generated crashes were compared to real-world reconstructed crashes. Automated emergency braking (AEB) systems were then applied to the crashes, to study the impact different data sources have on crash avoidance and mitigation. The results show that those artificially generated crashes are very different from real-world crashes, with lower severity outcomes and criticality.

The second objective of this thesis is to understand if existing reference driver models represent a competent and careful human driver. These models are intended to be benchmarks for ADS safety performance. The models studied in this thesis—from the UN Regulation No. 157—did not perform as the competent and careful drivers they are intended to represent when applied on near-crash cut-ins through counterfactual simulations. Specifically, one model generally showed delayed responses to critical scenarios, compared to humans. The other model instead showed non-human-like behavior, reacting substantially earlier than humans.

The impact of the findings is twofold. First, they can help the development of virtual safety assessment methods by discouraging the use of everyday driving data and near-crash data in counterfactual crash generation. Second, the findings on reference driver models make it clear that models used in regulations must be validated using a range of data types. To continue the work on reference driving models, future work aims at studying how urgency in traffic scenarios impacts drivers' behaviors. The concept of comfort zone boundaries (CZBs) will be used to study the limits that drivers are able and willing to tolerate in routine driving, and the inclusion of CZBs in the models will be investigated. This research has the potential to contribute to the improvement of reference driver models and virtual safety assessment methods.

Keywords: virtual safety assessment, reference driver model, counterfactual simulations, crash surrogates, conflict and crash avoidance.

Acknowledgements

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 and Mikael Ljung Aust for their support. And thank you to my examiner Marco Dozza for many insightful and thought-provoking conversations. I am 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 is always helpful, kind and supportive.

Thank you to my family, who never feel as distant as geography might suggest. And to my friends—Maria, Valeria, Alessio, Erik, Jonas, and Amr—and everyone who helped me settle in Gothenburg, making this city such a special place for me. To my friends Christopher, Noemi, Morgan, Kathi, and Ron, spread across Europe but always a safe haven. And a special thank you to Clarissa, who always helps me see the good I have to offer.

Finally, I acknowledge the funding project FFI, which, through the QUADRIS project (contract number 2020-05156) funded this research work and Autoliv Research for funding the initial part of my work.

List of publications

This thesis is based on the following publications:

Paper I

Olleja, P., Bargman, 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: Setup and execution of the simulations, and conducted data analysis. Wrote the first draft of the paper and collaborated with the co-authors to finalize the manuscript.

Paper II

Olleja, P., Markkula, G., & Bärgman, J. (submitted). Validation of human benchmark models for Automated Driving System approval: How competent and careful are they really? *arXiv preprint arXiv:2406.09493*. *Under review at Accident Analysis & Prevention.* https://doi.org/10.48550/arXiv.2406.09493

Author's contribution: 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.

List of additional work

Other relevant presentations and publications, but not included in this thesis:

Jokhio, S., Olleja, P., Bärgman, J., Yan, F., & Baumann, M. (2023). 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

Jokhio, S., Olleja, P., Bärgman, J., Yan, F., & M., B. (2023). 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

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

Contents

1 Introduction

Traffic safety plays a major role in the development of the mobility of the future. Road crashes are historically 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 being involved in a crash, incurring serious injuries or death, 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. Solutions for safer passenger cars address the in-crash phase (while a crash is occurring) or the pre-crash phase (typically a few seconds before the crash). In-crash solutions improve the vehicle's ability to absorb the energy of a crash as well as the restraint systems' ability to protect the occupants. Pre-crash solutions, on the other hand, improve the vehicle's ability to avoid the crash in the first place. Advanced driver assistance systems (ADASs) and, more recently, automated driving systems (ADSs), are two examples of the latter. The pre-crash phase covers a wide range of driving situations, from low-risk to high-risk (imminent crashes). This range, however, can be divided into two levels of system intervention, crash avoidance (high-risk) and conflict avoidance (low-risk). Crash avoidance is the ability of the subject vehicle's system 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 could, if not treated early enough, increase the probability of a crash. However, the border between these two levels is not well-defined.

Automated emergency braking (AEB) is an example of an ADAS that acts in the pre-crash phase when safety criticality is high—thus it is referred to as a crash avoidance system. However, as their development continues, ADASs are increasingly able to handle driving tasks even in low-criticality situations (Antony & Whenish, 2021; Nidamanuri et al., 2021). Unlike ADASs, ADSs promise to 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 is regulated by standards and regulations that define their safe operation, a task requiring well-defined procedures for testing hardware and implementing software. The international standard that defines functional safety for electrical and electronic systems installed in vehicles is ISO 26262 (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". 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). The standard categorizes the infinite possible interactions that ADSs can encounter into a finite number of scenarios. The result is a scenario-based approach that divides the driving task into three aspects: perception, judgement and control. Each aspect is associated with the physics principles that ADSs utilize or are influenced by: wave propagation for sensor detection of the environment—perception; kinematics for path planning—judgment; and vehicle dynamics for executing the driving commands—control. The scenario categories can be used to obtain quantitative results from ADS testing. An example of scenario-based ADS testing is described in the UN Regulation No. 157 (R157; UNECE, 2023). This regulation contains provisions for the approval of automated lane-keeping systems (ALKSs) using scenario-based testing, as part of EU Regulation (Regulation 1143/2014). UNECE stands for United Nations Economic Commission for Europe.

The performance of systems such as ADASs and ADSs must be tested to ensure their safe operation. However, it typically takes years for these systems to reach deep enough market penetration to enable valid retrospective safety evaluations (Gulino et al., 2022; Smit et al., 2019; Wimmer et al., 2019). Consequently, to aid in their development and regulation, the assessment method must be prospective (carried out prior to the system's release on the market). These methods, often performed in virtual environments, predict the impact of a system if it were available in traffic (Alvarez et al., 2017) and thus play a fundamental role in the development of ADASs and ADSs. Moreover, they 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). When quantifying the safety impact of a system a comparison between the treatment (i.e., the use of the system) and some baseline (i.e., the non-use of the system) must be made. Traffic safety research borrows the terms baseline and treatment from the medical field. Systems applied to reduce or avoid crashes are considered analog to medicines for treating a disease. Therefore, a traffic scenario without the safety system in assessment is considered a baseline scenario, and one in which the system is applied is a treatment scenario. The baseline should, at least in theory, represent the real-world traffic situations considered important for safety (and that the systems aim to address) as accurately as possible. That is, the choice must be based on the evaluation scope—what should be assessed and what the purpose is. Fahrenkrog et al. (2024) 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, and should also "point out the gap that is addressed by the evaluation" (p. 25; Fahrenkrog et al., 2024).

Wimmer et al. (2023) and Fahrenkrog et al. (2024) define three fundamentally different approaches to baseline generation for prospective safety assessments. The first uses unmodified real-world events as the baseline (A in Fig. 1). The second approach (B) consists of increasing the number of available scenarios by modifying individual real-world scenarios. The third approach (C) has two sub-approaches: in one, statistical data are aggregated to a few representative cases, and the other uses stochastic models to increase the coverage (create more cases) for the baseline. The illustration in Fig. 1 shows the three approaches described in Wimmer et al. (2023).

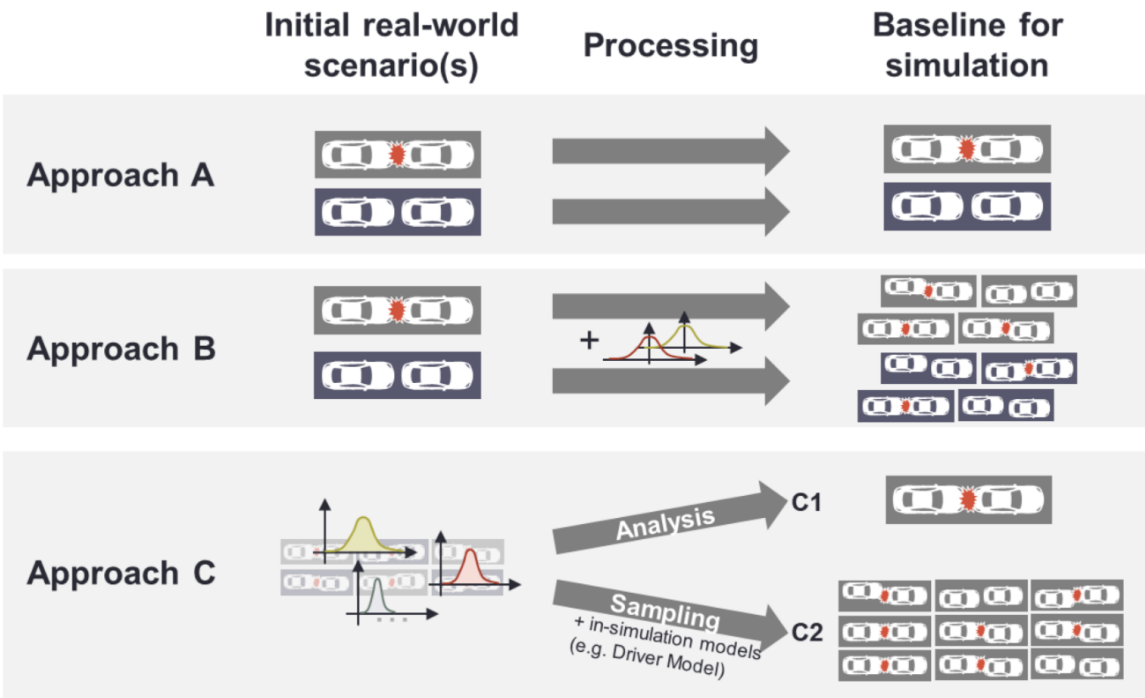

Fig. 1 Approaches for creating baseline data for virtual safety assessments from Wimmer et al. (2023; reproduced with permission from the authors). In the column "Initial real-world scenario(s)" approaches A and B have individual original scenarios as a starting point, while approach C uses distributions of parameters that describe the scenarios. The processing for approach A consists in digital representation of the unmodified original scenarios (the second and third column). Approach B is similar to A, with the difference that here the processing can include parameter variations for part of the scenario, resulting in more scenarios than what was initially available. For approach C, the first sub-approach (C1) aggregates parameter sets to generate a few representative scenarios; sub-approach C2 typically uses sampling techniques and models of road users to obtain large datasets of interactions.

While baselines are all in some way derived from the real world—either directly via statistical models, or indirectly via behavior models—the data they use 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). These databases provide relatively high-fidelity data, and they are commonly used as baseline when assessing crash avoidance systems (Erbsmehl, 2009). Some of them capture the characteristics of real-world crashes in detail, including pre-crash time-series data of the 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 (Schubert et al., 2013). However, in-depth crash databases can only describe a relatively small part of the complexity of the driving task—and 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) collect data during drivers' everyday driving. They can complement in-depth crash databases by providing much more complete coverage of all safety criticalities (Bärgman, 2016; Hankey et al., 2016), albeit at the expense of capturing few (if any) crashes, and very few high-severity crashes (Ehsani et al., 2021). As crashes are rare in NDSs, near-crashes are commonly used as surrogates of crashes (Guo et al., 2010; Hydén, 1987; Laureshyn & Várhelyi, 2018; Victor et al., 2015; 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 (Bärgman et al., 2015; Davis et al., 2011) by modifying the real-world driving kinematics which successfully avoided the crash. This method of obtaining relevant safety-critical scenarios is, however, not always straightforward and is therefore an active research field (see, e.g., Fahrenkrog et al., 2024; Wimmer et al., 2023).

One safety assessment method that uses modifications of safety-critical situations from real-world driving is counterfactual simulations (Davis et al., 2011). Counterfactual simulations typically use crashes (baseline scenarios) and compare them to the simulated driving situations (treatment scenarios). More specifically, in counterfactual simulations, ADASs and ADSs are used to complement or replace the road users of the baseline scenario to answer the question: What if, in this safety-critical situation, the vehicle(s) had been equipped with the system? If the outcome is better than what happened in the real world (the baseline), the system is considered to improve safety for that specific driving situation. Of course, what is meant with the terms "outcome" and "better" needs to be defined, and they 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 in a different way. For example, delta-vs can be obtained from the impact speeds (Kullgren et al., 2003), or the risk of injuries can be predicted (Gennarelli & Wodzin, 2006; Kullgren, 2008)—either metric can be compared between baseline and treatment.

Some virtual safety assessment methods also need driver behavior models (ISO 2021). A driver behavior model is here defined as a mathematical representation of driver behavior. In virtual safety assessment methods, such mathematical driver models are used to represent the behavior of drivers in traffic. This may include everyday driving (e.g., car-following), models of crash causation (e.g., off-road glance behavior), and driver responses to critical events (e.g., hard braking to a lead-vehicle braking hard). A combination of these models can then be used to virtually generate crashes, although it must be made sure that it represents what it is supposed to represent (i.e., that the generated crashes aligns with the evaluation scope, Fahrenkrog et al. (2024), and are validated, (Bärgman et al., 2024)). For treatment simulations, there are also models that describe how drivers would respond to a system, for example a forward collision warning (FCW). For counterfactual simulations specifically, part of the trajectory can be replaced by behavior models, generating alternatives of what actually happened. Note that, driver models can be used in two of Wimmer et al.'s approaches (B and C2; see Fig. 1), as part of the baseline generation process (Wimmer et al., 2023). Many different driver behavior models can be found in the literature (see Chapter 2 for a more detailed description). One type of driver model that has recently received attention is the one-driver reference driver model (Rothoff et al., 2019); a model type that also is included in regulations (UNECE, 2023). The term "reference driver model" refers to a mathematical representation of one human driver or a population (distribution) of human drivers; the model is considered a benchmark (or a reference) for safe driving. With this definition, reference driver models represent the level of safety performance that a human can reasonably achieve. A one-driver reference driver model describes how one driver with some specific skill set drives, rather than describing how a population of drivers drives (Markkula et al., 2016; Pelella et al., 2023; Svärd, Bärgman, et al., 2021)—the latter can also be called a population reference driver model.

As ADSs will be expected to take control of the driving task without supervision, they will in essence replace humans—at least under specific conditions. ADSs developers could, therefore, use reference driver models as part of their virtual safety assessment chain in order to compare the safety performance of ADSs against what they aim to replace: humans. The reference model's performance should be the target to reach and preferably exceed (ISO 2020; Wood et al., 2019).

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. As an example, one-driver reference driver models have been proposed in R157 (see Section 2.3.3 for a more detailed description of the models). The models in R157 are stated to be mathematical representations of a competent and careful human driver, and they are used to provide guidance to define avoidable and unavoidable crashes in three types of scenarios (UNECE, 2023). The first is a deceleration (rear-end) scenario, in which the ego vehicle (the vehicle downstream in traffic, controlled by the driver model) and the principal other vehicle (POV) 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, in which the POV moves into the ego vehicle's lane from an adjacent lane, and the third is a cut-out scenario, in which 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 highD (Krajewski et al., 2018) safety-critical scenarios, finding, among other things, that one of the models often reacted to the lateral perturbations of surrounding vehicles, even when they were not changing lanes. These scenarios, however, may not be high-risk enough to be useful for a thorough validation of reference driver models.

1.1 Aims and objectives

The overall aim of this PhD project is to assess reference driver models and contribute to their development for use in virtual safety assessments, which in turn can be used in the development of ADASs and ADSs and in the approval of ADSs. The aims of this licentiate are to investigate the importance of data choice for accurate safety assessment and modeling and to assess reference driver model validity.

To achieve the aims of this licentiate work, two objectives were set:

- To quantify the influence of the choice of data in pre-crash virtual safety assessment, comparing the simulation outcomes of a simple scenario generation process across crashes, near-crashes, and everyday driving data.
- To determine whether the reference driver models defined in the R157 represent a competent and careful driver, by applying them to near-crashes from an NDS.

Further work will look into the design of reference driver models in more detail. New strategies for answering the question: What is a reference driver? will be explored. Specifically, comfort zone boundaries (CZBs) will be investigated as a way to set limits for what drivers feel comfortable while driving—to define the moment at which a competent and careful reference model would act to an evolving traffic situation. This has the potential to be improve future reference driver models.

2 Data, methods, and driver models

This section describes the data, methods, and driver models typically used in virtual safety assessments, and specifically in this work. First, an overview of some commonly used data for virtual assessment and modeling is presented, accompanied by a detailed description of the data used in Papers I and II. Second, the crash generation process and counterfactual simulations—the safety assessment method used in this work—are described in more detail. Finally, driver models, particularly the reference driver models used in Paper II, are described.

2.1 Data

The type of data used in virtual safety assessments plays a major role in the evaluation of systems and driver models. Two different types were used in this licentiate work: in-depth crash databases and NDSs. This section presents the key elements of both dataset types, emphasizing those that determine their suitability for use in virtual safety assessments.

2.1.1 In-depth crash databases

Data from in-depth crash databases are often used to virtually assess the crash-avoidance capabilities of systems and driver models. These databases consist of collections of detailed crashes, meticulously analyzed and reconstructed (estimated) to generate a digital version of the events leading to the crash, and of the crash itself (Bakker et al., 2017). Examples of in-depth crash databases are GIDAS in Germany (Otte et al., 2003), RAIDS in the UK (Cuerden & McCarthy, 2016), RASSI in India (Rameshkrishnan et al., 2013), CIDAS in China (Chen & Dai, 2018), and CISS in the USA (Zhang et al., 2019). Rather than confining itself to one country, IGLAD (Bakker et al., 2017) is a database that aims at harmonizing the crash databases from various countries. Data from in-depth crash databases are commonly used for virtual safety assessment, both as input data (Bjorvatn et al., 2021) and as validation data (Bärgman et al., 2024).

The in-depth reconstructed crashes used in this study come from GIDAS. In 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) is created. The PCM includes detailed pre-crash kinematics for up to five seconds prior to each crash (Schubert et al., 2013). The GIDAS rear-end crashes (i.e., when the ego and the POV are in the same lane for some time before the crash) 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 which the ego vehicle did not show signs of an evasive braking maneuver. The latter were extracted as those were the concrete scenarios that are the most similar to modeling of unresponsive drivers.

2.1.2 Naturalistic driving studies

NDSs are another common data source for virtual safety assessment. They can be divided into on-site and in-vehicle NDSs. On-site NDSs consist of location-based data collections, typically using a camera at a fixed location or on a drone and recording 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 typically collected from vehicles equipped with additional 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 situations that occur can be captured, regardless of the physical location.

NDSs can contain crashes, albeit rarely (van Nes et al., 2013). Further, there is less information for some crash aspects (e.g., injury outcomes and forces exchanged during the crash) than is available for in-depth reconstructed crashes. However, unlike in-depth reconstructed crashes, NDSs include recordings that can be used to extract time-series data of the kinematics of the instrumented vehicle and other road users (Hankey et al., 2016; Krajewski et al., 2018). These data provide more accurate descriptions of the vehicle trajectories than those obtained from crashes reconstructed from in-depth crash data, which are mostly based on assumptions rather than recordings. NDSs also include other safety-critical driving situations, such as near-crashes. These situations need to be identified and separated from the many hours of uneventful driving typically collected by NDSs. This work is usually done by automatic kinematic triggers (Hankey et al., 2016), which activate when harsh braking is detected or when surrounding vehicles get unusually close to the instrumented vehicle (to name two examples). Additionally, drivers can manually flag events they consider relevant with a button press. Expert annotators also play a role in identifying and classifying safety-critical events. NDSs also include non-safety-critical driving, which can be used to study false activations of systems—but that is beyond the scope of this work.

This work uses data from two different NDSs. The first is the Strategic Highway Research Program 2 (SHRP2), a large NDS collected over two years in the USA (Blatt et al., 2015; VTTI, 2024). The dataset includes trips made by 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 CAN bus were also available. The data, collected between 2010 and 2013, are still 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). The highD data, collected between 2017 and 2018, consist of recordings made by drones over sections of German highways. There were 60 recordings with an average duration of 17 minutes each. The trajectories of 110,000 vehicles were extracted from the recordings.

2.2 Methods

2.2.1 Generation of crashes from non-crash data

As mentioned, the type of data used in virtual safety assessment depends on the type of assessment to be made. For instance, the baseline for the virtual safety assessment of crash avoidance systems (e.g., AEB) is typically a set of crashes to which the system is then virtually applied. As described in the Introduction, the crashes can be generated by a variety of methods (Wimmer et al., 2023). In general, the main objective of several of these methods is to increase the number of baseline cases in order to overcome the scarcity of relevant original data; critically, the generated cases must be relevant and realistic. There are two main approaches that use statistical modeling methods to increase the number of baseline crashes: traffic-simulation-based and in-depth-database-based. The former uses extensive simulations that aim to use road users' behavior models to recreate their interactions in traffic scenarios. These simulated interactions sometimes (albeit rarely) result in crashes. In this approach, the road-users' behaviors are usually based on stochastic behavior models derived from NDSs (Feng et al., 2021; Li et al., 2019). In the second approach, in-dept crash databases are used to generate the baseline (Bärgman et al., 2017; Scanlon et al., 2021). Distributions of pre-crash kinematics from the real crashes are sampled and used to generate new crashes. (There are actually different ways new crashes can be generated—behavior-model-based or purely stochastic—but the latter is beyond the scope of this work; see Wimmer et al. (2023) for details.)

Both of these statistical approaches have limitations. The traffic-simulation-based approach requires lengthy, resource-intensive simulations to generate enough crashes across the relevant severities. As a result, safety surrogate measures are often used instead of traffic simulations (Åsljung et al., 2021; Westhofen et al., 2023).

The second approach, in-depth-database-based assessment, on the other hand, is limited by the scarcity of crashes in the databases for use in scenario generation. Additionally, crash databases tend to be biased towards higher-severity crashes (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). This bias could 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), in particular when used for validating the generation of additional cases. To overcome the limitations of this approach, (Wu, 2024) proposed a method that combines data from NDSs with pre-crash data, creating a set of models that generate crashes representative of real-world crashes across all levels of severity.

Paper I investigates the challenges associated with increasing the number of baseline cases by using counterfactual simulations to generate rear-end crashes from SHRP2 and highD car-following near-crash scenarios. When generating baseline crashes using behavior models through counterfactual simulations, any evasive maneuver by the original driver would need to be removed, to have a "clean slate" for applying the driver model. Using the same computational behavior models for the baseline generation and the treatment ensures that any positive or negative effect of the system is due exclusively by 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. To truly replace the evasive behavior of the driver, the problem here is deciding at what point the original driver's evasive actions are to be replaced by those of the model: too soon, and the modified event may differ greatly from what originally happened—the link to a real-world case gets weaker; too late, and the model may start intervening after the original evasive maneuver onset—the baseline is not a "clean slate" anymore. Typically, to avoid this problem the onset of the original driver's evasive maneuver is identified, and the following trajectory is removed (Bärgman et al., 2017; Bärgman et al., 2015; Scanlon et al., 2021). It is replaced with a new trajectory—the result of assumptions and simulation design choices. One example of an evasive maneuver is the driver's braking reaction (see Fig. 2A). Eliminating the original trajectory after its onset (Fig. 2B) means that the ego vehicle in the simulation typically keeps the speed constant instead of slowing down harshly (actually, typically with constant speed—without any deceleration); the original driver is effectively replaced with an unresponsive one. This procedure was used for Papers I and II.

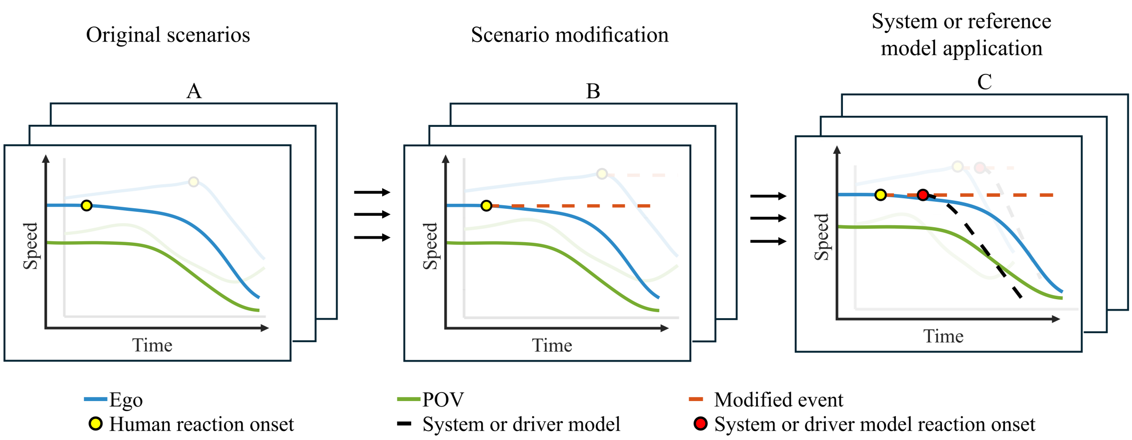

Fig. 2 Counterfactual modifications of rear-end traffic scenarios by removing the evasive maneuver and replacing the original maneuver of the ego vehicle with a model (of a system or reference model), but keeping the kinematics of the POV.

2.2.2 Counterfactual simulations for ADAS and ADS assessment

This licentiate 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 can be distinguished, they also share similarities. Approaches A and B both use counterfactual simulations for the prospective safety evaluation of systems. That is, they apply a system to modify the kinematics in a set of concrete baseline scenarios, and evaluate the system's safety benefits by comparing scenarios with and without the system. However, unlike approach A, approach B also uses counterfactual simulations when generating the baseline. That is, some sets of original scenarios are modified to create counterfactual versions. The system under assessment is then (counterfactually) applied. Fig. 2 conceptually illustrates counterfactual simulations applied to concrete rear-end crash scenarios. As mentioned in the Introduction, the system or model that replaces the behavior of the ego vehicle modifies its kinematics, while leaving the kinematics of the POV unmodified.

Counterfactual simulations have been used to assess ADASs (Erbsmehl, 2009) and ADSs (Bjorvatn et al., 2021; Scanlon et al., 2021). However, counterfactual simulations can also be used to evaluate driver behaviors (Bärgman et al., 2017; Bärgman et al., 2015; Lee et al., 2018). The studies that are part of this licentiate thesis used counterfactual simulations with baseline approach B from Wimmer et al. (2023): using data from crashes, near-crashes and everyday driving, systems and driver models were applied to the modified 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: choosing a moment to apply the system under assessment (treatment) that does not fundamentally change the configuration of the safety-relevant event. That is, the application of a system, if not carefully thought through, can undermine the validity of the simulations, and consequently the safety assessment. As an example, imagine that a system including adaptive cruise control (ACC) is applied to the pre-crash kinematics of rear-end crashes—which only include data for a few seconds (typically five) 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). In other words, a vehicle equipped with ACC would have never ended up in a traffic scenario described by the few seconds of pre-crash kinematics typically available for analysis, but instead it would have started slowing down earlier and a conflict could have been completely avoided, without emergency actions (e.g., by an AEB system).

2.3 Reference driver models for ADS safety assessment

This licentiate thesis is part of a project which aims to contribute to the development of reference driver models. As a first step in that direction, the study in Paper II evaluated the validity of two existing computational reference driver models available in R157. This section briefly introduces computational driver models (with particular attention to reference driver models) and describes their role in ADS safety assessment. More details about the two reference driver models assessed in Paper II are also provided.

2.3.1 Computational driver models

As described earlier, computational driver models are mathematical representations of human drivers. These models often aim to simulate a human driver’s ability to avoid conflicts and crashes (see Introduction for the definitions of conflict and crash avoidance). The models can include reactive actions (e.g., evasive braking or steering—crash avoidance) and proactive actions (e.g., predicting possible conflicts in the near future and adjusting the vehicle's controls accordingly—conflict avoidance). Over the last few decades, many of these models 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). Driver models may either be intended to represent a single driver (e.g., an average driver or a competent and careful driver) or to capture driver variability (i.e., represent some group of drivers; Svärd, Bärgman, et al., 2021; Webb et al., 2020). The latter is often done through the use of distributions of model parameters (Markkula et al., 2016; Rasch & Dozza, 2020).

2.3.2 General description of reference driver models for ADSs

Driver models can generate a human benchmark for comparison with ADSs (Rothoff et al., 2019; Webb et al., 2020). This comparison supports one of the key concepts of ADS safety assessment: achieving a positive risk balance (PRB; Di Fabio et al., 2017). 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). Certainly, an ADS should cause fewer crashes than the average human driver (ISO 2020; Wood et al., 2019); the average driver's performance may not be considered good enough for ADSs safety assurance. Instead, 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. However, it is not obvious how to construct such models, and the results from Paper II indicate that more research is needed to create (and validate) valid reference driver models that are competent and careful (or equivalent).

A different approach to ADS safety assessment is using rule-based models that are not specifically based on human behavior, such as the responsibility-sensitive safety (RSS) model. 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; that is, in a safe position, the ego vehicle can always avoid causing a collision. If all the vehicles respected these rules, there would be zero crashes. The RSS, however, is not grounded in human behavior. It is therefore problematic to use the RSS rules directly in the algorithms for reference driver models.

In summary, the safety assessment of ADSs includes many components, one of which is the comparison of the ADSs' safety performance against human reference models. The key concept here is the use of humans as a reference—rather than what is objectively safe.

2.3.3 The UNECE reference driver models

The two reference driver models used in Paper II are in R157 named "Performance model 1" and "Performance model 2". 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 time to collision (TTC) and lateral position of the POV as metrics for the assessment of possible threats, and (if needed) reacts by braking. For a cut-in scenario (the only scenario considered in Paper II), the CCDM defines a "wandering zone". This zone is centered in the POV's lane; its width is the width of the POV plus an additional 0.375 m on both sides of the POV. The extra width allows for small lateral corrections, which are to be expected in normal lane-keeping. 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 time to collision is less than 2 s. The reaction time of the CCDM is divided into 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 predicts the possibility of a collision based on lateral and longitudinal safety checks. When it is predicted that the trajectories will overlap, the FSM applies the brakes. The braking reaction of the FSM does not always reach the maximum braking capability of the driver. That is, the model is capable of braking with any deceleration value between 0 and the maximum reachable deceleration (set to 6 m/s^2). The actual 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 (up to 1) based on the criticality of the scenario. The required deceleration increases gradually with the PFS, from 0 m/s^2 when PFS = 0 to 6 m/s^2 when PFS = 1. If the situation is critical and a collision is imminent, the CFS increases and the maximum required deceleration is reached with a jerk of 12.65 m/s^3.

3 Summary of papers

This section presents a summary of the included 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
ADASs can help reduce the number of crashes worldwide. However, it is not easy to quantify the safety of ADASs that are not yet on the market, as a considerable amount of data from real-world crashes is required. Previously, the use of crashes artificially generated from NDS has been proposed to meet this need, as NDS can typically provide much more data than in-depth crash databases. However, these artificial crashes should be validated against real crashes.

Method
Crashes were generated from two non-crash NDS datasets, SHRP2 and highD, by replacing the original driver at the time of the evasive maneuver with an unresponsive (sleeping) driver, resulting in a rear-end crash. Then, AEB systems were applied to the real and generated crashes.

Results
There were substantial differences between the real and the generated crashes. SHRP2-generated crashes and GIDAS crashes showed similar levels of criticality, while highD-generated crashes were less critical. Crashes generated from the NDS datasets did not match the level of severity of real crashes, since the AEB application avoided a higher percentage of generated crashes than real crashes.

Conclusions
This work studied the validity of crashes generated from non-crash NDS datasets for use 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 this process is *not* a viable option for increasing the data in virtual safety assessments. On the other hand, crashes generated from SHRP2 showed levels of criticality similar to that of real crashes. This result means that SHRP2 is more suitable for generating crashes than highD, even though more research is still needed in order to make the crashes more realistic.

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. (submitted)

Introduction

As ADSs are being developed, virtual safety assessment methods are being adapted to accommodate the shift from ADASs to ADSs assessment. One aspect of ADS safety assessment that has been proposed involves the use of reference driver models as a safety benchmark for ADS performance. That is, one of the requirements for ADS approval would be that the system is at least as safe as a competent and careful human. This work investigates the validity of the two reference models described in the UNECE Regulation No. 157 by evaluating their safety performance when applied counterfactually to near-crash traffic scenarios.

Method

The models were applied using counterfactual simulations to 38 near-crash cut-ins from SHRP2. First, videos of the near-crashes were manually annotated to extract the vehicles' trajectories. Then, the original braking evasive maneuver performed by the SHRP2 drivers was removed. The models were then counterfactually applied to the modified SHRP2 events.

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. The second model's reaction preceded the original onset of the evasive maneuver by 0.7 s on average. Noticeable differences between the models and the SHRP2 drivers were also found by analyzing the lateral position of the POV in the lane: the first model reacted well after the start of the POV's lateral motion to change lanes—the POV was closer to the lane mark than it was when the second model reacted.

Conclusions

This study evaluated the validity of two reference driver models described in R157. The first model was not careful enough, as it caused crashes that had not happened in the original real-world scenarios. The second 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. Main takeaways from this work include that more work is needed in the development of reference driver models for ADS assessment, and that, once developed, these models need to be validated on data that ranges across all levels of criticality, from everyday driving data to high-severity crashes.

4 Discussion

This PhD project aims to evaluate and develop reference driver models for virtual safety assessments, aiding ADAS and ADS development and approval. This licentiate focuses on the impact of data choice on safety assessments and the validity of reference driver models. Section 4.1 discusses the specific challenges presented by different data sources in creating a virtual assessment baseline, given how different types of data can impact the relevance of the virtual assessment results. Section 4.2 discusses the components that are relevant for the development of reference driver models and reflects on possible improvements to the models currently used in ADS safety assessments. Finally, Section 4.3 describes the limitations and future directions of this research.

4.1 The importance of data choice in virtual safety assessment

Virtual safety assessment methods for ADASs and ADSs are considered key for the systems' development and release on the market. These systems are intended to operate on public roads, so the assessment needs to be based on data from real-world traffic. However, the specific real-world traffic scenarios must be relevant for the type of assessment performed. For ADASs, the baseline situations selected are typically highly safety-critical. On the other hand, all criticalities—from everyday driving to critical situations—are relevant in the baseline for ADSs, which are responsible for the complete driving task. This section focuses on one of the aims of this licentiate work: quantifying the impact of the choice of data on the safety impact assessment of pre-crash safety systems. In practice, this means that I studied the impact of data of different levels of criticality (i.e., everyday driving, near-crashes, and crashes) on baseline scenario generation, and on the subsequent virtual safety assessment of an ADAS system. This section starts with a discussion of the need for generating crashes from non-crash data, as an alternative to data from in-depth crash databases. Second, the approach to generating crashes used in this work is discussed. Third, and finally, the validity of crash-generation approaches in general is discussed.

4.1.1 Crash generation in Paper I

Part of the aim of this licentiate 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., 2015; Wu & Jovanis, 2012). However, they can be used as surrogates only if appropriate sampling and validation methods are employed, and then only within a scenario type (Knipling, 2015). Everyday driving data, on the other hand, captures a much broader set of traffic scenarios, most of which are less relevant for safety than near-crashes and crashes, particularly when assessing an ADAS. Naturally, when assessing an ADS, which should be able to perform everyday driving in a safe way, less critical scenarios should be included in the assessment.

In an attempt to better understand the relevance for traffic safety of different types of data, the study in Paper I used traffic scenarios from SHRP2 and highD to generate crashes. The crashes were then compared to reconstructed real-world crashes from the GIDAS database (Otte et al., 2003). Paper I used the assumption of an unresponsive driver to generate worst-case scenario crashes from rear-end near-crashes (from SHRP2) and (more or less) normal driving data (from highD). The unresponsive driver replaced the driver of the following 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, even though theoretically the resulting crashes could be even more severe if the driver accelerates, because the probability of that situation occurring in the real world is low, as shown by Wu et al. (2024). The results of the comparison between crashes artificially generated from non-crash events and real crashes indicate that they are profoundly different in their outcome severity. In the real-world crashes, the vehicles had much larger speed differences, and the POV performed a stronger braking maneuver. Crashes generated from highD data, which at most contained very minor conflicts, were 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 the POV initiated braking was substantially higher in the highD-generated crashes than in the SHRP2-generated crashes. One reason for this difference is, as previously mentioned, likely to be found in the deceleration values reached by the POVs in the highD data, which were much lower (less harsh braking) than the ones reported in the real SHRP2 crashes.

However, it is not only the low deceleration of the POV that differs between the real-crashes and the crashes generated from the highD 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); the ego vehicle does not react early enough to the presence of the POV. The insufficient reaction may be due to, for example, 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 cannot be created, as only the situations with the POV decelerating harder than 2 m/s^2 were used for crash generation. However, it is not obvious how to generate catch-up crashes from everyday driving data such as highD, especially if the data consist of mainly dense traffic, where almost all vehicles are just following one another. The vehicles tend to travel at similar speeds and with relatively short “car following” time headway (THW); therefore, the severity of any generated catch-up crashes tends to be low. An exception to this tendency is the “end of traffic jam” scenario, that consists of a catch-up crash in which, for example, a vehicle encounters a traffic jam after having been driving at free-flow speeds. This scenario can, however, be 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. Such data would be valuable for future studies.

In Paper I, the validity of the generated crashes was assessed through the application of AEB systems. The AEBs' performance, measured in terms of crash avoidance and mitigation, showed substantial differences across different crash datasets. These differences were part of the argument against using that particular set of generated crashes. However, the research field of scenario validation is certainly larger than what is within the scope of Paper I. Accordingly, the next section considers the results of Paper I from the broader perspective of the scenario validation research field.

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

As mentioned, one challenge of using crashes generated from near-crashes (and even lower-severity conflicts) in virtual safety assessments is that the generated crashes do not capture the criticality of real crashes. This in turn means that the systems (like AEB in Paper I) avoid the generated crashes more easily, since they have more time to assess the threat and react to it. These results raise the questions: Are counterfactually generated crashes valid for 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: combining sleepiness with crashes generated by considering distributions of off-road glances and with crashes generated by the ego drivers not braking as hard as they could (given the physical limitations of the situation).

Generating crashes without assuming the drivers are unresponsive means that the crashes will be less severe. For example, crashes generated using a distribution of off-road glances would have drivers that react more quickly to a critical situation than crashes using a model of an unresponsive driver. Thus, the crashes generated using off-road glances (as in Bärgman et al., 2024) would be less frequent and severe than unresponsive-driver-generated crashes. Actually, Fig. 4 in Bärgman et al. (2024) shows that the crashes generated with the more complex crash-causation model have about half the delta-v of the crashes with the unresponsive drivers. This means that if a more realistic crash generation model had been used in Paper I, those generated crashes would have been even more different from the real-world crashes. It should also be noted that Victor et al. (2015), followed by Kusano and Victor (2022), leveraged the concept of an unresponsive driver in counterfactual simulations to define the metric of "maximum injury potential" of a traffic scenario—a metric that in its simplest form is derived by assuming an unresponsive driver. This metric enables comparisons of scenarios' potential severity, so that the relevance of various databases in virtual safety assessment can be demonstrated. Overall, the 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.

So far, this discussion has considered the fact that crashes generated counterfactually from everyday driving data do not seem to be realistic. However, as mentioned, other methods can be used to generate crashes based on everyday driving data, for example by using traffic simulations with behavior models based on everyday driving (Feng et al., 2021; Li et al., 2019). Can such methods generate realistic crashes? Although this work does not address that specific question, it is possible to speculate based on the results of Paper I. 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 amount of harsh deceleration maneuvers appears to be quite low—even though the highD dataset used in Paper I included trajectories from as many as 110,000 vehicles. If behavior models for traffic simulations are based on deceleration distributions such as from highD, it is hard to see how the simulations could generate realistic crashes (and whether they would generate crashes at all).

Regardless of the method used to generate crashes, what is truly important here is determining which metrics the generated crashes are valid for (and validated on). 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). I have not found any studies using the traffic-simulation-based approach that validated their results with respect to outcome severity that is truly related to safety (i.e., delta-v and injury risk). One of the most prominent works on traffic-simulation-based scenario generation is Feng et al. (2021; published in Nature Communications), but even they did not validate their method on severity outcome; it was validated only on crash frequency. Before traffic simulations are used for virtual safety assessment, they need to be validated on scenarios ranging from everyday driving, via near-crashes, to severe crashes—not least because of the issues of using everyday driving and near-crash data for scenario generation identified in Paper I. Do they really generate representative crashes across all levels of outcome severity? This question is important as the World Health Organization and the Vision Zero concept consider severe injury and fatalities to be what is important when addressing traffic safety (Tingvall et al., 2020; p. 9; WHO, 2021).

If everyday driving data cannot be used to generate realistic crashes across all levels of severities—and today we do not know whether they can when it comes to traffic-simulation-based scenario generation—what is to be done? Either more studies generating crashes across all severities from multiple data sources, such as the one by (Wu et al., 2024), are needed, or some way to “get from” low-severity scenarios to higher-severity scenarios is needed. This is where extreme value theory (EVT) may come in. In the traffic safety context, EVT can be used to find the tails of distributions of safety-related metrics, such as levels of deceleration, TTC, post-encroachment time (PET), and brake threat number (BTN), but also delta-v and injury risk. For example, EVT was applied to estimate the safety of intersection traffic scenarios using PET (Songchitruksa & Tarko, 2006) and injury severity (Arun et al., 2021). More generally, it was also used to assess the relevance of surrogate measures when predicting the frequency of collisions (Åsljung et al., 2017). However, these studies did not use EVT to generate the pre-crash time-series of crashes, instead they extrapolated the frequency and characteristics (e.g., TTC, delta-v) of crashes from near-crashes. More research is required to determine how to use EVT to generate crashes for virtual safety assessment, at which point the crashes will need to be validated on crash outcome severity.

4.2 The role of reference driver models in ADS development and deployment

This section discusses the components of reference driver models, and more generally the role of such models in the development and safe deployment of ADSs. When discussing reference driver models, a question that naturally arises is: What should a reference driver model represent? It was defined in this work as a mathematical representation of a human driver that is considered a benchmark for safe driving. However, this definition is far from a practical formulation of a model, which requires parameters and conditions that describe its behavior. To implement a model that is to represent human drivers, human driving behavior needs to be studied and described with mathematical equations. The discussion in this section aims to relate design choices in computational modeling of driver behavior to characteristics that are typical of human responses in real traffic. The section starts with a discussion of the results of Paper II. The findings are then linked to Paper I and contextualized within the wider scope of ADS safety assurance.

4.2.1 The UNECE models' performance

Paper II analyzed 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 in R157 are to be used as "guidance" for defining preventable and unpreventable crashes in specified traffic scenarios. It is not clear what "guidance" means here; there is a danger that implementers of ADSs could use these models as the actual target: "If our systems perform better than the models, we are good"—which, as we argue in Paper II, is a highly problematic approach.

In Paper II, the models were counterfactually applied to near-crash cut-ins from SHRP2 to evaluate their performance compared to the SHRP2 driver. 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 avoided a crash 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 not human-like. The results from Paper II raise questions about the capability of such models to accurately capture human behavior. First, a reference driver model should avoid basically all the crashes that were avoided by humans that the model is supposed to represent, such as a competent and careful driver. Second, the driver models should also behave in a predictable and non-erratic manner. A driver model that reacts to any risk of collision, even if far in the future, may not be representative of realistic human driving behavior—even a competent and careful one. Actually, there is a risk that such a model may avoid crashes normally unavoidable by humans, by reacting early in the scenario to some "cue" that a human would ignore. That is, the model could be acting on some benign, in-lane lateral perturbations by the POV, resulting in it missing the cues that actually cause reactions in human drivers (as it is already avoiding the crash at this time). Note that 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 potentially influence the way that ADSs are developed. Specifically, the FSM could set unrealistic requirements for ADSs and perhaps even impact ADS deployment (i.e., a system that is actually good enough might not be released).

4.2.2 Urgency, surprise, and comfort zones

The response of the UNECE models, and of crash avoidance-related driver models in general, to specific traffic situations is characterized by two main components. One is the situation criticality, referred to in this work as "urgency"—which dictates the type and intensity of the response to an unexpected event. The other component defines the start of the unexpected event, which is the moment in time when the model notices that something is not right, and is referred to in this work as "surprise".

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), BTN (Brannstrom et al., 2008), deceleration rate (Allen et al., 1978; Darzentas et al., 1980), and PET (Allen et al., 1978). These metrics can be calculated using the kinematics of the involved road users. Driver actions (e.g., braking or steering) are often based on threshold values of these metric. These thresholds can be defined to distinguish between critical and non-critical driving scenarios. However, defining these thresholds is not straightforward. It should be noted that although just applying thresholds to the basic metrics is done, there are often more elaborate modeling approaches to better computationally describe human behaviors. One such approach for driver model development is evidence accumulation. In this approach, models integrate information—such as some visual cue metric—over time, to predict the future behavior of another road user, for example. The onset of the reaction is dictated by the accumulation of visual evidence (what the road user observes). When the difference between prediction and reality becomes large enough, the reaction is initiated (Markkula, 2014; Markkula et al., 2018). Evidence accumulation models have been shown to be more human-like than simple threshold-based models (Markkula, 2014; Svärd et al., 2017).

One promising approach to defining urgency metrics thresholds uses the concept of comfort zones and their boundaries. The concept of comfort zones 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). Comfort zones and their 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 boundaries of the comfort zones, the limits of what the driver may consider acceptable without feeling discomfort, can be used to define thresholds in engineering-based metrics. For example, if a collision in a rear-end situation can be avoided with a slight deceleration—below the limit of comfortable braking—the situation would not be considered urgent. In both Paper I and in Yang et al. (2024), CZBs were used to tune and assess the ADAS AEB. 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 Paper I and in Yang et al. (2024) 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 situations, at least for most drivers. In Paper I, Fig. 5a 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, the driver of the ego vehicle 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 almost never reached a deceleration of 5 m/s^2, and the THW (Fig. 1 of Paper I) was rarely short enough to be considered unsafe.

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., CZB) perspectives, has already been demonstrated. All crashes are urgent by definition: at some point in time the driving situation becomes critical, and the driver is not able to avoid a collision even by braking well beyond the levels of comfortable crash avoidance. In Paper I, the SHRP2 near-crashes showed high levels of urgency.

In Paper II, urgency is also key in the assessment of the UNECE models when applied to cut-in scenarios. However, the two models implement the concept differently, and, based on the results of Paper II, not in the way that human drivers consider urgency. The CCDM uses TTC in the longitudinal direction to define the criticality of a situation, 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 CCDM assumes that a situation is critical as soon as a vehicle exceeds the wandering zone (see Sect 2.3.3). In one way this can be seen as the ego driver feeling comfortable as long as the POV does not leave the zone. The use of this wandering zone can be interpreted here as a form of driver satisficing (Summala, 2007): the ego driver does not need to optimize the vehicle's lateral positioning to stay in the comfort 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. In the analysis in Paper II, it was observed that this design approach did not adequately account for non-urgent situations (when a collision was not immediately predicted), since the FSM does not consider urgency in the timing. However, it uses urgency in the braking reaction itself: the level of deceleration is modulated based on the criticality of the scenario, performing slight decelerations when the scenario is not very critical.

The second component that characterizes the response of driver models in critical situations, surprise, could also lead to a reaction if the situation is urgent enough. The two UNECE models consider this component differently as well. The CCDM perceives a potentially dangerous cut-in situation by determining whether 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 driver). In contrast, the FSM bases its determination on whether the cut-in vehicle is moving laterally towards the ego vehicle, causing the future paths of both vehicles to encroach. Note that due to the way that the FSM is modeled, the situation does not need to be critical or unexpected for the model to start braking. This also means that maybe this is not necessarily a surprising situation. For example, for most drivers it would not be unexpected that a POV two lanes over changes lanes into the adjacent lane. Therefore you could even say that the FSM does not implement "surprise" as defined in this work. The FSM is more 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 situation is not very critical).

The "non-impaired road user with their eyes on the conflict" (NIEON) model, based entirely on the concept of surprise, was introduced by (Engström et al., 2024). 'Surprise' in their work, based on information theory (Shannon, 1948) and Bayesian surprise (Itti & Baldi, 2009), is defined as an observation that is not explained or predicted by the prior belief of 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". The dependency on context makes NIEON particularly interesting for the issue of "when to start the clock" of the reaction time. One aspect that is not covered in NIEON (at least the way it is described in the literature), however, is how the reaction can change based on the level of urgency of the situation—if it is not urgent there may not even be any reaction. This means that an obvious question is: What level of urgency is the ego driver willing to comfortably tolerate before an action has to be taken? Studies that test the boundaries of comfort zone in relation to scenarios with different levels of urgency could help fill this gap, and could be considered as a complement to surprise. Olleja et al. (2023), for example, used a percentile-based approach as a first step towards defining components for driver models based on CZBs.

4.3 Limitations and future work

This section explains the main limitations of the studies in this licentiate work. The first focus is on the limitations in the conceptualizations and actuation of the methods and the second on the limitations that affect the data. Finally, this section also discusses opportunities for future development related to this work.

The crash-generation process of Paper I relies on the assumption that the ego driver is completely unresponsive to the braking maneuver of the POV (basically sleeping). It is not realistic to assume that all drivers are sleeping (unless the object of study is the worst-case scenario of human driving behavior (Kusano & Victor, 2022). Adding additional 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 crash-generation process (Bärgman et al., 2024). However, Paper I demonstrates that, *even in the worst-case scenario* (unresponsive drivers), the lower-severity datasets do not show the high level of criticality and severity seen in the crash datasets. This is partially due to the more benign situation, but it should be noted that it is probably to a large extent due to the censoring of low-severity crashes in the crash data. That is, property-damage-only crashes are completely missing in GIDAS, and low-severity injuries are underreported. Future studies should investigate the contribution of benign situations (e.g., slight lateral movement within a lane, without the intent to change lane) and of the censoring (e.g., the effect of censoring of data for both modeling and validation) with respect to the difference in urgency and outcome severity. One option would be to apply a "transformation" to the generated crashes to account for the censoring of GIDAS—making the crashes comparable (Bärgman et al., 2024).

Papers I and II are limited to one type of driving scenario each. Paper I focused on rear-end scenarios only, primarily because the AEB functions applied to the crashes were designed to work in those scenarios. Future work could focus on other scenarios, such as cut-ins and cut-outs.

Paper II focused on assessing the UNECE models for cut-in scenarios only. This was a deliberate choice 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 were limiting the number of cases extracted. The SHRP2 subset composed of cut-in near-crashes had been gathered and annotated in a previous study (Chau & Liu, 2021). They were annotated using a tool developed by Shams El Din (2020). For Paper II, the annotation tool was refined and the annotation process was partially repeated in order to increase the data quality. The annotation process combined the ego vehicle's video footage from the front-facing camera with radar and speed data into a digital reconstruction of the near-crash.

Unfortunately, the poor quality of the 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 position of the POV. Future work could further improve the video annotation tools to increase the quantity and quality of data or use new, better datasets. The sheer scale of SHRP2 makes it unlikely that a similar study 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 ones in Paper II to those traffic scenarios.

Future PhD work will focus on studying the component of urgency in human behavior, to support the development of reference driver models. The next studies will aim specifically at quantifying CZBs as a way to define thresholds for satisficing behavior in driver models. specifically at quantifying CZBs as a way to define thresholds for satisficing behavior in driver models. Practical examples include the use of data from driver monitoring systems to determine how off-road glance behavior (and therefore inattention to driving) changes as a function of context (e.g., time-gap to the car ahead and traffic density). It is expected that driving with short time-gaps or in highly dense traffic incurs a higher percentage of on-road glances compared to driving with large time-gaps in low density traffic. The shift in glance behavior may be modeled, and the model could define the transition from a more attentive to a less attentive driver. The transition is here considered as a boundary that defines when drivers feel comfortable looking away from the road.

Overall, the continuation of this PhD work will investigate new ways to operationalize urgency for one-driver reference driver models. Operationalizing urgency entails defining quantitative values describing the level of urgency that have been extracted from the traffic scenario—and in turn define actions taken by the driver model. The work on urgency will also include a review of the substantial body of research quantifying perceived safety (e.g., He et al., 2024; He et al., 2022; Kolekar et al., 2020; Prasetio & Nurliyana, 2023), as it is a research domain closely related to CZBs.

5 Conclusions

The research work presented in this licentiate thesis contributes to the field of traffic safety by supporting the development of virtual assessment methods for ADASs and ADSs. The thesis discusses the relevance of various types of data typically available for virtual safety assessments. It then focuses on the role that reference driver models have in the safety assessment of ADSs, assessing existing models and projecting the results into future work planned for the PhD project. This work strengthens the understanding of how different design choices can impact virtual safety assessments. One key takeaway is the importance of using relevant, valid practices. Specifically, it was found that crashes generated from everyday driving data are quite different from real-world crashes, and therefore they should be used with care in virtual safety assessments. It was also found that reference driver models should be properly validated across the full severity range of traffic scenarios before they can be considered for use in regulations.

The results of Paper I highlight the difficulties of generating crashes from non-crash data in NDSs. Increasing the amount of data available for the virtual assessment—at least with the methods used in this work—comes with the risk of generating a baseline that does not reflect the severity or criticality of real-world crashes. Such a baseline is problematic, since it potentially results in erroneous conclusions.

Paper II focuses on the assessment of two existing reference driver models, described in UNECE Regulation No. 157. One model reacted to safety-critical cut-ins later than humans, and even caused some crashes. The other model typically reacted earlier than humans, avoiding crashes but also resulting at times in behavior that was overly careful and not human-like. The results constitute a first step towards understanding what components are relevant to this type of driver model, to be used specifically as a benchmark for ADSs.

Future developments of this work will continue to pursue the overarching goal of improving virtual safety assessment methods. Specifically, future work will investigate if and how the use of CZBs can be used to define quantitative values of the urgency of a traffic scenario. This has the potential to contribute to the development of reference driver models.

References

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

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. *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, Detroit, MI, USA. https://www-nrd.nhtsa.dot.gov/pdf/ESV/Proceedings/25/25ESV-000346.pdf

Antony, M. M., & Whenish, R. (2021). Advanced driver assistance systems (ADAS). In *Automotive Embedded Systems: Key Technologies, Innovations, and Applications* (pp. 165-181). Springer. https://doi.org/10.1007/978-3-030-59897-6_9

Arun, A., Haque, M. M., Bhaskar, A., Washington, S., & Sayed, T. (2021). A bivariate extreme value model for estimating crash frequency by severity using traffic conflicts. *Analytic methods in accident research*, *32*, Article 100180. https://doi.org/10.1016/j.amar.2021.100180

Åsljung, D., Nilsson, J., & Fredriksson, J. (2017). Using extreme value theory for vehicle level safety validation and implications for autonomous vehicles. *IEEE Transactions on Intelligent Vehicles*, *2*(4), 288-297. https://doi.org/10.1109/TIV.2017.2768219

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

Bakker, J., Jeppsson, H., Hannawald, L., Spitzhüttl, F., Longton, A., & Tomasch, E. (2017). IGLAD-International harmonized in-depth accident data. *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, Detroit, MI, USA. https://www-nrd.nhtsa.dot.gov/pdf/ESV/Proceedings/25/25ESV-000248.pdf

Bärgman, J. (2016). *Methods for analysis of naturalistic driving data in driver behavior research* [Doctoral thesis, Chalmers Tekniska Hogskola]. 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., 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

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 in the L3pilot project)*. 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., & Hoover, M. (2015). *Naturalistic driving study: Field data collection*. 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. *2020 IEEE Intelligent Vehicles Symposium (IV)*, Las Vegas, NV, USA, 1929-1934. 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

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

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

Chau, C., & Liu, Q. (2021). *Driver Modelling for Virtual Safety Assessment of Automated Vehicle Functionality in Cut-in Scenarios* [Master's Thesis, Chalmers University of Technology]. https://hdl.handle.net/20.500.12380/304058

Chen, Q., & Dai, B. (2018). *Characteristics and casualty analysis of two-wheeler accidents in China, data source: the China in-depth accident study (CIDAS)* (0148-7191). SAE Technical Paper. 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

Cuerden, R., & McCarthy, M. (2016). *The methodology and initial findings for the Road Accident In Depth Studies (RAIDS) Programme*. 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. (2006). *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment* (DOT-HS-810-593). Virginia Tech Transportation Institute. 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. *International encyclopedia of transportation*, *7*, 20-38. 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., 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*, 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. *21st International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. https://www-nrd.nhtsa.dot.gov/pdf/ESV/Proceedings/21/09-0162.pdf

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

European Commission (2020). *Ethics of Connected and Automated vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility*. Publication Office of the European Union: Luxembourg. https://data.europa.eu/doi/10.2777/035239

Fahrenkrog, F., Fries, A., Das, A., Kalisvaart, S., Charoniti, E., Op den Camp, O., Sander, D., Bourauel, B., Skudlarek, H., Bärgman, J., Urban, M., Menzel, T., Glasmacher, C., Chajmowicz, H., Davidse, R., Denk, F., & Lorente Mallada, J. (2024). *Draft Prospective Safety Assessment Framework - Method*. Milestone Report MS10 of the Horizon Europe project V4SAFETY.

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), 748, Article 748 (2021). 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

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

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

Gulino, M. S., Fiorentino, A., & Vangi, D. (2022). Prospective and retrospective performance assessment of Advanced Driver Assistance Systems in imminent collision scenarios: the CMI-V r 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

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

He, X., Happee, R., & Wang, M. (2024). A new computational perceived risk model for automated vehicles based on potential collision avoidance difficulty (PCAD). *Transportation research part C: emerging technologies*, *166*, Article 104751. https://doi.org/10.1016/j.trc.2024.104751

He, X., Stapel, J., Wang, M., & Happee, R. (2022). Modelling perceived risk and trust in driving automation reacting to merging and braking vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *86*, 178-195. https://doi.org/10.1016/j.trf.2022.02.016

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*. Lund University, Department of Traffic Planning and Engineering.

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

ISO (2020). *Road vehicles — Safety and cybersecurity for automated driving systems — Design, verification and validation* (ISO/TR 4804:2020). https://www.iso.org/standard/80363.html

ISO (2021). *Road vehicles — Prospective safety performance assessment of pre-crash technology by virtual simulation* (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

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* (FHWA-HOP-16-009). United States. Department of Transportation. Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/50639

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

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 (ITARDA). https://www.itarda.or.jp/contents/10235/23_06microPCM_e.pdf

Knipling, R. R. (2015). Naturalistic driving events: No harm, no foul, no validity. *Driving Assessment Conference 2015*, Snowbird, Salt Lake City, UT, 197-203.

Kolekar, S., de Winter, J., & Abbink, D. (2020). Which parts of the road guide obstacle avoidance? Quantifying the driver's risk field. *Applied ergonomics*, *89*, Article 103196. https://doi.org/10.1016/j.apergo.2020.103196

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. *IEEE 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2118-2125. 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. *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. 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. *Proc of Int. IRCOBI Conf., Bern, Switzerland*, 3-14. 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. *Proceedings of the 18th International Technical Conference on Enhanced Safety of Vehicles*. https://www-nrd.nhtsa.dot.gov/pdf/ESV/Proceedings/18/18ESV-000404.pdf

Kusano, K., & Victor, T. (2022). Methodology for determining maximum injury potential for automated driving system evaluation. *Traffic injury prevention*, *23*(sup1), S224-S227. https://doi.org/10.1080/15389588.2022.2125231

Laureshyn, A. (2010). *Application of automated video analysis to road user behaviour* [Doctoral Thesis (compilation), Transport and Roads]. https://portal.research.lu.se/files/5337337/1522970.pdf

Laureshyn, A., & Várhelyi, A. (2018). *The Swedish Traffic Conflict Technique: observer's manual*. https://portal.research.lu.se/en/publications/fc968207-ed2b-49cc-9346-57d8c74cac41

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

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), eaaw0863. https://doi.org/10.1126/scirobotics.aaw0863

Liers, H. (2018, December 19). *Traffic accident research in Germany and the German in-depth accident study (GIDAS)*. 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

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*, 181-207. https://doi.org/10.1007/s00422-017-0743-9

Markkula, G., Engstrom, J., Lodin, J., Bargman, 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

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
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
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
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. *Proceedings: International Technical Conference on the Enhanced Safety of Vehicles*, *2003*.
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 in the i4Driving project)*. https://i4driving.eu/wp-content/uploads/2024/07/D4.1.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
Prasetio, E. A., & Nurliyana, C. (2023). Evaluating perceived safety of autonomous vehicle: The influence of privacy and cybersecurity to cognitive and emotional safety. *IATSS Research*, *47*(2), 160-170. https://doi.org/10.1016/j.iatssr.2023.06.001
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. *Proceedings of IRCOBI conference*. https://www.ircobi.org/wordpress/downloads/irc13/pdf_files/13.pdf
Rasch, A., & Dozza, M. (2020). 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
Regulation (EU) 2018/858 of the European Parliament and of the Council of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, amending Regulations (EC) No 715/2007 and (EC) No 595/2009 and repealing Directive 2007/46/EC (Text with EEA relevance)Text with EEA relevance. https://eur-lex.europa.eu/eli/reg/2018/858/2023-07-30
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*. 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. 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. *5th International Conference on ESAR 2012*. https://bast.opus.hbz-nrw.de/opus45-bast/files/630/31_Schubert_Erbsmehl.pdf

Shams El Din, A. H. (2020). *Statistical modelling of critical cut-ins for the evaluation of autonomous vehicles and advanced driver assistance systems* [Master's Thesis, Chalmers University of Technology]. https://hdl.handle.net/20.500.12380/301054

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

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. *Driving Assessment Conference*, *5*(2009), 313-319. https://doi.org/10.17077/drivingassessment.1337

Songchitruksa, P., & Tarko, A. P. (2006). The extreme value theory approach to safety estimation. *Accident Analysis & Prevention*, *38*(4), 811-822. https://doi.org/10.1016/j.aap.2006.02.003

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

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

Tingvall, C., Michael, J. P., Larsson, P., Anders Lie, Segui-Gomez, M., Wong, S. V., Kobusingye, O., Krafft, M., Wegman, F., Peden, M., Hyder, A., Khayesi, M., Dumbaugh, E., Cockfield, S., & Furas, A. (2020). Saving Lives Beyond 2020: The Next Steps. In K. E. Björnberg, S. O. Hansson, M.-Å. Belin, & C. Tingvall (Eds.), *The Vision Zero Handbook* (pp. 789-839). Springer Cham. https://doi.org/10.1007/978-3-030-76505-7

UN Regulation No. 157. Uniform provisions concerning the approval of vehicles with regard to Automated Lane Keeping Systems, (2023). https://unece.org/transport/documents/2023/03/standards/un-regulation-no-157-amend4

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

Victor, T., Dozza, M., Bargman, J., Boda, C. N., Engström, J., Flannagan, C., Lee, J., & Markkula, G. (2015). *Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk*. The National Academies Press. https://doi.org/10.17226/22297

VTTI. (2024). *InSight data access website: SHRP 2 naturalistic driving study.* 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 preprint arXiv:2011.00054*. 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. (2021). *Global Plan for the Decade of Action for Road Safety 2021-2030*. https://www.who.int/publications/m/item/global-plan-for-the-decade-of-action-for-road-safety-2021-2030

WHO. (2023). *Global status report on road safety*. 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*(sup1), 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, April 3-6). *Harmonized Approaches for Baseline Creation in Prospective Safety Performance Assessment of Driving Automation Systems*. 27th International Technical Conference on the Enhanced Safety of Vehicles

(ESV), Yokohama, Japan. 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://group.mercedes-benz.com/documents/innovation/other/safety-first-for-automated-driving.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* (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 Tekniska Hogskola]. https://research.chalmers.se/publication/542271

Wu, J., Flannagan, C., Sander, U., & Bärgman, J. (2024). Model-based generation of representative rear-end crash scenarios across the full severity range using pre-crash data. *arXiv preprint arXiv:2406.15538*. https://doi.org/10.48550/arXiv.2406.15538

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* (DOT HS 812 801). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812801