THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

# Improving pre-crash safety systems

Using computational driver behavior models and efficient sampling methods

XIAOMI YANG

Department of Mechanics and Maritime Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2025

Improving pre-crash safety systems
Using computational driver behavior models and efficient sampling methods
XIAOMI YANG
ISBN: 978-91-8103-184-3



Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5642
ISSN 0346-718X

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

Cover:
An overview of the four papers included in this work in relation to the overall aim.

Printed by Chalmers digitaltryck

Gothenburg, Sweden 2025

# Abstract

In-vehicle pre-crash safety systems have been used by the automotive industry for some time as integral parts of advanced driver assistance systems (ADAS) and automated driving systems (ADS), in order to improve traffic safety worldwide. Among the methods used to assess these systems, virtual safety assessment methods have been shown to have great potential and efficiency. These methods are likely to continue to play a very important role in assessing vehicle safety at all levels of automation. The ultimate aim of this thesis is to enhance the safety performance of conflict and crash avoidance systems through the use of computational driver behavior models. The work first addresses this aim by incorporating behavior models into pre-crash safety systems; the second part focuses on overcoming methodological challenges encountered in safety system development and assessments.

The first objective of this thesis is to investigate the safety performance of safety systems that include a driver model incorporating drivers' comfortable behaviors in its crash avoidance algorithm. Chinese car-to-two-wheeler crashes were targeted; automated emergency braking (AEB) algorithms which include drivers' comfort zone boundaries (CZB) were compared to a traditional AEB algorithm. The proposed algorithms showed larger safety performance benefits, indicating that including computational behavior models in the algorithms of pre-crash safety systems may reduce the number of crashes and injuries on our roads. It should also be noted that residual crash characteristics did not differ among different AEB implementations. If in-crash protection systems do not have to account for different AEB outcomes, then the systems' designs could be simplified, leading to a more effective allocation of resources.

The second objective is to develop a method for the efficient collection of human-participant data, for use in the development of safety systems that incorporate driver behavior. The resulting method, predictive Bayesian optional stopping (pBOS), enables early stopping—either when a specific statistical target is reached or when it is not likely that the target will be reached, given the available resources (e.g., financing or test-track time). The results show that traditional Bayesian optional stopping (BOS) outperforms traditional frequentist sample size determination—and pBOS outperforms traditional BOS when the experiments have less than a 50% chance of reaching the target with the allocated resources. Consequently, under the appropriate conditions, the use of pBOS in the development of pre-crash safety systems is likely to reduce the resources required, allowing them to be reallocated to other safety research or system development priorities.

The third objective is to develop and apply a method for efficient sampling in crash causation model-based scenario generation for virtual safety assessment. The method, which is machine-learning-assisted, actively and iteratively updates the sampling probability based on new simulation results. The method requires almost 50% fewer simulations than traditional importance sampling. In addition, the impact on efficiency of incorporating the following three features into the method was investigated: domain knowledge-based adaptive sample space reduction logic, stratification, and batch size (the number of samples per iteration). The results show that both knowledge-based logic and stratification can reduce the target estimation error, and a larger batch size is preferred for overall simulation efficiency. As with pBOS, active sampling in behavior model-based pre-crash safety system assessment may reduce development costs, allowing the reallocation of resources.

Keywords: advanced driver assistance systems, automated driving systems, counterfactual simulation, scenario generation, Bayesian optional stopping, active sampling, car-to-VRU, conflict and crash avoidance

*To my family*

*To Eric*

*To the past, current, and future*

# Acknowledgements

I want to express my sincere gratitude to my supervisor Jonas Bärgman and co-supervisor Carol Flannagan for their invaluable guidance, support, and encouragement throughout my studies. Without your support and help, I could not have come so far. I would also like to thank András Bálint for his support and supervision during the first two years, and to my examiner Marco Dozza for all the insightful discussions and meetings. I would like to thank Nils Lübbe for his co-supervision during my master's thesis and Henrik Imberg for his collaboration and support during Papers III and IV. I would like to thank Tina Mayberry for her language reviews. I would like to thank all my colleagues at the division for the stimulating discussions, enjoyable fika, lunch, and after-work time.

I want to acknowledge the SHAPE-IT project, which supported the majority of the research work under the European Union's Horizon 2020 research and innovation programme (under the Marie Skłodowska-Curie grant agreement 860410). I also want to thank Autoliv for funding the initial research, as well as the V4Safety project and the HiDrive project for funding part of the research work.

I would like to thank the friends and colleagues I have met. Thanks to Simon, Malin, and Erik for their support and encouragement during and outside of the HAM project. Thanks Shan and Huizhong for insightful discussions at UMTRI. Thanks to Wudi, Shuyu and Xiaolin for all their support during my downtime and thanks to Jonte, Pontus, Sarri, Anton, Chiara and Thomas for all the fun game time together.

I am indebted to my family for their endless support and understanding of all my decisions, and I feel very fortunate to be a part of such a loving and supportive family. I would like to thank Lotta, Peter, and Emelie for treating me like a family member and making me feel at home. I am grateful to my fiancé Eric for his unwavering love, kindness, understanding, open-mindedness, patience, encouragement, and humor, which have been a source of strength, happiness, and comfort to me throughout our time together. Eric, I love you.

# List of publications

## Paper I

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

Author's contribution: wrote the first draft of the paper; was responsible for the simulation setup, development, and implementation of the models and the evaluated AEBs; performed the simulations; did most of the analysis; and contributed substantially to the synthesis of results and conclusions.

## Paper II

Yang, X., Flannagan, C., & Bärgman, J. (2025). Strategic decision points in experiments: A predictive Bayesian optional stopping method. *Preprint*. https://doi.org/10.48550/arXiv.2503.00818

Author's contribution: wrote the first draft of the paper, was responsible for the simulation setup and development, performed the simulations, did most of the analysis, and led the synthesis and conclusions work.

## Paper III

Imberg, H., Yang, X., Flannagan, C., & Bärgman, J. (2024). Active Sampling: A Machine-Learning-Assisted Framework for Finite Population Inference with Optimal Subsamples. *Technometrics*, *67*(1), 46–57. https://doi.org/10.1080/00401706.2024.2374554

Author's contribution: contributed to the setup of the crash causation and counterfactual AEB simulations used as the application (motivating example), applied the active sampling (from the first author), structured and analyzed the results from the simulations, wrote a substantial part of the paper and reviewed and edited the other parts, and interpreted results and wrote conclusions jointly with the other authors.

## Paper IV

Yang, X., Imberg, H., Flannagan, C., & Bärgman, J. (2025). Evaluation of adaptive sampling methods in scenario generation for virtual safety impact assessment of pre-crash safety systems. *Preprint*. https://doi.org/10.48550/arXiv.2503.00815

Author’s contribution: wrote the first draft of the paper contributed to the simulation development, performed the simulations, did the analysis, and led the synthesis and conclusions work.

## Acronyms

| | | |
|---|---|---|
| ACC | - | Adaptive cruise control |
| ADAS | - | Advanced driver assistance systems |
| ADS | - | Automated driving systems |
| AEB | - | Automated emergency braking |
| AI | - | Artificial intelligence |
| ASSR | - | Adaptive sample space reduction |
| AUC | - | Area under the receiver operating characteristic curve |
| BOS | - | Bayesian optional stopping |
| CIDAS | - | China In-Depth Accident Study |
| CIL | - | Credible interval length |
| CZB | - | Comfort zone boundary |
| FCW | - | Forward collision warning |
| GIDAS | - | German In-Depth Accident Study |
| HBM | - | Human body model |
| HMI | - | Human machine interface |
| IIHS | - | Insurance Institute for Highway Safety |
| NDD | - | Naturalistic driving data |
| pBOS | - | Predictive Bayesian optional stopping |
| PCM | - | Pre-crash matrix |
| PCTSD | - | Pre-crash time-series data |
| PTWs | - | Powered two-wheelers |
| RMSE | - | Root mean squared error |
| ROPE | - | Region of practical equivalence |
| SHRP2 | - | The second Strategic Highway Research Program |
| SHUFO | - | Shanghai United Road Traffic Safety Scientific Research Center |
| TL | - | Tolerance level |
| TTC | - | Time-to-collision |
| TW | - | Two-wheeler |
| VRUs | - | Vulnerable road users |

# Contents

# 1 Introduction

Approximately 1.2 million people die annually in road traffic crashes worldwide, and more than half of the deaths are among vulnerable road users (VRUs), a category including pedestrians and all two-wheeler (TW) riders (World Health Organization, 2023). Powered two-wheelers (PTWs) and powered three-wheelers were involved in nearly 30% of crash fatalities reported in 2016 worldwide (World Health Organization, 2017). Due to the substantial negative impact of crashes on society, many researchers have studied both crashes and crash mechanisms in order to identify the problems that need to be solved (Anund et al., 2016; Bianchi Piccinini et al., 2017; Bucsuházy et al., 2020; Klauer et al., 2006; Petridou & Moustaki, 2000; Viano & Ridella, 1996; X. Wang et al., 2022).

Human factors may contribute to as much as 94% of all traffic crashes (NHTSA, 2015). Therefore, many argue that the number of traffic safety situations involving fatalities and injuries will be reduced substantially with the development of advanced driver assistance systems (ADAS) and automated driving systems (ADS) (Cicchino, 2017; Eichberger et al., 2011; Jermakian, 2011; Kyriakidis et al., 2015; Payre et al., 2014; Rödel et al., 2014). In fact, ADAS has already been shown to substantially improve traffic safety (Cicchino, 2017; Fildes et al., 2015). Data on ADS are only starting to come in, but the preliminary results are positive (albeit only based on low-severity crashes; see Kusano et al., 2024).

Conflict and crash avoidance is an integral part of both ADAS and ADS. Conflict avoidance refers to the vehicles' ability to avoid conflicts in everyday driving even as the traffic environment changes dynamically. Crash avoidance, on the other hand, refers to the vehicle's ability to avoid a crash (or mitigate its consequences) in a critical situation, when a crash is imminent unless an avoidance maneuver is initiated (Jermakian, 2011).

As part of their development process, conflict and crash avoidance systems need to be assessed to ensure that they actually improve safety (Jeong & Oh, 2017; Lemmen et al., 2012; Lindman & Tivesten, 2006; M. Zhao et al., 2017). In addition, regulatory constraints (ISO, 2021, 2022, 2024), and consumer testing programs (C-NCAP, 2018, 2020; Euro NCAP, 2022, 2023) also require methods to assess safety. Safety assessment can be either retrospective or prospective. Retrospective safety assessments evaluate existing conflict and crash avoidance systems using historic data (collected from end-user's cars in the traffic system); for example, fatalities, injuries, crash rates, or insurance claims can be compared for vehicles with and without the systems (Cicchino, 2017). However, this method cannot be used to assess the safety of a system (or system version) that is not yet on the market—enough data are not yet available. In contrast, prospective assessments are conducted before new systems (or new versions of existing systems) are introduced to the market, either during their development or before they are widely adopted. Virtual simulation is one such method, which in the last decade has gained attention for its efficiency and cost-effectiveness. This method for assessing crash avoidance systems is the core of this work.

## 1.1 Pre-crash safety systems

To understand and study pre-crash safety systems (subsuming both conflict and crash avoidance systems), ADAS and ADS should be defined in more detail, given that there is still some confusion about their differences. ADAS and ADS represent different levels of automation, as defined by the SAE (2021). ADAS are advanced technologies which assist drivers during driving tasks by providing information, warnings, and/or interventions related to events unfolding or by taking over part of the driving. The most popular ADAS include warning systems—like lane departure warning (LDW; Son et al., 2015) and forward collision warning (FCW; Dagan et al., 2004)—and automated systems—like automated emergency braking (AEB; Haus et al., 2019), adaptive cruise control (ACC, Benmimoun et al., 2013), and lane-keeping and centering (Tsoi et al., 2010). FCW, AEB, and ACC are longitudinal control systems, while LDW and lane-keeping and centering are lateral control systems.

In contrast, ADS aim to relieve drivers of the driving task by taking over entirely—at least for part of the drive (i.e., conditional automation; SAE, 2021). Examples of ADS systems already on the market today are found in the autonomous vehicles produced by Waymo (Scanlon et al., 2021; Waymo, 2022); autonomous cars with no drivers which transport paying customers; and the first cars approved for conditional automation in Europe (with a driver behind the steering wheel), produced by Daimler.

Furthermore, there are several vehicles on the market that are on the upper end of the ADAS level of automation, but driver supervision of the vehicle's automated driving task is still needed. That is, drivers' reactions and response times (to both the road situations and the safety systems) still influence road safety. Examples of such systems are vehicles from Tesla (Ingle & Phute, 2016; Tesla, 2022), which is branding its vehicles as self-driving or even fully self-driving, but still requires that the driver be attentive and in control. Actually, a recent report from the Insurance Institute for Highway Safety (IIHS) in the US (Matt, 2022) stated that 53% of General Motors' Super Cruise users and 42% of Tesla Autopilot users were comfortable not paying attention to the roadway while the vehicles drove—although the users are still ultimately responsible. These statistics highlight the challenges of designing in-vehicle technologies able to account for both driver behavior and traffic safety.

One example of a safety system that is typically designed without accounting for driver behavior is a traditional AEB system. It is often based on time-to-collision (TTC) (Kusano & Gabler, 2012) or required deceleration (Brännström et al., 2010; Coelingh et al., 2007). A TTC-based AEB usually triggers when the vehicle reaches a fixed TTC threshold, while a required-deceleration-based AEB takes the maximum braking level of a vehicle into consideration and aims to trigger at (or just before) that point of no return—after which the vehicle cannot possibly avoid the crash by braking. There have been many studies on the real-world effect of AEB (Fildes et al., 2015; Haus et al., 2019); although it improves safety substantially, it has the potential to improve safety even more. Many crashes would have been avoided if AEB had been triggered earlier. The problem is, of course, that triggering earlier may substantially increase nuisance interventions, which can be both irritating and dangerous. Alternatives to the traditional AEB algorithms take driver comfort into account, enabling AEB to trigger earlier without being a nuisance. This latter approach is the focus of Paper I of this work.

## 1.2 Virtual simulations for safety benefit assessment

Virtual simulation, one type of prospective safety benefit assessment of pre-crash safety systems, has gained attention for its efficiency and cost-effectiveness (ISO, 2024). Scenario-based virtual safety assessments have been widely tested and developed in recent years (Cai et al., 2022; Nalic et al., 2020; Riedmaier et al., 2020). In the context of virtual simulations, a “scenario” is a “temporal sequence of scene elements with actions and events of the participating elements occurring within this sequence” (Ulbrich et al., 2015). Scenarios can be divided into three categories based on the level of detail: functional, logical, and concrete (Menzel et al., 2018). Functional scenarios are the most abstract; they are typically based on traffic rules, expert knowledge, etc. Logical scenarios are more detailed and described by sets of parameter ranges or distributions. Finally, concrete scenarios are time-series scenarios with detailed dynamics of the involved road users. Because virtual safety assessments require scenarios with defined parameter values, concrete scenarios are used.

The concrete scenarios in virtual simulations represent relevant traffic infrastructure and road-users, with and without a virtual representation of the pre-crash safety system under assessment. The scenarios (typically pre-crash kinematics data or variants of those) without the pre-crash safety system are baseline scenarios, while the scenarios with the pre-crash safety system present and active are treatment scenarios.

Virtual simulations are called counterfactual simulations when they create treatment scenarios by applying the virtual representation of the pre-crash system to the baseline scenarios. The performance of the safety system under assessment can consequently be evaluated by comparing each baseline scenario with the corresponding treatment scenario using safety assessment metrics. The metrics commonly used to evaluate the benefits of a system are either crash-based or non-crash-based. Examples of the former include two closely related metrics, impact speed and injury mitigation (Doecke et al., 2020; Jermakian, 2011). Impact speed is used to calculate delta-v, the change in velocity during a crash; delta-v in turn is used to calculate injury mitigation (Dean et al., 2023; Viano & Parenteau, 2010). Crash rate (the number of crashes per unit of exposure, such as miles traveled) is another such metric. In contrast, TTC is a non-crash-based metric, also called a safety surrogate (Guo et al., 2010; C. Wang et al., 2021; C. Wang & Stamatiadis, 2014).

Counterfactual simulations have been used extensively to quantify the performance of pre-crash safety systems (Cicchino, 2022; Haus et al., 2019; Kibalama et al., 2022; Rosén, 2013; Sander, 2018; Scanlon et al., 2022; Xia et al., 2017). These simulations enact what would have happened if the systems had been present before the crashes happened. The detailed kinematic data from the pre-crash phase that counterfactual simulations require include measurements of real-world crashes from event data recorders (Gabler et al., 2004) or from in-depth studies of individual crashes (Sander & Lubbe, 2018). Examples of sources for these data include the German In-Depth Accident Study (GIDAS) pre-crash matrix (PCM; Rosén, 2013; L. Stark et al., 2019), the China In-Depth Accident Study (CIDAS) PCM (T. Wei et al., 2022), and the Shanghai United Road Traffic Safety Scientific Research Center (SHUFO) pre-crash time-series data (PCTSD; Deng et al., 2013; Ding et al., 2016). Pre-crash kinematics commonly capture dynamic information such as traffic participants' velocities, accelerations, positions, heading angles, etc., for several seconds before the crash. In the counterfactual simulation, the system can intervene and change the dynamics, which may lead to avoiding the crash or mitigating its severity.

The baseline scenarios without the safety system are usually (at least to date) obtained from measured or reconstructed manual driving situations—often from in-depth crash databases. However, a challenge when using original or reconstructed crash data is that the amount of crash data is limited. Conflict situations collected in traffic are rare, and data from crash reconstructions are even rarer. It is thus highly beneficial to generate (create) conflict situations (especially crashes) that are representative of the real-world scenario under study across the different levels of outcome severity. If done appropriately, scenario generation can produce a more comprehensive set of scenarios than those available from datasets of real-world data (Fahrenkrog et al., 2024). The scenario generation-based counterfactual simulation process is shown in Figure 1. The first step is to generate baseline scenarios, and the second is to apply the pre-crash safety systems in order to generate counterfactual treatment scenarios.

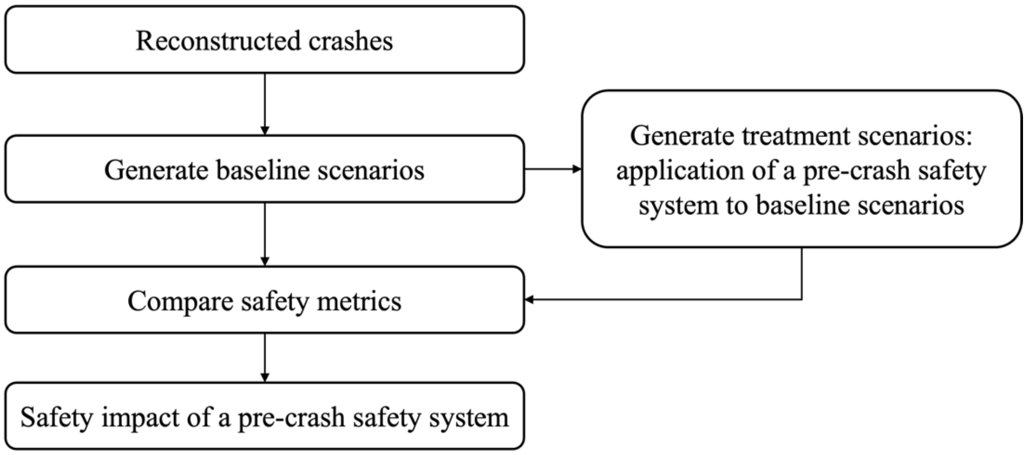


*Figure 1: A high-level illustration of the process of scenario generation using counterfactual simulation for pre-crash safety system assessment.*

Baseline scenario generation can be divided into three approaches: A, B, and C (Fahrenkrog et al., 2024; ISO, 2024). Approach A uses recorded data (e.g., by event data recorders; Donnelly et al., 2001), non-recorded data (e.g., reconstructed time-series from in-depth crash databases; Otte et al., 2012) as the baseline. The number of generated scenarios does not increase. Approach B uses measured or reconstructed crashes as a starting point and then creates variants of each of these crashes, increasing the number of generated scenarios. Unlike approach B, approach C does not use the pre-crash kinematics from individual original events directly, but instead typically creates logical scenarios that are parameterized by kinematics or driver models to create concrete scenarios.

In practice, approach A commonly uses reconstruction software to reconstruct crashes from in-depth crash investigations (Sharma et al., 2007; X. Zhang et al., 2008). With approach B, there are two ways to generate variants. The first is to vary the kinematic scenario information (describing the physical states of the road-users involved, such as accelerations, speeds, and positions). The second way is to vary the parameters of behavior models that represent some human behavior relevant to the simulation. For approach C, concrete scenarios can also be generated in two ways. First, they can be generated from parameters of road participants' dynamics (e.g., initial speed) and the road network (e.g., lanes and other constraints)—a purely "kinematic" approach. Second, as in approach B, they can be generated using driver models relevant to the simulation. However, unlike in approach B, there is no specific real-world scenario used as the starting point; instead, the driver model (and models of other road users) and road environment are logical scenarios, and concrete scenarios are generated through parameter variations and simulations.

What is important for this work is the way that the scenarios generated by approaches B and C can be based on models of driver behavior and on the parameter distributions that feed into these models. Such models should, at least, capture mechanisms of crash causation and drivers' responses to the critical situations that unfold. (Examples of behavior models for scenario generation are driver braking level and on/off-road glances.) That is, the scenario generation process incorporates crash causation mechanisms into the simulations to better mimic crashes in the real world. Thus, driver models, which represent driver behaviors, should be considered when developing crash causation models that computationally describe the mechanisms behind crashes for scenario generation. The driver models may contain parameters to represent drivers' responses (e.g., braking or steering) and states of attention (e.g., whether the driver is looking off-road or is drowsy) in critical situations. Driver response models are mathematical descriptions of how drivers respond to the critical event that emerges from the crash causation mechanisms.

Until now the term *safety assessment* has been used broadly, but it should be noted that there are two specific types of pre-crash safety assessment: verification and validation (Amersbach & Winner, 2019; Åsljung et al., 2017; Beglerovic et al., 2017) on the one hand, and safety impact assessment (Cicchino, 2022; Rosén, 2013; Sander et al., 2019) on the other.

Scenarios in approaches B and C can be generated by either parameter ranges or parameter distributions from real-world data, for either kinematics or driver models. Importantly, if ranges are used, the generated scenarios only represent the scenarios covered, without information about their exposure, which is the probability of the individual concrete scenario occurring in the real world (Cai et al., 2022; Nalic et al., 2020; Riedmaier et al., 2020). In contrast, scenarios generated by sampling from distributions of real-world data can—at least in theory—be representative of real-world crash scenarios, enabling safety impact estimations that quantify the safety impact (e.g., estimations of overall crash avoidance rate, impact speed reduction, and injury risk reduction).

These two approaches address two very different questions. Although both approaches can be part of safety verification and validation, accurate exposure data must be a part of scenario generation if the scenarios are to be used to quantify the safety impact. On the other hand, scenario generation that does not provide estimates of exposure is typically used in system validation and verification (Khastgir et al., 2017), since it is only necessary to know whether the system performs as expected. Note that safety impact assessment can be part of safety assessment and argumentation—for example, assessing if a system has a positive risk balance (Kauffmann et al., 2022). This work solely addresses scenario generation that includes exposure, in order to generate scenarios representative of the real world. The quantification of safety impact pre-crash safety systems in this work is called *safety impact assessment*.

Also note that when scenario generation considers exposure, the data from which the distributions are created must be relevant for the specific assessment (e.g., in terms of scenarios, geographical area, and human demographics). Joint distributions are often needed, as parameters are often correlated (J. Kim & Mahmassani, 2011). In fact, concrete scenarios can be generated from either concrete or logical scenarios, when distributions of specific parameters are combined—either parameters directly describing kinematics of the participating road users or parameters of the driver models that are used in the simulation to generate scenarios. However, to realistically quantify the system's impact on safety, the distributions must be accurate and relevant (Bärgman et al., 2017).

To conclude, the generation of baseline scenarios (typically crashes) is a core component of all virtual pre-crash safety assessments. Scenario generation has been developed and used in many projects, including L3Pilot (Bjorvatn et al., 2021), V4SAFETY (European Commission, 2022), and PEGASUS (Winner et al., 2019), and has been reported in numerous scientific publications (Nalic et al., 2020; Riedmaier et al., 2020). Scenario generation for safety impact assessment should generate crashes that are representative of the population. Representative scenario generation is also at the core of papers III and IV in this thesis.

## 1.3 Supporting the improvement of safety systems

This thesis addresses two ways to support the improvement of safety systems. The first is to directly enhance the performance of the safety system itself, and the second is to improve the methodologies that support the development of these systems. Even though the improved methodologies do not improve safety system performance directly, they can increase the efficiency and precision of safety system development, thereby enhancing system performance within the limited timeframe of its development.

### 1.3.1 Improving AEB designs

As described above, AEB is one of the safety systems that have shown a substantial impact on safety by both avoiding crashes and mitigating injuries in crashes (Chajmowicz et al., 2019). As mentioned, most available AEB system algorithms do not take drivers' or road users' behaviors into account and are instead designed based on vehicle limitations (e.g., the time it takes to decelerate to a stop). However, understanding driver behaviors is often essential when developing safety systems, as drivers' responses can influence some safety systems' performance and drivers' acceptance influences their choice of use. For example, a driver's reaction time to an FCW significantly influences whether a crash is avoided, since a quicker reaction time allows for more effective evasive maneuvers, thereby reducing the risk of collision (Lubbe, 2017). In addition, drivers' perceived safety, feeling of comfort, and level of trust all impact the acceptance and use of new or developing systems (He, 2024; J. D. Lee & See, 2004; Molnar et al., 2018; T. Zhang et al., 2020).

Driver models, widely used in the development of pre-crash safety systems, are commonly included in the systems' assessment. One example is Waymo's NIEON ("Non-Impaired, with Eyes ON the conflict"), a reference driver model that represents "consistently performing, always-attentive drivers" (Scanlon et al., 2022). Driver models can be incorporated directly into the system design, such as the FCW timing that is based on the average human driver's reaction time (T. L. Brown et al., 2001). However, the explicit inclusion of driver models in system design is still relatively rare; as noted, AEB systems do not typically include driver behaviors in the algorithm design in published work—although there are some exceptions (Brännström et al., 2010; Sander, 2018).

One type of driver model with strong potential to improve safety system performance and acceptance is based on comfort zone boundaries (CZB). Drivers' CZBs for car-to-car scenarios describe, for example, drivers' comfortable braking and steering limits, which are typically quantified based on naturalistic studies (Dingus et al., 2006) or controlled experiments (Bärgman, Smith, et al., 2015). There are also a few studies that include drivers' CZBs in car-to-car AEB algorithms. For example, Sander (2018) re-simulated accidents with car-to-car AEB algorithms with different CZB parameter settings. Drivers' CZBs in car-to-pedestrian crossing scenarios were also investigated by Lubbe & Davidsson (2015). They analyzed brake onset time and brake deceleration levels in pedestrian crossing scenarios; their results show that 90% of drivers braked before 2.2s TTC for pedestrian speeds of 2m/s.

As PTWs are becoming more prevalent on public roads, especially in developing countries, there is a substantial need to develop relevant safety systems, such as AEBs for car-to-PTW crashes. In fact, PTWs are a critical safety concern in Southeast Asia, where they account for 43% of all traffic deaths (World Health Organization, 2019). Chinese statistics indicate that there were 9,923 fatalities and 48,518 traffic injuries that involved motorcycles in China in 2022 and those crashes were mostly motorcycles with cars (National Bureau of Statistics of China, 2023).

However, compared to drivers' CZBs, much less is known about those of PTW riders. There appears to be only one published study that explicitly quantifies riders' CZBs. Kumar Akinapalli et al. (2023) reported a maximum deceleration of around 5.5 m/s$^2$ in a study of 58 participants who rode PTWs for 32km in India. Paper I investigates the specific crash avoidance benefits of CZB-based PTW AEBs through virtual simulations.

### 1.3.2 Improving methodologies

Incorporating driver behavior models into system designs and virtual safety assessments presents several challenges: two key challenges in incorporating driver behavior models into system designs and virtual safety assessments are presented here. First, the models must be accurate and based on data from real drivers. Conducting experiments, a common way to obtain the data, is resource-intensive and costly. Therefore, methods for optimizing experimental resource allocation are being sought. While existing methods allow an experiment to be stopped early when an experimental goal is reached, no method exists to support redirecting experimental resources when an experimental goal is unlikely to be reached. The second challenge is that integrating driver behavior models into virtual safety assessment adds complexity; the increase in the number of simulation parameters and simulations can lead to combinatorial explosion and simulation load issues. Therefore, efficient sampling methods are needed. Current approaches require advance knowledge of the shape of the response surface—and lack the ability to dynamically update that knowledge as simulations are conducted. In summary, there are two methodological gaps related to the challenges of developing efficient methods for developing driver models (for use in safety systems or virtual simulations) and scenario generation-based safety assessment. These two gaps and the related literature are further described in the two subsections below.

#### *1.3.2.1 Sample size determination*

An example of the significant resources and funding required to collect data on driver behavior in road traffic can be found in the second Strategic Highway Research Program (SHRP2) naturalistic driving study. Conducted by the Transportation Research Board in the United States, the study collected data from over 3500 drivers over a period of two years, comprising more than 32 million miles of continuous data (Antin et al., 2019). The authorized budget supporting SHRP2 implementation exceeded US $170 million (FHWA SHRP2 Implementation Team, 2024). Driving simulators are also commonly used for driver behavior data collection. Although the exact cost of advanced simulators is not publicly detailed, building them and running the experiments (including participant recruitment) are known to be costly. Further, during system development the system is iteratively updated, necessitating new experiments in case drivers' behavior and/or acceptance changes. A method to determine the smallest needed sample size to reach a particular statistical goal dynamically during an experiment, enabling the reallocation of experimental resources, would help reduce development time and cost. Consequently, product development could be faster, reducing time to market and potentially resulting in less expensive systems for the end users; a more efficient method also allows the developers to reallocate resources to further optimize system performance.

Current methods determine the minimum sample size when the goal of an experiment (for example, to test a hypothesis or reach a value with a certain precision for a parameter of interest) can be reached without exceeding the available resources (for example, the number of participants or the test track time available). Without enough samples, the research question cannot be answered—but conducting a study with more samples than needed is a waste of resources. Therefore, efficient sample size determination methods are needed.

In the traditional frequentist framework, sample size determination methods set the sample size *before* the experiment starts, and once the experiment begins, the frequentist paradigm requires that they all be collected. This framework relies on comparison of a statistic from the observed data with an estimated distribution of that statistic under the designed experimental protocol. Stopping data collection when enough data have been gathered can save experimental resources but is not possible with frequentist statistics because it violates the assumption of the test.

The Bayesian framework, in contrast, takes a different approach that can accommodate stopping an experiment early. The Bayesian approach estimates the distribution of a statistic based on prior knowledge updated with new data; the estimate does not depend on a particular experimental paradigm. Bayesian optional stopping (BOS) methods can stop an experiment early when a goal is reached (Rouder, 2014). These methods can guide researchers to the smallest necessary sample size during an experiment. However, they are not designed to assess whether the goal *can* be reached as the experiment is executed. As a result, BOS methods might run all available experimental trials without reaching the goal. In the context of resource constraints on data collection, there is a need to improve sample size determination methods to allow for early termination of experiments when it becomes apparent that collecting enough samples with the available resources is not feasible. Such a method is proposed in Paper II.

#### *1.3.2.2 Sampling methods*

The use of driver models to generate scenarios for pre-crash safety system assessments addresses the issue of the small sample sizes available in crash data (Bärgman et al., 2024), since the models can provide comprehensive coverage while still being representative of real-world crashes. While approaches B and C can employ sampling to vary the dynamics or driver model parameters, as scenarios and models become more complex—and the number of parameters increases correspondingly—it is intractable to run all combinations of all parameters, because of limited time and computational resources. To mitigate the issue of combinatorial explosion, sampling methods can be used to choose a subset of all combinations of parameters in order to estimate, for example, the population mean, distribution, or quantiles.

There is a range of different sampling methods, which vary in efficiency. The most basic method, simple random sampling, is good for homogeneous and uniformly selected populations in which each sample has an equal opportunity to be selected. This type of sampling is widely used in survey studies (Golzar & Noor, 2022). However, it is not very efficient, as it often requires many samples for good accuracy and repeatability.

Instead of sampling based solely on the data, importance sampling uses information from a distribution defined by a prior to oversample parameters from parts of the parameter space that contribute more to the estimate (i.e., the variable you want to estimate). When the goal is safety impact assessment, the target parameter is typically a finite population inference target. As such, it can be a total, ratio, or correlation coefficient estimate. Importance sampling then reweights the data to achieve an unbiased estimation for the target. This method has been shown to be efficient. It has been applied in many different domains, including scenario generation for safety assessment (see, for example: de Gelder & Paardekooper, 2017; O'Kelly et al., 2018; X. Wang et al., 2021).

Sampling is particularly important for improving the efficiency of large-scale data problems (such as virtual safety simulations), since they typically require that a massive number of scenarios be generated and tested. However, because traditional importance sampling relies on prior knowledge, it can be inefficient when that knowledge is unknown or incorrect. In fact, choosing the importance distribution for importance sampling is challenging, as driving is a complex task, requiring the driver models to be complex. No importance sampling method described in the literature enables adaptation of the sampling model during data collection, although this possibility would substantially improve the method's efficiency and robustness. Papers III and IV in this thesis propose and apply, respectively, a novel way to perform adaptive importance sampling using machine learning methods.

## 1.4 Aims and objectives

The overarching aim of this research is to improve the safety performance of pre-crash safety systems by using computational driver behavior models. We apply the models in two ways: 1) incorporating them into the design of pre-crash safety systems and 2) incorporating them into virtual safety impact assessments of the systems.

Three objectives were defined to fulfil this aim. The first objective directly addresses safety by incorporating a computational driver model into safety system algorithm design; the second and third objectives overcome the methodological challenges described above, enabling faster system deployment and the reallocation of resources to other pre-crash safety system development initiatives. The three objectives are:

1) to investigate how the inclusion of a computational driver behavior model in algorithms of pre-crash systems impacts the systems' performance.
2) to improve the efficiency of human-participant data collection, thereby also improving the efficiency of the development of pre-crash safety systems that include computational behavior models.
3) to improve scenario generation efficiency when using computational driver behavior models to generate crashes for virtual safety impact assessment.

In this work, four papers are included. An illustration of their contributions to the objectives is shown in Figure 2.

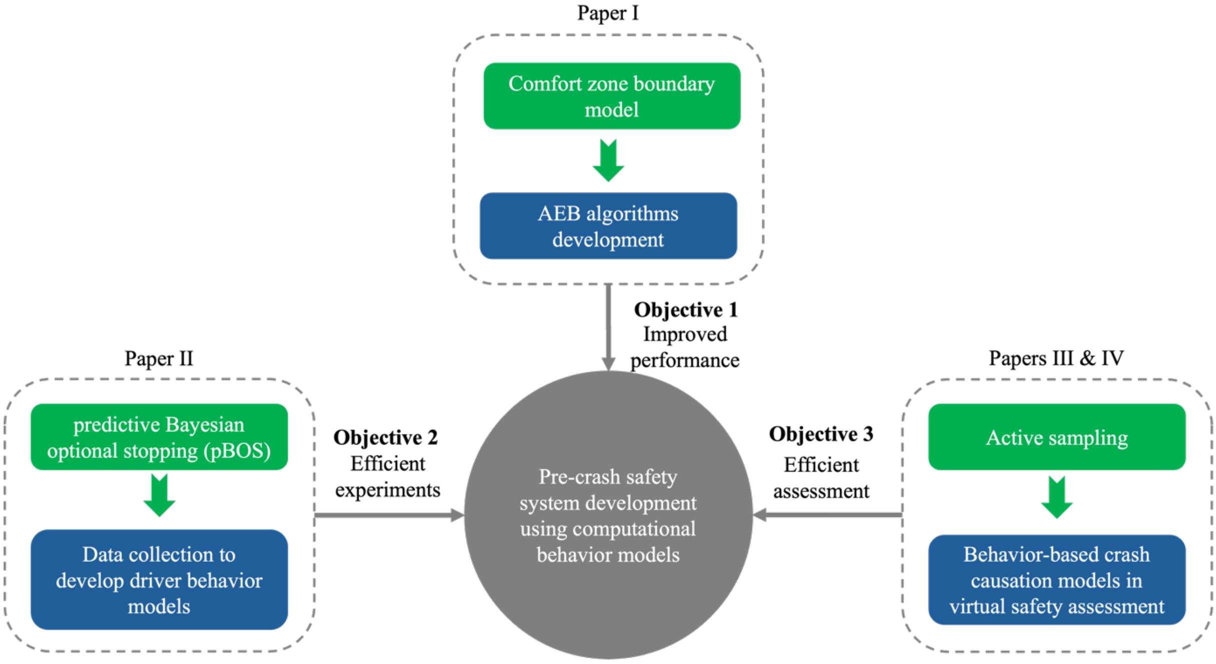


*Figure 2: An overview of the four papers included in this work in relation to the three objectives and the overall aim (gray circle).*

# 2 Methods

There are many methodological aspects to developing and accessing pre-crash safety systems, ranging from the data and road-user behavior models used for system assessment to efficient methods for both data collection and system assessment. The following sections cover the four key methodological aspects of this work: Section 2.1 introduces the baseline data used for virtual safety assessment in general, as well as the data used specifically in this work. Section 2.2 describes some uses of road-user models in virtual safety assessment in general, as well as some aspects of behavior model development and their application in this work. Section 2.3 provides background information about frequentist and Bayesian statistics and their respective approaches to sample size determination—information needed to understand this work. Section 2.4 provides more information about the sampling methods used in scenario generation and the application of the machine-learning-assisted active sampling method developed in this work.

## 2.1 Data for virtual safety assessment

Developing and accessing pre-crash safety systems requires robust and high-quality data. The baseline data used in the literature for counterfactual simulations are described in Subsection 2.1.1. The baseline crash data used in this thesis are specified in Subsection 2.1.2. Subsection 2.1.3 describes the metrics used to evaluate data quality, the importance of data quality checks, and the data quality check applied in this thesis work.

### 2.1.1 Baseline data for counterfactual simulations

Crash data can be used directly as baseline scenarios, and in most cases the reconstructed pre-crash kinematics data from crashes are used in counterfactual simulations. The crash datasets can be region-based, like US data (Kusano & Gabler, 2014), German data (L. Stark et al., 2019), Chinese data (T. Wei et al., 2022), or Indian data (Pisharam et al., 2022). Pre-crash kinematics can also be used to generate scenarios for the assessment of pre-crash safety systems (Esenturk et al., 2021; Song et al., 2022; C. Stark et al., 2020).

In addition, data from sources other than crashes can be combined to generate critical baseline scenarios. For example, Waymo (Scanlon et al., 2022) combined real-world crash and near-crash data, ADS testing data, and expert knowledge in order to identify critical scenarios for the safety verification and validation of ADS. In another work, Wu (2024) merged crash data from two different datasets with near-crashes from naturalistic driving data (NDD).

### 2.1.2 Data used in this work

This study made use of the following data types: reconstructed crash data, NDD, test-track and simulator data. Paper I used 93 SHUFO PCTSD car-to-PTW crashes. The SHUFO crash database consists mainly of real crashes collected in the Shanghai Jiading distinct in China; the selection criteria are at least one passenger car involved, at least one airbag deployed, at least one severe injury, or an economic loss of at least US$3500 (Deng et al., 2013). The SHUFO PCTSD consist of reconstructed pre-crash kinematics data based on the SHUFO crash database, with a structure similar to that of the GIDAS PCM. The data contain detailed kinematics information for a few seconds before the crashes happened, which can be used for counterfactual assessment. Paper II used test-track and simulator data collected by Puente Guillen & Gohl (2019) and Papers III and IV used 44 Volvo rear-end crashes which were selected from an internal Volvo crash dataset. The main selection criterion was that the cost of the crash exceeded €4500 (Isaksson-Hellman & Norin, 2005).

In Paper I, SHUFO PCTSD were used for the performance assessment of different AEB systems targeting car-to-TW crashes in China, and the CZB thresholds were based on NDD and test-track data. The glance and maximum deceleration data used in Papers III and IV were extracted from NDD.

### 2.1.3 Data quality

The quality of the data used in this work can be quantified by two metrics: accuracy and completeness. Data accuracy refers to the verisimilitude of the data. Possible sources of inaccuracy include reconstructed crashes that differ from the original events and misreporting. Completeness has two aspects: the data must include sufficient information/variables to represent the real world accurately, and the samples must not lack critical information (Sander et al., 2024).

Accuracy is to a large extent related to bias. Understanding bias is important for interpreting and using data correctly, since most data include some form of bias. Consequently, compensating for and mitigating bias are challenges a researcher should consider before using data for modeling or assessment. Data bias commonly occurs when data are gathered and analyzed. For example, GIDAS data are collected through police reports for traffic accidents involving personal injury. However, police may not be notified about crashes when there are no obvious injuries or participants agree to private settlements. As a result, less severe crashes are under-represented in GIDAS data (Hautzinger et al., 2004). This type of bias is not unique to GIDAS but is present in all crash databases (Sander et al., 2024). Geographical location can also introduce bias, as different countries or regions may have varying levels of underreporting (Ahmed et al., 2019). Additionally, the type of road user can affect data bias. For instance, cyclist crashes have the highest levels of underreporting of all road crashes (Elvik & Mysen, 1999). Since data form the basis for practically all road safety research, their quality significantly influences the accuracy of safety assessments and, consequently, the design of safety systems. Therefore, we evaluated the quality of the data used in this work to understand whether they limit the validity of our assessment; further, we tried to avoid bias during the analysis before the assessment.

As the SHUFO PCTSD were reconstructed based on SHUFO data, quality checks can be performed by comparing participant number matching and the cumulative distribution of pre-crash simulation times for dynamic data (Junaid et al., 2024). Physical constraints should also be considered in quality checks, to ensure realistic values for yaw angles, yaw angle rates related to steering, accelerations, and acceleration jerks.

For data completeness, the pre-crash time series should ideally start at least five seconds before the crash (Junaid et al., 2024). It can start even earlier, but the data are unlikely to be relevant for crash avoidance. Cases with durations less than two seconds were excluded from the assessment in Paper I to improve completeness. One limitation of Paper I is the small sample size of 93 cases, which motivated the development of methods for scenario generation to increase completeness with a larger sample size. For Papers III and IV, only a small subset of the Volvo database was used (44 crashes). Actually, the motivation for developing the sampling method proposed in Paper III and applied in Paper IV was also to increase completeness, both in terms of scenario (interpolation) and severity coverage; scenarios of all severities were created with the method, even though the original data are biased towards higher-severity crashes (in part due to the inclusion criterion on repair costs; Isaksson-Hellman & Norin, 2005). That is, incorporating a driver model-based crash causation model can potentially compensate for the severity bias, as such a model commonly contains a wider range of distributions for driver behavior. In our application, the crash causation model includes drivers' off-road glances from 0s to 6.6s. Typically, the longer the glance duration the more severe the crash. The generated crashes are likely to cover a wider severity range than the original 44 cases, while accuracy is maintained. However, validating coverage and representativeness is beyond the scope of this work. Relevant work on completeness and validation can be found in the work by Bärgman et al. (2024).

## 2.2 Road-user behavior models

Road-user behavior models are used in both the development and assessment of pre-crash safety systems. This section first describes different road-user behavior models in system design in general, followed by our specific application in Subsection 2.2.1. Subsection 2.2.2 introduces the usage of road-user behavior models in virtual safety assessment in general and in our scenario generation application. Subsection 2.2.3 describes different experiment types for data collection for modeling road-user behavior, and emphasizes the importance of measuring the variability of driver responses—especially when developing safety systems.

### 2.2.1 Driver behavior models as a part of vehicle system algorithms

We do not know exactly what algorithms are in the safety systems of production vehicles and we do not know if the algorithms include road-user behavior models such as CZBs, as car manufacturers do not usually disclose that information. (Safety systems with CZBs include driver behavior, rather than being based purely on the performance boundaries of the involved vehicle.) Traditional required-deceleration-based AEB systems only include vehicle's deceleration constraints, so if the trigger is activated before the CZB is reached, drivers may see these as false-positive activations (i.e., when the activation is not warranted, from the driver's perspective). Nonetheless, some studies about driver models in pre-crash safety system designs have been reported in the literature. For example, a driver model was used in the pre-crash safety system design presented by Dozza et al. (2020). They first designed the FCW system to issue a warning if there was a mismatch between the actual braking and steering and those predicted by the driver model. AEB was then activated if there was no reaction by the driver to the warning. For the same FCW system, assessments of the reaction time of different driver response models were further performed by Kovaceva et al. (2022). They found that different driver response models affected the safety impact of the behavior-based FCW. Further, drivers' CZBs have been used in driver models for car-to-car AEB (Brännström et al., 2010; Sander, 2018), pedestrian AEB (Edwards et al., 2015) and cyclist AEB (Duan et al., 2017). The potential safety benefits of CZB-based AEB could also be included in pre-crash safety system designs for higher-level automation, like ADS. Wei et al. (2019) explored these benefits by including human-like behavior in vehicle motion control in order to ensure smooth, comfortable trajectories with automated driving algorithms. This finding is highly relevant, as it is an example of human-like driving being preferred by occupants of ADSs.

This work includes CZBs in pre-crash safety system (AEB) algorithms. A CZB-based AEB triggers when a crash cannot be avoided without exceeding the road user's CZBs, which represent the maximum comfortable steering and braking maneuvers. The values for these maneuvers are based on the literature (Bärgman, Smith, et al., 2015; Brännström et al., 2014; C-NCAP, 2018; Costa et al., 2019; Kiefer et al., 2003). We compared five different CZB-based AEBs made up of different combinations of steering and braking maneuvers. The AEB that includes only the CZB model for driver braking, for example, will trigger when the crash is unavoidable if the driver only brakes comfortably (as shown in Figure 3). In contrast, the AEB that includes the CZB models for both driver and rider braking and steering will trigger at the moment when the crash is unavoidable even if both driver and rider brake and/or steer comfortably. In this implementation, the braking and steering maneuvers (and their respective CZBs in the AEB algorithm) are simulated separately, but they can be considered in combination in future studies.

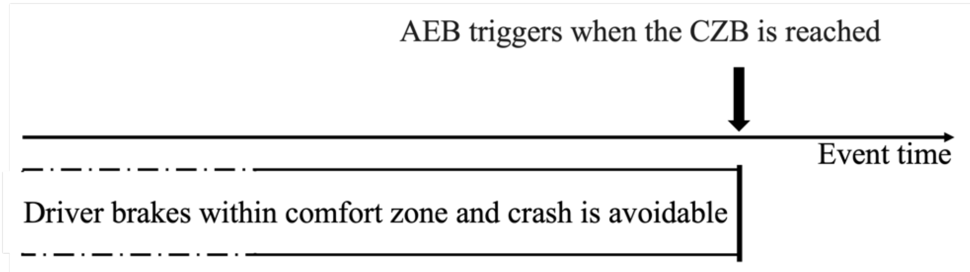


*Figure 3: Illustration of CZB-based AEB including only the driver's braking CZB.*

FCW is another example of a pre-crash safety system that incorporates driver behavior in the system design. How fast drivers react to FCW influences its safety performance, which means that reaction time needs to be considered when designing the FCW algorithm. Although there is existing literature on drivers' responses to FCW (Bakowski et al., 2015; Ruscio et al., 2015; Wege et al., 2013), the responses vary substantially between systems. When a new algorithm for an FCW (or other warning system, such as a take-over request in ADS; Morales-Alvarez et al., 2020) is under development, driver responses to the warning should be investigated, since the literature indicates that factors such as modalities and trigger criticality can affect driver response (Lubbe, 2017; M. Wang et al., 2020). Further, brake reaction time, brake deceleration, and brake jerk are dependent on the human machine interface (HMI; Benderius et al., 2018; J. Zhang et al., 2021). Further experimentation is essential to better understand the factors affecting drivers' responses to FCWs. An FCW design which achieves more consistent driver responses would perform more reliably.

### 2.2.2 Driver behavior models as part of crash causation models for scenario generation

Driver models are commonly used in traffic and counterfactual simulations (Hughes et al., 2015). Many studies have been conducted to investigate the risk factors for driving and elucidate reasons for crashes (Anund et al., 2016; Dingus et al., 2016; Klauer et al., 2006; G. Zhang et al., 2016). Since driver error contributes to almost all crashes, drivers' risk factors merit special attention. One such risk factor, driver inattention (when the driver is not focused on driving), is a factor in over half of police-reported crashes (Stutts et al., 2001). Inattention plays an especially important role in rear-end collisions: Knipling et al. (1993) suggested that driver inattention was implicated in 90% of rear-end collisions occurring on straight roadways. Driver glance behavior can be used in crash causation models as a crucial indicator of attention in rear-end scenarios on highways. The work by BMW is an example; see the L3Pilot report (Bjorvatn et al., 2021) and the work by Fries & Fahrenkrog (2021). Other examples of safety assessments that include behavior-based crash causation models are assessments of off-road glance behaviors related to the HMI design (J. Y. Lee et al., 2018; Victor et al., 2015), and off-road glances in relation to levels of automation (Bärgman & Victor, 2019). These studies illustrate the importance of driver models in scenario generation. Scenario generation can also use more simplistic driver models or different levels of braking response, such as delays in the reaction time to collision warnings (Kusano & Gabler, 2012). Bärgman et al. (2024) compared a glance-based crash causation model, a traditional response model (based on Kusano and Gabler's 2012 paper), and an improved traditional response model. Their results show that the choice of model significantly impacts the generated scenarios; further, the glance-based crash causation model performed the best, as its generated scenarios were most similar to those of the original data.

This work includes driver off-road glance and braking behavior as a core crash causation mechanism. That is, as off-road glances are the major cause of highway rear-end crashes and drivers can react with different braking levels to a frontal braking vehicle, the same glance-and-deceleration-based crash causation model used by Bärgman et al. (2024) was used to generate rear-end crashes in this study. We found no documented evidence of a correlation between glance and deceleration, so they were assumed to be independent. Further, in this study the braking behavior by the driver of the following vehicle was removed and replaced by a simple looming model (Bärgman, Lisovskaja, et al., 2015) and the glance-and-deceleration model described above. In the simulations, the driver was assumed to be attentive after looking back at the road and to react by braking according to the looming model. The braking level was based on the fitted maximum deceleration distribution. Each original crash generated 1005 combinations: 67 (glance) x 15 (deceleration). With 44 original crashes, 44,220 baseline cases were generated—a challenge to simulate.

### 2.2.3 Data collection for developing driver behavior models

Driver behavior models such as those described above require data collected from drivers. Some studies are observational and their data may be used for exploratory or descriptive analysis. Others involve experiments with more specific targets (as opposed to just seeking to understand behavior better), such as testing hypotheses or quantifying the response surface to inputs that have complex variations. Experiments can also be designed to address questions about the variability of responses by a driver over time or between different drivers. Additionally, as the design of a system can vary and be updated iteratively, the response patterns can also vary over time. It is therefore necessary to explore driver behavior in response to a safety system during its development. By gaining insight into human driver behavior and responses, developers can design systems that better anticipate and respond to different scenarios on the road, ultimately enhancing overall traffic safety. The more consistent (less variable) the drivers' responses, the more reliable the system's performance.

While more data generally improve analysis, cost efficiency calls for identifying the smallest sample size which can still achieve the experimental goal. The sample size needed is usually determined by the precision of a key statistic, such as mean response or the difference in mean responses to two designs or conditions. Achieving response consistency (across people or time) is a related goal, typically measured using sample variance and related metrics (Davidian & Carroll, 1987). It is important to note that the term "precision" is often used to refer to the general concept of variability; however, technically, "precision" is defined as the inverse of variance. Both frequentist confidence intervals and Bayesian credible intervals are closely related to sample variance (and thus precision). The term "precision target" (following Kruschke, 2014) refers to any variability-related metric as a target for sample size determination.

## 2.3 Sample size determination in traffic safety experiments

Efficient sample size determination methods are essential for achieving experimental goals within resource constraints. Subsection 2.3.1 introduces frequentist and Bayesian statistics in the context of sample size determination; Bayesian sample size determination methods, including the proposed predictive Bayesian optional stopping method, are introduced in Subsection 2.3.2.

### 2.3.1 Frequentist vs Bayesian approaches to sample size determination

Frequentism and Bayesianism represent two distinct statistical paradigms. Both are trying to develop an understanding of some aspect of the world (i.e., predictable patterns in some domain) and both have a sample of data to work with. Frequentists assume that the available data represent a random sample of the data that could have been collected in a given experiment. That is, there is one underlying, typically parametric model of the world, but the specific data could change if the data collection were repeated. In contrast, Bayesians treat the collected data as fixed and consider the underlying data model as unknown. Bayesian probability represents the degree of belief about each specific model (i.e., the specific parameter values) being the true underlying data model. Bayesian models begin with a prior, which represents the understanding of the true data model before data collection. The models are then updated with data and the degree of belief about it is also updated.

The fact that these two philosophies take very different approaches to hypothesis testing has consequences for sample size determination. Frequentists compute a statistic using the current sample and then compare it to a distribution of that statistic expected over many repetitions of the same experiment under certain assumptions about the data generation model. The set of assumptions is called the "null hypothesis," which is typically that some parameter equals zero or that there is no difference between two (or more) groups. If the current sample statistic is unlikely under the null hypothesis, then the null hypothesis is rejected. For a particular question and analysis, the definition of "unlikely" is chosen as $\alpha,$ which is also the Type I error rate. A Type I error represents the probability of incorrectly rejecting a true null hypothesis. In contrast, a Type II error $\beta$ represents the probability of failing to reject a false null hypothesis. For a given sample size, there is a tradeoff between Type I and Type II errors. That is, if the chosen $\alpha$ is small, the chance of incorrectly rejecting a true null hypothesis will be small, but $\beta$, the chance of failing to reject a false null hypothesis, will be large. If $\alpha$ is large, the reverse is true.

Power, defined as 1-$\beta$, refers to the probability that a statistical test will correctly reject a false null hypothesis. Power is commonly used to determine a target sample size, in combination with a selected $\alpha$ and assumptions (based on estimates) about the true underlying parameter values (e.g., the difference between two groups) as well as the assumptions about the true variability in that difference. Importantly, because the frequentist hypothesis-testing framework depends on comparison of a sample statistic to the long-term distribution of that statistic under repetitions of the same experiment, the target sample size must be determined *before* the experiment. Moreover, once the experiment starts, it must be fully carried out, since stopping early will violate the assumptions behind the repeated-experiments comparison distribution.

In contrast, Bayesian analysis produces a posterior distribution, which is the probability distribution of a set of target parameters. Hypothesis testing, which is not central to Bayesian methods but is often carried out, is done by comparing the distribution of a given parameter to a benchmark value. Often, Bayesian hypothesis testing is analogous to the frequentist null hypothesis: for example, a parameter or a difference between groups may be benchmarked against a value of zero. However, unlike the frequentist approach, Bayesian hypothesis testing does not depend on an assumption about repeated identical experiments. Instead, the posterior can be updated repeatedly as new data are added, and it is possible to "peek" at the posterior at any time. The critical point here is that, unlike the frequentist approach, if the posterior has met a certain goal (e.g., reaching a specified width of a credible interval) the experiment can be stopped early (Rouder, 2014).

### 2.3.2 Bayesian sample size determination methods

Bayesian optional stopping (BOS) is an approach to sample size determination in the Bayesian context. This approach can save experimental resources, which can then be allocated to other purposes. BOS involves stopping when a specific experimental target, which is linked to particular research questions, is reached. A target can be either effect-size-based or precision-based. Given that precision-based targets support the improvement of safety systems by testing and identifying systems with more consistent driver responses, and that the appropriateness of effect-size based stopping criteria remains under debate (de Heide & Grünwald, 2021; Rouder, 2014), we employ precision-based targets in the application of pBOS.

Even though BOS can halt an experiment early when a target is reached, it can encounter the same issue as frequentist sample size determination: all resources may be exhausted without the target ever being met, which represents another efficiency problem. A method that can estimate how likely it is that the experiment will reach the goal can resolve this issue. Kruschke proposed conducting rehearsal simulations at a single timepoint early in an experiment, to provide a basis for estimating the sample size needed. Extending this approach to repeat the process after every new sample, we discovered that the rehearsal simulations can overestimate or underestimate the variance of future estimates. Thus, we added a calibration step to compensate for the misestimation. Combining this approach with traditional BOS resulted in predictive Bayesian optional stopping (pBOS; Paper II). pBOS aims to save and/or reallocate experimental resources in two kinds of early stopping. One, like BOS, stops the experiment when the target is reached, based on collected data; the other stops the experiment when the target is not likely to be reached, based on rehearsal simulations. The whole process is iteratively updated with newly collected data, and the misestimation of variance for future data based on current data is compensated for with a regression model.

## 2.4 Sampling methods in scenario generation

Turning to scenario generation for virtual simulations, Subsection 2.4.1 describes traditional sampling methods and provides a short overview of the proposed active sampling method. Additional practical considerations that can make sampling more efficient are described in Subsection 2.4.2.

### 2.4.1 Sampling in scenario generation for safety assessment

Scenario spaces for safety impact assessments of pre-crash safety systems are vast. Quantifying the impact of a system relative to a baseline requires taking into account not only the parameter ranges of logical scenarios, but also the scenario distributions—in order to account for exposure. Because generating a massive number of scenarios can be computationally challenging, efficient sampling methods play a crucial role in the selection of the scenario parameters to be generated or simulated.

The most basic method, N-wise sampling, samples all possibilities by discretizing parameters to achieve sufficient parameter space coverage, as demonstrated in the work by Ponn et al. (2019). Another method, simple random sampling, ensures unbiased representation, with all possible generated scenarios having equal probability. Only a subset of the space is sampled, which makes scenario generation more efficient. Roesener et al. (2017) used the method with NDD sampled from a kinematic parameter distribution to generate car-following scenarios. However, this process can still be somewhat inefficient, especially when the data lack crashes or contain too few crashes (D. Zhao et al., 2017). To improve the efficiency, researchers have employed importance sampling as an accelerated approach (de Gelder & Paardekooper, 2017; X. Wang et al., 2021; D. Zhao et al., 2017). For example, X. Wang et al. (2021) applied importance sampling with a reachability analysis to pedestrian-crossing scenarios, generating realistic scenarios that were also physically feasible. This unbiased method uses prior knowledge to estimate what combinations of parameters will generate scenarios that influence the outcome the most. Consequently, the method typically saves resources by not sampling from the parameter space that results in non-critical scenarios.

A non-exhaustive list of literature that uses sampling in scenario generation is shown in Table 1. This list is not intended to contain all relevant literature; rather, it illustrates the variety and combinations of sampling methods used in scenario generation (ISO, 2024).

*Table 1: A non-exhaustive list and classification scheme for research articles that use sampling in scenario generation.*

| **Sampling method** | **Study and application**[1] | **Baseline scenario generation**[2] | **Safety outcome metrics**[3] | **Data/Knowledge type used** |
|---|---|---|---|---|
| N-wise sampling | Ponn et al. (2019): Lane-keeping assist algorithms | none[4] | Non-crash metrics | German motor construction guideline |
| Simple random sampling | Roesener et al. (2017): Car-following scenarios | Approach C | Non-crash metrics | NDD |
| Importance sampling | X. Wang et al. (2021): Pedestrian-crossing scenarios | Approach C | Crash rate and delta-v | NDD |
| | de Gelder & Paardekooper (2017): ACC algorithms | Approach C | Crash rate | NDD |
| | D. Zhao et al. (2017): Cut-in scenarios | Approach C | Crash rate and delta-v | NDD |
| Bisection with logic[5] | Bärgman et al. (2024): Rear-end scenarios | Approach B | Delta-v | Reconstructed crashes |

Although much used, traditional importance sampling is challenging to implement in practice, as prior knowledge about what impacts crash severity does not always exist. When we know little and guess wrong, importance sampling may perform poorly (Elvira et al., 2019).

[1] When the application is 'algorithm', the publication evaluates the algorithm; when the application is 'scenario', the publication generates scenarios without assessing a specific system or algorithm.

[2] Approaches A, B, and C, based on the ISO scenario categorization standard (ISO, 2024).

[3] Non-crash metrics, crash rate, or delta-v depending on the type of generated scenarios and their corresponding safety impact metrics.

[4] N-wise sampling does not quite fit into the ISO 2024 definition of approach A, B, or C; the ISO states that all approaches must be data-driven, but N-wise sampling, as used in this publication, is not.

[5] Simulating all parameter combinations except those with known outcomes, by identifying regions without crashes and with maximum impact speeds through iterative bisection (a way to include logic in the sampling that is simpler than what was done in Paper II).

Unlike importance sampling, the active sampling method proposed in this work does not require prior information to the same extent; instead, it uses machine learning to update the sampling scheme based on the new simulation results. The sampling process is shown in Figure 4. The initialization starts with initial simulations that represent the most severe crashes for each case, (choosing the longest off-road glances and the lowest decelerations). An initial sampling scheme is then built on the initial simulation results (whether there was a crash or not, what the impact speed was, and what the injury risk was) based on training and optimization of machine learning models. Two machine-learning models predict the results of all remaining simulations based on the results of the simulations already performed. One model, for classification, is used to model whether there is a crash or not; the other, for regression, models other assessment targets, including impact speed reduction and injury risk reduction. We also investigated several different machine-learning models, including random forest regression, gradient boosting, and k-nearest neighbors. After the initial sampling scheme is built, new draws of simulations can be selected, and training and optimization of machine learning models are updated. All selected simulations contribute to the estimation of the target, and the target precision is evaluated at each iteration. If the target is reached, sampling can stop. Otherwise, the sampling continues, and this update process iterates. This iterative sampling strategy, inspired by machine learning, was applied in the scenario generation in this work.

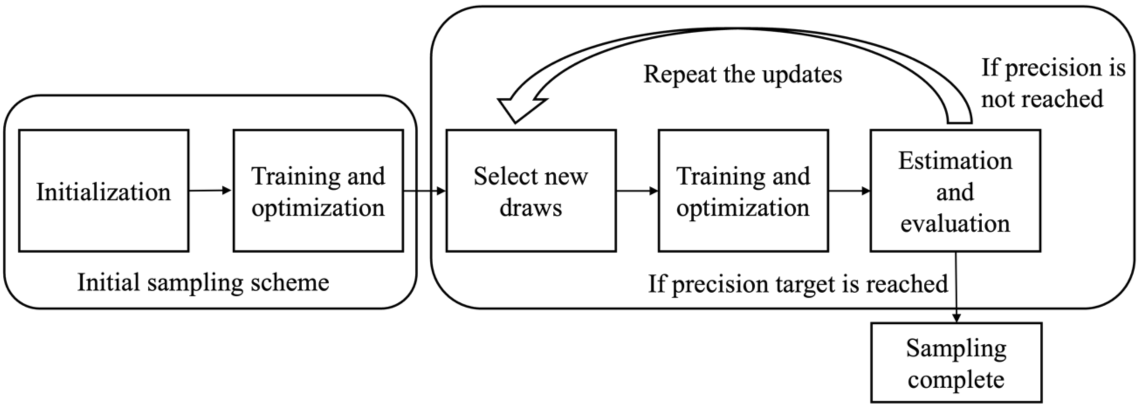


*Figure 4: Sampling process of the proposed active sampling method*

Targets for sampling are chosen explicitly within the context of safety assessment. Different indicators can be used as targets depending on the research focus. For example, metrics like TTC, brake threat number, and steering threat number can be used to evaluate the situation's criticality (Brännström et al., 2008). Because this work focuses on safety impact assessment and crash situations (rather than near-crash and comfort scenarios), the evaluation metrics in this work are impact speed reduction (or, rather, relative speed at impact), crash avoidance rate, and injury risk reduction.

### 2.4.2 Additional considerations in the practical application of active sampling

In some practical applications, features of the outcome space can be used to improve the sampling efficiency. For example, in pre-crash safety system assessments, if a behavior model-based simulation results in no crashes, then—with domain knowledge of the specific application—it might be known that any simulations with parameter values (e.g., level of deceleration and duration of off-road glances) generating less severe conditions cannot result in crashes, either. Similarly, if a particular simulation results in the maximum outcome severity and the driver did not react at all, it can also be known that more severe conditions will result in the same maximum-severity outcome. Since simulations for which the outcome is known need not be run, these two pieces of information can improve efficiency. Further, rule-based domain knowledge combined with machine-learning prediction methods might make active sampling even more efficient. Paper IV investigated this combination to determine whether it could improve the practical efficiency of simulations in safety assessments.

Another practical consideration for virtual simulations is that sampling efficiency (CPU time) calls for small batches (the number of samples updated and run in parallel per iteration) and computing efficiency (wall-clock time) calls for large batches. Paper IV also investigates the consequences of different batch sizes (i.e., how many simulations are run in parallel) on sampling efficiency.

# 3 Summary of papers

## 3.1 Paper I

### Evaluation of comfort zone boundary based automated emergency braking algorithms for car-to-powered-two-wheeler crashes in China

Xiaomi Yang, Nils Lübbe, and Jonas Bärgman (2024)

#### Introduction

With an increasing number of powered-two-wheelers (PTWs) in China, crashes involving PTWs constitute a large proportion of Chinese traffic crashes. In-vehicle automated emergency braking (AEB) systems have been shown to be effective in preventing or mitigating car-to-car, car-to-pedestrian, and car-to-cyclist crashes, but few have studied these systems for conflicts between cars and PTWs. As previous studies have also shown the benefits of including road users' comfort zone boundaries (CZBs) in AEB algorithms for other types of AEB, the benefits of their inclusion in TW (two-wheeler) AEB should also be investigated.

#### Methods

A CZB model with different thresholds was used to assess the performance impact of including CZBs in AEB algorithms for car-to-PTW conflicts. The CZB-based AEBs trigger when road users cannot avoid a crash with comfortable braking or/and steering maneuver(s). Five different CZB-based AEBs were compared with a traditional required-deceleration-based AEB using counterfactual simulation, applied to pre-crash kinematics data. The different AEB systems were also compared with each other, to assess their safety performance regarding crash avoidance rate and injury mitigation. Residual crash characteristics were studied with respect to impact speed and location.

#### Results

The CZB-based AEB that considered driver braking and steering avoided 66.7% of the original crashes, performing substantially better than the traditional AEB, which avoided 48%. When only interventions that occurred earlier than those of the traditional AEB were considered, all CZB-based algorithms (understandably) performed better than the traditional AEB. Furthermore, the residual crashes for the different AEBs had similar impact speed distributions and impact location distributions.

**Conclusions**

Road user participants' CZBs included in AEB system design improve system performance, in terms of both crash avoidance/mitigation and fewer potential nuisance interventions. The similarities in residual crashes after different AEB implementations may simplify the future design of in-crash protection systems—at least for the Chinese market, since the data were obtained in China.

## 3.2 Paper II

### Strategic decision points in experiments: A predictive Bayesian optional stopping method

Xiaomi Yang, Carol Flannagan, and Jonas Bärgman (2025)

#### Introduction

Efficient sample size determination is crucial for experiments. On one hand, without enough samples, a research question cannot be answered; while on the other hand, more than enough samples waste experiment resources. The traditional frequentist sample size determination method estimates the smallest needed sample size based on assumptions about sample variance and effect size. The estimation is made before the experiment starts, and once it has started all the samples must be collected. In contrast, traditional Bayesian optional stopping (BOS) can stop the data collection when enough have been collected by peeking at the data. However, this method risks collecting all the data without answering the research question. A more efficient sequential sample size determination method is needed to stop the experiment early when it is unlikely that enough samples can be collected with the available resources.

#### Methods

We propose a predictive Bayesian optional stopping (pBOS) method that combines traditional BOS with Kruschke's rehearsal simulations. Rehearsal simulations evaluate the chance to reach a predefined target based on predictions about the data to be collected. The predefined target, which is related to the research question, can be hypothesis testing-based or precision-based. We evaluated a precision-based target. The predicted data model is based on the data posterior distribution. However, the prediction of possible future data is typically misestimated for the precision-based target, so we use a regression model to compensate for the misestimation. Like BOS, pBOS can stop an experiment early when the statistical target is reached, but it can also do so when it is unlikely that the target is attainable with the given constraints.

#### Results

The regression model for misestimation compensation shows good fit with the data. The performance of the proposed pBOS was evaluated using the area under the receiver operating characteristic curve (AUC) and cost benefit metrics. pBOS shows a range of 0.79 to 0.98 for the AUC with most priors and a cost benefit up to 118% better than that of traditional BOS, which is itself more efficient than frequentist sample size determination.

**Conclusions**

This study demonstrates the significant advantages that pBOS offers over traditional BOS and frequentist sample size determination. Utilizing iterative analysis and prediction, the method allows researchers to stop experiments when the research question cannot be answered with allocated resources or when sufficient data have been collected, which is not possible with traditional frequentist methods or BOS. Thus, pBOS has the potential to enhance the efficiency and resource management of experimental designs, particularly in traffic and transport research involving studies with human participants.

## 3.3 Paper III

### Active sampling: A machine-learning-assisted framework for finite population inference with optimal subsamples

Henrik Imberg, Xiaomi Yang, Carol Flannagan, and Jonas Bärgman (2024)

#### Introduction

Sampling has been a research topic for a long time. However, as virtual simulations prove to be a crucial tool for the development and assessment of safety systems, the need to complete such simulations in a timely, cost-effective way has emphasized the importance of efficient sub-sampling methods. Traditional methods usually require prior knowledge of the underlying data, which is often scant or unavailable. There is therefore a need to develop new sampling methods that do not require prior knowledge but perform at least as well as traditional methods.

#### Methods

A machine-learning-assisted framework for an optimal sampling method is proposed and applied in this work. This active sampling method iteratively updates the sampling scheme based on the previously chosen subsamples until the newly chosen samples have reached a desired estimated mean precision. The active sampling is implemented with and without model prediction uncertainty and the results are compared with those of the importance sampling methods. The three methods were applied to an assessment of the benefits of an automated emergency braking system (with respect to the parameters mean impact speed reduction and crash avoidance rate).

#### Results

The proposed active sampling method performed much better with model uncertainty than without. Further, it also performed better than traditional importance sampling methods—especially when it was optimized on the characterized parameters. Three different variance estimation methods for stopping the simulation (sampling) based on precision were also illustrated and their results were assessed. Of these the classical survey method was the most efficient, performing as well as bootstrapping for different batch sizes.

### Conclusions

The proposed machine-learning-assisted framework was applied to a motivating example for a pre-crash safety system assessment, showing the benefits of the proposed active sampling method over traditional sampling methods. The method has the potential to be used in scenario generation and virtual simulations across conflict and crash avoidance safety assessments as well as across scenarios, facilitating faster assessment and consequently better systems. In addition, three variance estimation methods for active sampling were assessed.

## 3.4 Paper IV

### Evaluation of adaptive sampling methods in scenario generation for virtual safety impact assessment of pre-crash safety systems

Xiaomi Yang, Henrik Imberg, Carol Flannagan, and Jonas Bärgman (2025)

#### Introduction

Virtual safety assessment is crucial for evaluating the safety benefits of systems like advanced driver assistance systems and automated driving systems. However, the number of crash scenarios increases combinatorially with the number of varied parameters, making complete enumeration practically unfeasible. Efficient sampling methods, such as importance sampling and a recently proposed machine-learning-assisted active sampling, have been demonstrated but lack detailed evaluations of their implementation. This study investigates the practicalities of implementing these methods and provides recommendations for selecting and implementing the appropriate sampling method.

#### Methods

This study uses a different target than previous work, ensuring that each of the original crashes contributes equally to the sampling. To improve sampling efficiency, it introduces domain-specific logic into the sampling process—a feature that, for some scenarios, substantially reduces the number of required simulations. Stratification is also examined, with both sampling methods implemented with and without stratification. Additionally, we evaluated the effects of different batch sizes.

#### Results

Our results indicate that incorporating domain-specific logic for adaptive sample space reduction can significantly improve the efficiency of both importance sampling and active sampling methods. Active sampling outperforms importance sampling when domain-specific logic is excluded. Domain-specific logic improves importance sampling significantly, and the severity importance sampling performs no worse than active sampling. Stratification improves the efficiency of both sampling methods, regardless of whether domain-specific logic is included. A smaller batch size enables better sampling efficiency than a larger batch size.

#### Conclusions

The findings suggest that stratification is recommended, if applicable, for its high efficiency in both sampling methods. When domain-specific logic can be used to reduce the sampling space, importance sampling and active sampling can be equally effective. However, when no domain information is available, active sampling is more efficient than importance sampling. Additionally, larger batch sizes are preferable when resources are not extremely limited and the required sample size is large.

# 4 Discussion

This work aims to improve the performance of pre-crash safety systems that are part of both ADAS and higher levels of automation. Section 4.1 discusses the first objective of this thesis, investigating the effect of CZB-based driver models in a pre-crash safety system. Section 4.2 explores virtual safety assessments and their inclusion of driver models. Section 4.3 considers different sample size determination methods for data collection. The method proposed in Paper II addresses the second objective of this thesis: developing efficient data collection methods for system development. Section 4.4 examines statistical methods for scenario generation, focusing on the sampling method developed in Papers III and IV. This method addresses the third objective: to improve scenario generation efficiency when using driver behavior models to generate crashes for virtual safety assessment. Section 4.5 looks at the future of machine-learning-assisted methods for scenario generation-based virtual safety assessment, highlighting the potential of AI applications to improve traffic safety research.

## 4.1 The use of driver models in pre-crash safety systems

Recall that the proposed CZB-based PTW AEB algorithm in Paper I is for an in-vehicle safety system. The CZBs are those of the driver—so in fact, they refer to what the car driver assumes the PTW rider is comfortable doing, rather than what the PTW rider actually finds comfortable. Most of the CZB-based PTW AEB algorithms proposed in this work were better at avoiding and mitigating crashes than the traditional (required-deceleration-based) AEB algorithm, while potentially reducing the number of false positives. On the one hand, the PTW AEBs could trigger earlier than the traditional AEB without being considered a nuisance when the road user's CZBs are reached earlier than the vehicles' physical constraints. On the other hand, the traditional AEB's trigger could be considered a potential nuisance if the vehicle's point of no return for braking occurs earlier than the user's CZB for braking or steering. In that situation, road users still have the potential to avoid the crash themselves.

As a result, CZB-based algorithms are likely to perform better, with fewer false positives, when a steering adjustment would suffice to avoid the crash even when hard braking would not. Examples include straight front-to-front (also called head-on) and overtaking scenarios, as well as scenarios with a small overlap between two road users in a rear-end situation, such as when a PTW is driving on the edge of the road and a car is driving in the middle of the lane; the driver only needs to make a small, comfortable steering adjustment (which can be performed quite late) to avoid a crash.

The larger the difference in trigger times between the traditional and CZB-based AEBs, the greater the nuisance of the traditional one for the driver. However, CZB-based AEBs may delay the trigger to an extent that the performance is reduced: the more constraints, the lower the performance (see Table 4 in Paper I). Currently, nuisance interventions are typically not a problem for AEBs; they have a relatively high acceptance rate (Mohd Ishanuddin et al., 2021, 2022; Reagan et al., 2018). Thus, it is likely that car manufacturers include rules in the AEB algorithms that avoid interventions in, for example, the overtaking situations described above. It would be reasonable to keep those rules or change them to CZB-based solutions, which may make it possible to improve the overall performance by allowing earlier triggers at even lower decelerations (thus lowering the risk of rear-end conflicts between the following vehicle and the vehicle behind it even further). Nuisance considerations need to be investigated further in future research.

Although Paper I explored the use of CZB driver models in the design of a crash avoidance safety system, driver models have substantial potential to improve conflict avoidance systems as well, both in ADAS and higher-level automation vehicle systems. Research has found that in the domain of ADAS, crash avoidance systems have a higher usage rate than conflict avoidance systems (Reagan et al., 2018); including driver behavior models in ADAS conflict avoidance systems like lane-keeping systems may increase their acceptance and usage. In fact, many ACC systems already consider drivers' CZBs—by allowing the drivers to explicitly set a (time or distance) gap to the lead vehicle which they are comfortable with. Moon & Yi (2008) designed an ACC system based on driver behaviors and such ACC system showed similar following performance compared to human's in both high-speed driving and low-speed traffic jams (Moon & Yi, 2008).

Researchers have also investigated the use of driving algorithms based on natural driving styles in higher-level systems such as ADS and studied driver acceptance of the algorithms. Bellem et al. (2016, 2018) investigated how highly automated vehicles should drive to ensure driving comfort. They conducted a simulator study to identify comfortable driving strategies based on three different driving styles. Based on their analysis, they provide a recommendation for a comfortable automated driving style. Wei et al. (2019) used human driving behavior in their ADS trajectory algorithm and tested the algorithm in four typical scenarios. The results show that the vehicle with more human-like driving behavior is expected to have a higher acceptance rate with both drivers and other road users. Thus, it may be argued that including driver behavior (e.g., explicit information about drivers' CZBs) in higher-level vehicle automation systems could improve drivers' acceptance, perhaps as a complement to the system's machine-learning-based driving. However, this conjecture needs to be investigated further.

Another type of driver behavior model that can be incorporated into CZB modeling to improve drivers' acceptance is the drivers' perceived risk of the driving situation. It captures the level of risk perceived by drivers (subjective risk), as opposed to actual risk (objective risk), which relates to the criticality of the situation (Griffin et al., 2020; Kolekar et al., 2020b). As perceived risk decreases, drivers are significantly more likely to accept safety systems (He et al., 2022); thus, computational behavior models are likely to be more acceptable if they operationalize perceived risk in terms of CZB. That is, if CZB thresholds can account for perceived risk (rather than merely actual risk), those models and thresholds could be used to increase driver acceptance of pre-crash safety systems.

In addition to contributing to CZB modeling, perceived risk models can also be used to guide the development of HMI and system design in order to improve their acceptance by drivers. S. Kim et al. (2024) investigated how HMI influences drivers' perceived risk in driving simulator studies. The results show that drivers' perceived risk was lowest when HMI designs used both visual and auditory modalities to provide information (about other road users detected by the vehicle and about maneuver decisions made by the vehicle). Kolekar et al. (2020a) found that coupling a driver's risk field model (representing the driver's belief about the probability of an event occurring) to a controller that maintains the perceived risk below a certain threshold results in human-like driving behavior. While the current research indicates that incorporating computational models of perceived risk into system design promises to improve system acceptance (at least in part by making the models' behavior more human-like), further research is needed.

As pre-crash safety systems continue to develop, they will impact vehicles' safety performance not only in terms of crash avoidance, but also in terms of the crash population. Estimating the characteristics of residual crashes may help developers optimize in-crash protection systems when conflict and avoidance systems are widely available on our roads—that is, not only for today's crash characteristics, but also for those of the future. Paper I shows that the impact speed and impact location distributions of the residual crashes were similar across the investigated AEBs (the proposed CZB-based PTW AEBs and the traditional required-deceleration-based PTW AEB). Compared to the original crashes, the residual crashes had lower impact speeds and a higher proportion of corner impacts. These results are supported by others' work; simulation results also show that corner impacts increased after AEB implementation

(Jeppsson & Lubbe, 2020). The increase in corner impacts reported in Paper I indicates that A-pillar protection (i.e., the pillars on the left and right of the front windscreen) and side airbags may be needed in a future where TW AEBs are ubiquitous. Additional information about future crash characteristics can impact the way that in-crash protection systems are designed.

It is important to note that the CZBs' values used in this work are fairly high; they were chosen so that the CZBs of the vast majority of drivers would be exceeded before the models' values are reached in a real situation, in order to minimize the risk of nuisance interventions. Sander (2018) tested the pre-crash safety system's performance by applying comfortable braking and steering thresholds of 3 $m/s^2$, 5 $m/s^2$ and 7 $m/s^2$ to crashes. He found that, compared with a CZB of 5 $m/s^2$, system performance increased around 10% for 3 $m/s^2$ and decreased around 10% for 7 $m/s^2$. Further sensitivity analyses should be performed using the CZB threshold values for both critical and non-critical situations in order to assess system safety and false-positive performance across a set of drivers; car-to-VRU scenarios should be included.

At the time of writing, we found no relevant studies about PTW riders' CZBs—either their own or from the driver's point of view; we therefore used the drivers' CZB model for the PTW riders. We argue that this is reasonable, as most drivers are likely to project their own experience as drivers onto the actions of the PTW rider (although drivers who are also PTW riders might not). Some recent studies have collected data on PTW riders. For example, Kumar Akinapalli et al. (2023) collected short-term naturalistic data from 58 participants riding a PTW for a round trip of 32km in India. Their data show a maximum deceleration of around 5.5 $m/s^2$, and the 75$^{th}$ percentile of maximum deceleration is around 2.75 $m/s^2$. According to this study, our choice of 5$m/s^2$ covers most PTW riders' longitudinal CZBs. This information can be used in the future to set CZBs for PTW riders. It is not clear to what extent the application of the driver's CZB model to other road users actually applies to the driver's perception of a nuisance intervention. This aspect of CZB models should be investigated in future studies.

Knowledge gaps notwithstanding, CZB-based AEBs must include CZBs that are as accurate as current research permits; the boundary constraints need to come from reliable data. In this work the constraints are drawn from different driver behavior studies. One study by Bärgman, Smith, et al. (2015) provides data from guided experiments on a test track with 22 participants. Comfort zones were defined for comfortable driving and hurried driving. The authors report a maximum comfortable lateral acceleration of around 5 $m/s^2$ (during left turns). In another field test study, participants who were instructed to drive normally reached a maximum braking deceleration of 5 $m/s^2$ (Hugemann & Nickel, 2003). Further, Sander (2018) selected CZBs for the AEB evaluated in his work according to experimental data and NDD (Bärgman, Smith, et al., 2015; Dingus et al., 2006; Moon & Yi, 2008), suggesting that 5 $m/s^2$ is a reasonable starting value for maximum deceleration in AEBs. The inclusion of accurate CZB models could improve pre-crash safety systems.

## 4.2 The use of driver models in traffic safety assessment

In addition to being included in pre-crash safety system designs, driver models can be included in crash causation models. The crash causation models can then be used as part of the virtual assessments of pre-crash safety systems. In rear-end crashes, one of the most common crash types, visual attention and braking behavior are crucial factors. Therefore, Papers III and IV adopted a crash causation model based on glance and deceleration in order to create rear-end crashes for manually driven cars (without automation). The same model was used by Bärgman et al. (2024), who compared crashes generated with this model and crashes generated with a traditional reaction model by applying scenario generation approaches to a set of reconstructed rear-end crashes. Their results suggest that crash causation model-based scenario generation is a promising method for virtually assessing traffic safety in general, and crash avoidance systems in particular.

In contrast, the traditional reaction model is based on the work of Kusano & Gabler (2012). In fact, many driving simulator studies that investigate a visual-manual task or delays related to cognitive load use the distribution of driver's reaction times to the onset of a lead vehicle braking as the main safety metric (Markkula et al., 2016). Bärgman et al. (2024) show that, compared to the impact speed distribution of the reaction model-based generated crashes, the distribution of the crashes generated through virtual simulation with glance-and-deceleration-based crash causation more closely resembles the original impact speed distribution. Similarly, a driver reaction model was used to generate scenarios for crashes with obstacles in front of the subject vehicle; however, the outcomes of the generated baseline were not compared with the outcomes of the original crashes to validate their method (Funke et al., 2011). BMW has also developed a promising model which is currently available as open source, but has not yet been validated in depth on crash outcome (Eclipse Foundation, 2024). The stochastic cognitive driver model was applied in a passive cut-in scenario, and the generated data are similar to the real-world data (Fries et al., 2022).

Although not directly relevant for this thesis, another use of driver models is worth mentioning in this section: as reference driver models for the assessment of higher-level automation. These models are currently used to represent a specific type of driver, like a competent and careful one. They do not necessarily need to represent all human behaviors during different driving tasks, but instead could be used to assess whether a crash is preventable by some specific type of driver. In addition to the reference model defined in Regulation 157 (UNECE, 2021), Mattas et al. (2022) suggested a "fuzzy safety model", which they argue can mimic defensive drivers (who tend to avoid emergencies in advance) by capturing their comfortable braking behaviors for conflict avoidance. Another example of a reference driver model is the Waymo NIEON model described earlier (Scanlon et al., 2022). It is possible that accurate CZB models may be used as components in these reference models, perhaps enabling better safety assessment of pre-crash safety systems (Olleja, 2024). More work is needed to operationalize CZB collection across scenarios—for use in both pre-crash safety system design and in the assessment of such systems—in order to facilitate improved pre-crash safety systems.

## 4.3 Sample size determination for data collection

A challenge with using driver models is obtaining the human behavior data to support them. Since experimental setups are costly and data collection involving human participants can be time-consuming, it is important to be as efficient as possible. This section looks at current methods for estimating the minimum sample size required to achieve a specific experimental target (e.g., credible interval length: CIL) and discusses the new method proposed in this work. The methods are further explained and compared.

This thesis introduces an efficient sequential sample size determination method, pBOS, and compares its performance to traditional BOS and a frequentist fixed sample size method. The effectiveness of pBOS, and how various pBOS feature settings influence its performance, are also investigated. The results can provide potential users of pBOS with guidelines for its use.

The traditional BOS method is typically more cost-efficient than frequentist sample size determination methods, as shown in Figure 8 in Paper II. Additionally, Sadia & Hossain (2014) concluded from their simulations that the sample size when using BOS with appropriate prior information is often smaller than the sample size when using the frequentist method. The BOS method is more cost-efficient because it can terminate early when a target is reached, thereby reducing the number of experimental trials. Given that BOS is generally more efficient than frequentist methods, in this work pBOS performance was primarily benchmarked against BOS. Figure 8 in Paper II also shows that pBOS achieves better cost-benefit performance than BOS when the likelihood of reaching the target is less than 50%, because in those situations pBOS can stop experiments early, permitting the redirection of resources.

However, the performances of pBOS and BOS are influenced by the simulation setup, which—along with a discussion about pBOS feature choices—is covered in the following. The setup comprises four simulation features: CIL, $N_{min}$ (the smallest sample size for rehearsal simulations), prior, and tolerance level (TL). After the CIL, $N_{min}$, and prior have been selected, an initial analysis can be conducted to assess how likely it is that the experiment can reach the goal with the given resources. The choice between pBOS and BOS can then be made, and the last feature, TL, can be set.

The choice of value for each feature depends on the specific application. Paper II provided an example that certain choices of these parameters. Here, we provide a short summary and further discuss the general principles of feature value selection. The selection of CIL should be application-specific and based on expert knowledge, the literature, and, possibly, pilot experiments. As shown in Figure 8 in Paper II, a smaller $N_{min}$ results in a higher cost-benefit. However, as shown in Figure 7, a larger $N_{min}$ leads to a higher AUC (the area under the receiver operating characteristic curve is bigger), suggesting that although a larger $N_{min}$ improves the accuracy of the stopping decision, the cost-benefit results are more relevant to the choice of $N_{min}$. Given that the major advantage of pBOS over BOS is that it can stop experiments early so resources can be redirected when appropriate and the increase of AUC with a larger $N_{min}$ is not significant, we recommend choosing a small $N_{min}$ (either as a constant or proportional to the maximum resources).

Naturally, the choice of prior is crucial, as it is an essential component of any Bayesian model. The advantages of Bayesian analysis are often gained by being able to use justified prior information to improve estimates. Using noninformative (i.e., very weak) priors will result in inferences similar to the frequentist approach, thereby forgoing some of the main advantages of the Bayesian method (Lemoine, 2019). We strongly suggest that when using BOS or pBOS, the prior be chosen based on available knowledge, as the more accurate the prior, the better the pBOS performance (as shown in Figure 9 in Paper II). Note, however, that the results also show the dangers of believing too much in previous data: the choice of a wrong highly informative prior results in poor pBOS performance.

As described in Paper II, after selecting the three pBOS features, the user must choose between pBOS and BOS. This choice depends on the difficulty of reaching the target, as pBOS outperforms BOS under specific conditions. The first rehearsal simulations are conducted once $N_{min}$ data have been collected, at which point the probability of reaching the target is estimated based on the available resources (typically the maximum number of participants in human factors research). If the probability of reaching the target is less than 50%, pBOS is recommended, due to its stricter target criteria. Conversely, if the probability exceeds 50%, BOS is preferred. It is easy to revert to using the traditional BOS from an initial pBOS implementation: if the TL is set to 0, early stopping based on rehearsal simulations (which is what pBOS is all about) is disabled. Otherwise, the TL can be selected according to Figure 9 in Paper II, which shows the performance of pBOS relative to BOS across different values. Since the probability of reaching the target is estimated and corresponds to the CIL percentile, Figure 9 provides insights into the estimated performance of pBOS across different TL values for the same CIL percentile value. Selecting the TL that achieves peak pBOS performance (according to Figure 9 in Paper II) ensures that the chosen TL maximizes the efficiency and effectiveness of the pBOS method.

As noted in the Methods section, there are two types of targets used for BOS: effect-size-based and precision-based. For example, the Bayes factor is a commonly used effect-size-based target (Moerbeek, 2021; Schönbrodt & Wagenmakers, 2018; Stefan et al., 2022). Effect-size-based testing, also called hypothesis testing, is particularly appealing in data collection for pre-crash safety system development because it helps determine whether a pre-crash safety system is effective (or which system is more effective in a comparison). Using precision-based target for data collection is a way to make sure you collect enough data to estimate the between- (and possibly within-) driver variability with 'enough' precision. The CIL in Paper II is a precision-based target which can be used to estimate whether the parameter of interest is precise enough for practical use. Both types of targets could be integrated into the pBOS decision-making process.

That said, there is still a debate about the appropriateness of effect-size-based stopping criteria. Although Rouder (2014) suggests there is no problem, issues arise when researchers interpret Bayesian results from a frequentist point of view. In contrast, Heide & Grünwald (2021) argue that for most types of priors, Bayes factor-based optional stopping can violate prior calibration and frequentist type I error guarantees. The use of precision-based stopping criteria in Paper II avoids the controversy surrounding effect-size-based criteria. Whether effect-size-based stopping criteria can effectively be applied (without bias) in this context is an area for further research.

The proposed pBOS method, while not suitable for naturalistic studies, may often be effective for expensive experiments with limited resources, a set budget, and/or a controlled maximum number of trials. It is particularly beneficial in simulator studies—especially those assessing driver responses, since the heterogeneity of human participants means the responses can be highly variable. Consider a situation in which the experimenters aim to minimize between-driver variability in glance behavior (e.g., the distribution of eyes-off-road glance durations) associated with a modified HMI, such as an infotainment system. If drivers' glance behavior toward the HMI can be relatively consistent across drivers, then driver monitoring systems (and concomitant distraction-mitigation strategies) can be much more effective. A CIL target regarding the variability of eyes-off-road glances can be set which quantifies the desired precision of the relevant parameters. Assuming pBOS is used appropriately, the experiment's run can be stopped if reaching the target is too unlikely. The experimenters can then make further modifications to the HMI, updating the prior with data collected from the previous experiment, and run another test (with a new set of participants) to investigate whether the new design reduces between-driver variability (i.e., reaches the target CIL). At any time, design iterations can stop and the design with the lowest between-driver variability can be chosen. Note that, for such a study, it is also possible to use a factorial experiment design instead of a sequential experiment design. Although not within the scope of the current work, it may be possible for an adaptation of the pBOS method to be an integral part of factorial precision-target experiment designs.

Another use of precision-based experiment designs is the development of a safety system based on CZB, as in Paper I. During the work for that paper, it became clear that there are very few studies quantifying CZBs, and those that exist rarely capture both between-driver and within-driver variability. One exception is the Bärgman, Smith, et al. (2015) study, which attempted to establish CZBs in terms of acceleration for left-turn-across-path/opposite direction in car-to-car scenarios. However, that study used traditional frequentist sample size determination. If a precision-based BOS experiment design were used to determine the between-driver distribution of CZB in some comparable variable space (e.g., lateral acceleration), the variability of the CZB could be specified in advance: the CIL target could be the desired precision of the mean or the median, or even a certain percentile of the lateral acceleration. One could then either run the experiment until this precision is achieved or (assuming it was appropriate to use pBOS) stop the experiment if it becomes unlikely that the desired precision can be reached with the available resources. In the latter case, it is necessary to either: a) acknowledge that the desired precision is not achievable and revise the target, b) allocate more resources and continue the experiment, or c) find a CZB metric with less variability that still captures the essence of driver comfort. Determining a CZB distribution using a precision-based experiment design ensures that the threshold (e.g., the 80th percentile CZB across drivers; see Olleja et al., 2023) accurately describes the true underlying distribution, according to the CIL precision deemed 'good enough'.

Although the sample size needed to reach a specific target is typically smaller for sequential sample size determination methods than for frequentist methods, sequential methods can be challenging and less familiar to researchers in the traffic safety domain. Since pBOS and BOS iteratively analyze newly collected samples as an integral part of the actual experiment (unlike the frequentist fixed sample size method), they require more computation. Further, frequentist methods use equations with relatively well-known inputs such as effect size, sample variance, and confidence interval coverage rate to calculate the sample size (Ryan, 2013). Sequential methods can be more complicated, but they have better cost-benefit performance. Increasing the awareness of the methods' advantages can motivate people to use them.

A disadvantage of the frequentist approach is that sample size must be determined up front and remain unchanged; if the collected data are inconclusive, the end result is unsatisfactory. In contrast, BOS continues until a goal is reached, but if the goal is very difficult to reach the experiment can be never-ending. The proposed pBOS method addresses this issue by allowing early stopping (and subsequent resource reallocation) when it is unlikely that the target will be reached with the available resources. Theoretically, pBOS might predict a high likelihood of reaching the target throughout the data collection, even when the target is never met. However, this scenario is unlikely because the predictions are updated with each new data point, so that the trend of the predictions can be monitored. Further, even in this worst case, the final results should be relatively close to the target; otherwise, pBOS would have stopped the experiment early. Additionally, even if the target was not met, the collected data can still be valuable.

## 4.4 Efficient sampling methods for scenario generation

It is advantageous to improve efficiency in all aspects of traffic safety research, not just in safety system development. In the virtual environment, the sampling space is typically large, so there is a need for an efficient sampling method. Scenario generation can be applied to both safety impact assessment and safety verification and validation. Safety impact assessments focus on quantifying a safety system's performance, while safety verification and validation verify system functionality as part of company-internal decision-making and regulatory processes for approving use on public roads (Ma et al., 2022; Thacker et al., 2004). For the former, the representativeness of generated scenarios is particularly important, as a system's overall safety impact (e.g., in terms of crash or injury risk reduction) must be accurately quantified. In contrast, safety verification and validation emphasize the system's safe deployment, to ensure its reliability in diverse situations (such as corner cases and worst-case scenarios).

There are many scenario generation methods that can be used for both applications, such as those in Table 1. At the highest level of categorization one can distinguish between knowledge-based and data-driven methods. Riedmaier et al. (2020) found data-driven methods superior for scenario representativeness, assessment transferability, and scenario space expansion. (Baseline approaches A, B, and C are all data-driven approaches.) Knowledge-based methods, which rely on standards and guidelines, often miss some scenarios. However, data-driven approaches are not without their disadvantages: they require extensive data, complex processing pipelines, and computational resources. This work is focused on data-driven methods, and, consequently, knowledge-based approached will not be covered further.

The sampling methods for scenario generation in this work are intended for use in safety impact assessments. Among the various methods, previous literature has demonstrated that N-wise sampling covers the whole sampling space—but it is inefficient. Simple random sampling can be more efficient than N-wise sampling and it is easy to implement, but it is less efficient than importance sampling (Owen & Zhou, 2000; Tokdar & Kass, 2010). As a result, importance sampling is the go-to method used in most applications (de Gelder & Paardekooper, 2017; Jesenski et al., 2020; Wulfe et al., 2018; Xu et al., 2018; Di. Zhao et al., 2018). However, as mentioned, for optimal performance importance sampling requires accurate information about the parameters of the underlying distribution before implementation, and many applications lack that knowledge. Therefore, the motivation for this work is first, to improve scenario generation efficiency by developing sampling methods that perform well even without such information, and second, to investigate the impact of incorporating domain knowledge on the efficiency of the sampling methods.

The machine-learning-based active sampling method proposed, implemented, and applied in Papers III and IV outperformed the traditional importance sampling method in a straightforward application, requiring fewer simulations to reach the same estimated error level. Unlike the active sampling method, the importance sampling method has a fixed sampling scheme, so newly collected samples do not add more information to the scheme itself. Further, if the fixed sampling scheme is very inaccurate, the sampling method may perform even more poorly. In Paper III, two different importance sampling methods were compared with the proposed active sampling. The two importance sampling methods performed differently, but both were worse than active sampling (as shown in Figure 5 in Paper III).

Interestingly, in a more complex application discussed in Paper IV, adding knowledge of certain patterns in the data (which we call “logic”) benefited importance sampling more than active sampling; in some cases, importance sampling even outperformed active sampling. In other words, when important sampling was enhanced by known structures in the data, thus improving its underlying model, it performed particularly well. This result further strengthens the conclusion that importance sampling works well in cases where the parameter space is well understood but less well when the structure of that space is unknown (and possibly complex).

In addition, the results further show that stratification also improves the efficiency of both importance sampling and active sampling. Stratification balances case representation, accounts for heterogeneity, and ensures that targets are estimated accurately, which together promise a reduction in variance. In fact, Park et al. (2024) and Jing et al. (2015) also showed that stratification is an effective variance reduction method in settings with substantial heterogeneity across strata. Interestingly, even though stratification improves the efficiency of both importance sampling and active sampling, it improves the latter more. This result may be an indication that active sampling without stratification has already learned the optimal structure through iterations, so the benefit of stratification is less marked. Active sampling is always better with stratification, but the difference decreases as sample size increases, because the machine learning model updates the prediction models over iterations. However, the difference for importance sampling with and without stratification is relatively constant as the sample size increases (no model update occurs). This suggests that, as we write in Paper IV, when stratification is not available, active sampling is preferred over importance sampling, as it requires a smaller sample size to reach the same precision target.

In importance sampling, the prior information determines the parameter density for sampling. The prior information used in both importance sampling methods in our application comprised a glance-deceleration distribution and maximum impact speed for each case (for severe importance sampling in Paper III). Active sampling, on the other hand, needs a set of samples to initialize the sampling schemes, instead of prior information to build them. The initial samples in the scenario generation application in Paper III were made up of the most severe scenarios for each case. However, when no information about the outcome can be estimated, it may be more difficult to pick initial samples that provide the desired information. In those applications, random sampling or space sampling can be used to pick the initial samples. Because active sampling iteratively updates the sampling scheme, the more simulations run, the more accurate the prediction model and the more efficient the sampling scheme. As a consequence, active sampling may perform less well than importance sampling for a small sample size but is likely to perform better as sample sizes increase. If the total sample size required is small and prior information is known, importance sampling is likely a good choice; however, if the sample size required is large and/or prior information is missing, active sampling is likely the better choice.

It should nonetheless be acknowledged that there are several other ways to sample efficiently. For example, de Gelder & Paardekooper (2017) used importance sampling based on kernel density estimation to generate scenarios using available real-world driving data. The method showed that importance sampling has the potential to identify critical scenarios, but the method requires a large amount of data, while the active sampling approach typically does not.

One of the challenges of using active sampling in simulation is that simulations are often run in batches to reduce wall-clock time. That is, parallel computations improve time efficiency (Horowitz, 2014), but active sampling benefits from sequential results. The best balance is likely to be application-dependent and require some exploration. In our application in Paper IV, we tested batch sizes of 44, 132, and 440 (one, three, and ten times the number of original crashes). We found that in many of the comparisons between batch sizes of 44 and 132, the results were similar, but the largest batch size substantially decreased the benefits of active sampling. It is crucial to balance the reduced sampling efficiency against the corresponding reduction in computational load, particularly in time-consuming simulations. Although virtual simulations to assess the impact of pre-crash safety systems can be time-consuming, other types of safety assessments, such as in-crash simulations with human body models (HBMs; Pipkorn et al., 2022), are even more computationally demanding and costly. Methods such as active sampling may therefore be particularly useful in in-crash safety assessments that use HBMs.

The performance of different sampling methods can be evaluated by measuring the root mean squared error (RMSE) of the key target estimates (e.g., estimated injury risk) relative to the ground truth. Figure 5 in Paper III shows the RMSE for eight different sampling methods. In practice, however, the ground truth is unknown; thus, methods (such as the three methods for estimating variance in Paper III) that can determine stopping criteria without ground truth information are needed. Descriptions about the stopping criteria used, and instructions for how to set them, are not always included in scenario generation literature. For example, de Gelder and Paardekooper (2017) planned to provide instructions for stopping as part of future work.

In this work, the bootstrap method (Efron, 2007), the Martingale method (B. M. Brown, 1971), and the classical survey method—also known as the Sen-Yates-Grundy estimator (Sen, 1953; Yates & Grundy, 1953), were applied to estimate variance during the 'experiment' (simulating not having ground truth data). The Martingale method and the classical survey method are parametric. In contrast, the bootstrap method is non-parametric; it is also more time-consuming and computationally demanding than the other two methods. The 95% coverage results of Paper III show that the three methods perform differently depending on batch size. For large batch sizes, the Martingale method performed worse than the other two, but those two performed well for both small and large batch sizes. Thus, for large batch sizes, we recommend the classical survey method and for small batch sizes, we recommend the Martingale method.

We also explored the incorporation of domain-knowledge-based logic into both active and importance sampling. The logic was based on the fact that less extreme parameters lead to less severe outcomes, while more extreme parameters lead to more severe outcomes. In Paper IV, using this logic, we showed that the addition of logic improved importance sampling more than it did active sampling, to the point where importance sampling outperformed active sampling for some targets. As shown in Paper IV, domain knowledge-based adaptive sample space reduction (ASSR) logic in our application enhances the performances of the proposed active sampling method and traditional importance sampling. The effectiveness of the ASSR logic depends on the specific application, and particularly on the monotonic relationship in crash causation modeling between parameters and crash outcomes. Computational crash causation models include parameter distributions of the crash mechanisms under study. Generally, non-crash scenarios are generated with parameters from the less severe (left) tail of the distribution, while crash scenarios are generated from the more severe (right) tail. There is a lower bound in the parameter distribution where the first crash occurs, and an upper bound beyond which the crash severity does not increase (at least for rear-end crashes). Identifying these bounds helps avoid unnecessary simulations.

Until now, domain knowledge has been discussed in terms of using ASSR logic to reduce the number of simulations. There is, however, another type of domain knowledge that is needed when using sampling methods for finite population inference (for example): domain knowledge about the conditions for stopping the sampling. When the sampling is optimized on a safety impact assessment metric (e.g., impact speed reduction, crash avoidance rate reduction, or injury risk reduction), a precision target based on the estimated standard error for that metric can be chosen as a stopping criterion (Altman & Bland, 2005). The stopping criteria may be based on, for example, a region of practical equivalence (ROPE) or coefficient of variance (such as 2.5% of the estimated mean value of the assessment metric). The choice of ROPE needs to be based on domain knowledge, which may consist of a few experts agreeing on what would constitute a practical equivalence for the specific assessment metric; that is, within what bounds (in the unit of the metric) could two values be considered the same.

Naturally, domain knowledge is used extensively in research, but Kerrigan et al. (2021) indicated that in the application of machine learning to various problems, there is often a lack of documentation about what domain knowledge is included and what the sources of the knowledge are—in spite of the fact that such documentation is needed for transparency. In this work we have tried to describe the rationales for the different domain knowledge-based decisions as clearly as possible.

Although this work has focused on pre-crash safety system development, the sampling method proposed is also well suited for use in the development of in-crash protection systems. For example, this method is likely to substantially reduce the number of simulations required (and thus the overall simulation time) if applied to modeling injury risk assessment using HBMs (Östh et al., 2022). Because HBM simulations are very time-consuming, this application would also represent an increase in efficiency.

## 4.5 Artificial intelligence for traffic safety development

Artificial intelligence (AI) is increasingly being used in the development and assessment of pre-crash safety systems. It is already integrated into automated vehicles' autonomous driving functions, processing data from sensors, perceiving other road users' intentions (including path prediction), and controlling driving decisions (Grigorescu et al., 2020; Z. Li et al., 2024; J. Zhao et al., 2024). The role of AI in making system assessments more efficient, thereby improving the overall effectiveness of system development efforts, is particularly noteworthy in the context of this work. This section first describes how AI was used in this work and then, more broadly, discusses relevant applications for AI in traffic safety development.

### 4.5.1 The use of AI in this work

In this work, AI (in the form of the machine learning models in Papers III and IV) was used to enhance the efficiency of scenario generation by improving sampling methods. Recall that these models could update the sampling strategy with new predictions based on new data. As previously described in Paper III, we proposed a machine-learning-assisted sampling method that uses already-collected data to predict the outcomes of unknown data. This prediction, based on a machine learning method, was shown to have a sampling efficiency higher than that of traditional sampling methods when domain-specific knowledge (e.g., ASSR logic and stratification) is not applied.

As mentioned in Section 4.4, machine learning methods can be combined with domain knowledge-based logic to improve safety assessment efficiency. Actually, current AI algorithms for autonomous driving often rely on a combination of machine learning methods and rule-based methods, and many argue that this trend will continue

(Aksjonov & Kyrki, 2023; Finesso et al., 2016; Likmeta et al., 2020). After combining domain knowledge-based logic in the sampling application in Paper IV, we, as previously described, found that sampling efficiency improved for both traditional importance sampling and machine-learning-assisted sampling.

### 4.5.2 Future AI applications related to this work

Another aspect of domain knowledge is understanding crash causation in order to develop crash-causation models, which embody the relationship between input parameters and crash outcomes. Irrelevant parameter spaces can be excluded from sampling, and additional knowledge about whether a specific configuration leads to a crash can be incorporated into machine learning models to provide more accurate predictions.

Analyzing large datasets can facilitate the understanding of crash causation, leading to the identification of factors that influence crash occurrences, resulting in lives and costs saved. The larger and more heterogeneous the dataset, the easier it is to uncover numerous factors (including road users, vehicles, the environment, and the infrastructure) and their underlying relations. Prati et al. (2017) applied a Chi-square automatic interaction-detection decision-tree technique and Bayesian network analysis to official statistics in order to investigate factors predicting the severity of bicycle crashes in Italy. That study illustrates the effectiveness of data mining techniques for understanding crash severity. Chen et al. (2018) developed crash prediction models using refined temporal data (hourly records) to characterize the time-varying nature of contributing factors. Their findings underscore the importance of both time-varying (e.g., hourly traffic volume) and site-varying factors (e.g., curvature and speed limit) in influencing crash likelihood. Das et al. (2020) applied association rules mining to discover crash patterns during rainy weather using crash data from Louisiana (2004 to 2011). They found that “single-vehicle run-off road” is the predominant crash type during such conditions. Collectively, these studies suggest that big data and AI technology can, with human guidance, effectively help identify crash factors to be incorporated into crash causation models, which can be further used in scenario generation for safety assessment.

AI can be used to explore large datasets to model driver behavior. Miyajima & Takeda (2016) illustrated the usefulness of data-driven approaches and large datasets by applying machine learning methods to extensive real-world driving data collected over more than 15 years. M. S. Wang et al. (2016) explored the detection of drowsy behavior using artificial neural network and random forest algorithms, finding that recording lateral and longitudinal accelerations at 20-second intervals was optimal for detecting drowsiness with the random forest algorithm. This research highlights the importance of selecting appropriate input parameters and AI algorithms for behavior detection. Tran et al. (2018) proposed a driver distraction detection system using a camera and deep convolutional neural networks capable of distinguishing between normal and distracted driving. Together, these studies show that AI and big data can be used to detect risky behaviors, such as drowsiness and distraction, and to model driver behavior. These insights can be applied to developing safety systems that alert drivers and to creating virtual safety assessments through scenario generation.

Since machine learning methods learn from data to make decisions or predictions, they can also be used to create test and assessment scenarios. For example, Y. Li et al. (2023) developed two types of generative adversarial network models capable of learning from the observed data space, and used them to generate test scenarios for automated vehicles. In addition to generating concrete scenarios directly, machine learning methods can also generate concrete scenarios by classifying or clustering logical scenarios. For instance, Montanari et al. (2021) used the supervised machine learning method recurrent neural network to support scenario detection, and Kruber et al. (2023) used the unsupervised random forest clustering technique to categorize scenarios.

In general, AI has demonstrated its potential for traffic safety development. It has been used in various ways to generate testing scenarios for safety assessment. As we have not found any accounts of crash outcome validation performed on machine-learning-based scenario generation, future validation based on crash outcome is required to confirm the utility of these methods for generating representative baseline crashes (for subsequent use in virtual safety impact assessment). In addition, there are various AI applications for investigating crash factors and modeling both crash causation and driver behavior. All these different aspects, to a large extent feeding into scenario generation for virtual safety assessment, can benefit traffic safety development.

## 4.6 Limitations and future work

This section provides a short overview of the main limitations of each paper, starting with the limitations related to sample size and sensor and vehicle model simplifications in Paper I. The limited application issue for Paper II is then discussed, followed by the limitation of applied crash causation modeling for complex scenario generation and the potential use of more complex estimation targets (e.g., parametric distribution functions) for Papers III and IV. Finally, a reflection on the expansion of the methods' applications to conflict avoidance is provided.

Both the counterfactual simulation and the scenario generation methods in this work can only be used when there are enough available crash data. Importantly, the accuracy and representativeness of the simulation results are also influenced by the amount of data. Admittedly, the CZB-based pre-crash safety systems in Paper I were assessed using only 93 car-to-TW crashes from the SHUFO database. However, the focus was not on providing an accurate assessment of the safety impact; rather, it was on assessing the impact of including CZB-based behavior models in AEB algorithms. We argue that even if the number of crashes is relatively low, the conclusions about the benefits of incorporating such models are valid and would not change if more crashes were included.

Further, the AEB algorithms that include CZB models were assessed with SHUFO crash data collected in a rural area of Shanghai, China. Because crashes can have different characteristics across countries and regions, the AEB performance might differ when applied to European data. However, the contributions of Paper I are not about the explicit performance of the AEB (which is likely to be different depending on, for example, infrastructure, road-user demographics, and traffic culture), but rather an assessment of the benefits of the inclusion of CZB-based AEB over traditional deceleration-based AEB. Consequently, the origin of the data and the traffic system characteristics should be of secondary importance.

Similar to the crash characteristics, driving behaviors can vary across countries and regions as well. The CZB models used in this work are based on studies and experiments conducted in Europe. As there are no available studies of CZB models for Chinese road users, it is not possible to investigate their inclusion in the AEB algorithm designs even though Chinese crash data were used for the AEB assessment. This mismatch between the data used for developing the CZB-based AEB algorithm and the crash data used in the assessment, is one of the limitations of the study. However, as the main aim of Paper I is to demonstrate the general benefit of using CZB in AEB algorithms, the behavioral characteristics, too, are of secondary importance. Because the CZBs were defined as upper boundaries instead of distributions, they are likely to contain most Chinese drivers' CZBs as well.

Another main limitation of all four papers is that the sensors used in the virtual simulations were idealized. It was assumed that there were no uncertainties or errors in the detection and tracking of other road users; they only had geometrical constraints (e.g., sensor field of view, when applicable). However, in the real world, sensors are not ideal; as a result, the performance of the pre-crash safety system in the simulations is likely to be substantially better than in the real world. Possible issues with detection and/or tracking should be considered in future simulations in order to estimate system performance more realistically. In addition, the vehicle models used in Paper I were simplified using a single-track bicycle model for both the car and the PTW, which may have influenced the simulation results somewhat—but here too we do not think that more realistic models would have changed the overall conclusions of the paper.

While the proposed pBOS can improve experiment cost-effectiveness, it is limited to particular applications whose experiment target is difficult to reach with the available resources. Further, the performance of the method is influenced by the prior, which in some applications is unknown; in this circumstance the advantage of pBOS is reduced.

The rear-end crash scenario in Papers III and IV, although relatively simple, was ideal for demonstrating scenario generation—as well as the active sampling method in scenario generation. However, there is much work needed to develop and validate crash causation models for other scenarios. The glance-deceleration model is a suitable crash causation model for highway rear-end crashes such as the data used in Papers III and IV, since lack of visual attention toward the forward roadway is a significant crash mechanism. However, the model is not suitable for more complicated scenarios, such as car-to-VRU interactions, because the causation mechanisms of car-to-VRU crashes are more complex (see, e.g., Habibovic et al., 2013). More research is needed to enable scenario generation for these and other interactions (particularly those including intersections). For rear-end crashes as well as some other scenarios, there is a dominant crash causation mechanism (e.g., inattention); for others, several mechanisms play significant roles (e.g., car-to-VRU crashes). In the latter case, each mechanism needs to be modeled. In addition, the proportion of crashes that each "causes" in the real world needs to be considered when the scenario generation framework is created for a particular concrete scenario (Bärgman et al., 2024). In summary, to be able to apply the active sampling method for behavior-model-based representative scenario generation for other scenario types, parameterized crash causation models that are validated against real-world crash outcomes are needed.

Further, the proposed active sampling method presented in Papers III and IV optimizes the sampling on the mean targets. Although it is also possible to optimize the sampling on the estimation of the parameters in parametric distribution functions (e.g., a log normal impact speed or injury risk distribution), it was not done in this work. In order to do so, one must understand which distribution is likely to be a good fit. More complex distribution mixes can be included, but more work is needed to integrate them into the proposed sampling method.

The scope of this thesis is limited to crash avoidance safety assessment. In the future, conflict avoidance and crash avoidance should be investigated together. Although crash avoidance is important for higher levels of automation, the conflict avoidance components of higher levels of automation are very crucial for safety (Scanlon et al., 2022). Future work should study both actual safety system evaluation (as in Paper I) and methods for sampling and scenario generation for conflict avoidance systems.

# 5 Conclusions

This thesis explores ways to enhance traffic safety through computational driver models, focusing on pre-crash safety systems (integral to ADAS and ADS) and the development of safety assessment methods for these systems. The research is described in four papers addressing the three objectives: 1) evaluating driver models based on comfort-zone boundaries (CZB) in crash-avoidance system design, 2) exploring efficient data collection for use in safety system development, and 3) optimizing sampling methods for scenario generation using a crash causation model based on driver behavior.

Paper I demonstrates that CZB-based AEB systems outperform traditional AEBs by nearly 50%, reducing both crash risk and nuisance interventions. The findings indicate that integrating driver behavior models into crash-avoidance systems can substantially enhance their safety performance. Specifically, incorporating CZBs into pre-crash safety systems is a promising approach for improving ADAS and ADS, potentially saving lives. Additionally, the study revealed similarities in residual car-to-PTW crashes across algorithms, which may simplify the design of future in-crash protection systems.

Efficient data collection is crucial for the rapid and cost-efficient development of good safety systems. Paper II introduces the predictive Bayesian optional stopping (pBOS) method, which improves cost-benefit ratios by up to 118% over traditional Bayesian methods. pBOS is particularly advantageous when the statistical target of an experiment is unlikely to be reached. The improved efficiency can lead to better pre-crash safety systems by using improved driver behavior model components (e.g., with respect to the precision and accuracy of driver behavior metrics), obtained through more appropriate allocation of experimental resources. Employing pBOS for data collection in the development of pre-crash safety systems can, under certain conditions, allow resources to be reallocated towards further system optimization or other safety system development efforts.

Papers III and IV propose and evaluate an optimal sampling method for scenario generation in virtual safety assessments. This machine-learning-assisted active sampling method enhances efficiency by requiring fewer simulations than traditional methods while maintaining precision, even with limited prior information. Further, Paper IV demonstrates that including domain knowledge by incorporating adaptive sample space reduction logic and stratification in the sampling process can substantially improve the sampling efficiency. These improvements streamline scenario generation, potentially accelerating the development and deployment of systems, ultimately leading to safer roads. That is, as with pBOS, the appropriate use of active sampling by developers of pre-crash safety systems can lower development costs and/or allow resources to be reallocated towards further optimization of system performance or other safety system development initiatives.

Both pBOS and the optimal sampling method have potential beyond their applications in pre-crash safety, such as improving the efficiency of in-crash safety system development using HBM models for injury risk assessment. Overall, the advancements presented in this thesis can lead to faster development cycles and more effective safety systems, contributing to the ultimate goal of saving lives on our roads.

# References


Ahmed, A., Sadullah, A. F. M., & Yahya, A. S. (2019). Errors in accident data, its types, causes and methods of rectification-analysis of the literature. *Accident Analysis and Prevention*, *130*, 3–21. https://doi.org/10.1016/j.aap.2017.07.018

Aksjonov, A., & Kyrki, V. (2023). A Safety-Critical Decision-Making and Control Framework Combining Machine-Learning-Based and Rule-Based Algorithms. *SAE International Journal of Vehicle Dynamics, Stability, and NVH*, *7*(3), 287–299. https://doi.org/10.4271/10-07-03-0018

Altman, D. G., & Bland, J. M. (2005). Standard deviations and standard errors. *BMJ*, *331*(7521). https://doi.org/10.1136/bmj.331.7521.903

Amersbach, C., & Winner, H. (2019). Defining required and feasible test coverage for scenario-based validation of highly automated vehicles. *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019*, 425–430. https://doi.org/10.1109/ITSC.2019.8917534

Antin, J. F., Lee, S., Perez, M. A., Dingus, T. A., Hankey, J. M., & Brach, A. (2019). Second strategic highway research program naturalistic driving study methods. *Safety Science*, *119*, 2–10. https://doi.org/10.1016/j.ssci.2019.01.016

Anund, A., Ihlström, J., Fors, C., Kecklund, G., & Filtness, A. (2016). Factors associated with self-reported driver sleepiness and incidents in city bus drivers. *Industrial Health*, *54*(4), 337–346. https://doi.org/10.2486/indhealth.2015-0217

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

Bakowski, D. L., Davis, S. T., & Moroney, W. F. (2015). Reaction Time and Glance Behavior of Visually Distracted Drivers to an Imminent Forward Collision as a Function of Training, Auditory Warning, and Gender. *Procedia Manufacturing*, *3*, 3238–3245. https://doi.org/10.1016/j.promfg.2015.07.875

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

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

Bärgman, J., & Victor, T. (2019). Holistic assessment of driver assistance systems: how can systems be assessed with respect to how they impact glance behaviour and collision avoidance? *IET Intelligent Transport Systems*, *14*(9), 1058–1067. https://doi.org/10.1049/iet-its.2018.5550

Beglerovic, H., Ravi, A., Wikström, N., Koegeler, H.-M., Leitner, A., & Holzinger, J. (2017). Model-based safety validation of the automated driving functio highway pilot. *8th International Munich Chassis Symposium*. https://doi.org/10.1007/978-3-658-18459-9_21

Bellem, H., Schönenberg, T., Krems, J. F., & Schrauf, M. (2016). Objective metrics of comfort: Developing a driving style for highly automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *41*. https://doi.org/10.1016/j.trf.2016.05.005

Bellem, H., Thiel, B., Schrauf, M., & Krems, J. F. (2018). Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits. *Transportation Research Part F: Traffic Psychology and Behaviour*, *55*, 90–100. https://doi.org/10.1016/j.trf.2018.02.036

Benderius, O., Berger, C., & Malmsten Lundgren, V. (2018). The Best Rated Human-Machine Interface Design for Autonomous Vehicles in the 2016 Grand Cooperative Driving Challenge. *IEEE Transactions on Intelligent Transportation Systems*, *19*(4), 1302–1307. https://doi.org/10.1109/TITS.2017.2749970

Benmimoun, M., Pütz, A., Zlocki, A., & Eckstein, L. (2013). Impact Assessment of Adaptive Cruise Control and Forward Collision Warning. *ATZ Worldwide*, *115*, 74–80. https://doi.org/10.1007/s38311-013-0051-5

Bianchi Piccinini, G., 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

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 (L3Pilot Deliverable D7.4)*. https://l3pilot.eu/fileadmin/user_upload/Downloads/Deliverables/Update_14102021/L3Pilot-SP7-D7.4-Impact_Evaluation_Results-v1.0-for_website.pdf

Brännström, M., Coelingh, E., & Sjöberg, J. (2010). Model-based threat assessment for avoiding arbitrary vehicle collisions. *IEEE Transactions on Intelligent Transportation Systems*, *1524–9050*, 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 1*, *7*(1), 87–106. https://doi.org/10.1504/IJVS.2014.058243

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

Brown, B. M. (1971). Martingale Central Limit Theorems. *The Annals of Mathematical Statistics*, *42*(1), 59–66. https://www.jstor.org/stable/2958462

Brown, T. L., Lee, J. D., & McGehee, D. V. (2001). Human performance models and rear-end collision avoidance algorithms. *Human Factors*, *43*(3). https://doi.org/10.1518/001872001775898250

Bucsuházy, K., Matuchová, E., Zůvala, R., Moravcová, P., Kostíková, M., & Mikulec, R. (2020). Human factors contributing to the road traffic accident occurrence. *Transportation Research Procedia*, *45*. https://doi.org/10.1016/j.trpro.2020.03.057

Cai, J., Deng, W., Guang, H., Wang, Y., Li, J., & Ding, J. (2022). A Survey on Data-Driven Scenario Generation for Automated Vehicle Testing. *Machines*, *10*(11), 1101. https://doi.org/10.3390/machines10111101

Chajmowicz, H., Saadé, J., Cuny, S., Rosén, E., Zhao, Y., Ito, D., Mizuno, K., Chajmowicz, H., Saadé, J., Cuny, S., Rizzi, M. M. C., Rizzi, M. M. C., Kullgren, A., & Algurén, B. (2019). Prospective assessment of the effectiveness of autonomous emergency braking in car-to-cyclist accidents in France. *Traffic Injury Prevention*, *20*(sup2), S20–S25. https://doi.org/10.1080/15389588.2020.1730827

Chen, F., Chen, S., & Ma, X. (2018). Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. *Journal of Safety Research*, *65*. https://doi.org/10.1016/j.jsr.2018.02.010

Cicchino, J. B. (2017). Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. *Accident Analysis and Prevention*, *99*(2017), 142–152. https://doi.org/10.1016/j.aap.2016.11.009

Cicchino, J. B. (2022). Effects of automatic emergency braking systems on pedestrian crash risk. *Accident Analysis and Prevention*, *172*, 106686. https://doi.org/10.1016/j.aap.2022.106686

C-NCAP. (2018). *C-NCAP Management Regulation 2018*. http://www.c-ncap.org/

C-NCAP. (2020). *C-NCAP Management Regulation 2021*. http://www.c-ncap.org/

Coelingh, E., Jakobsson, L., Lind, H., & Lindman, M. (2007). Collision Warning With Auto Brake - a Real-Life Safety Perspective. *Proceedings of the 20th International Conference on Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration (No. 07-0450)*. https://trid.trb.org/View/1365284

Costa, L., Perrin, C., Dubois-lounis, M., Serre, T., Costa, L., Perrin, C., Dubois-lounis, M., Serre, T., & Nacer, M. (2019). Modeling of the powered two-wheeler dynamic behavior for emergency situations analysis. *VSDIA 16th International MINI Conference on Vehicle System Dynamics*. https://hal.science/hal-02072140/file/VSDIA%202018-Costa-Serre.pdf

Dagan, E., Mano, O., & Stein, G. P. (2004). Forward collision warning with a single camera modeling acceleration momentary time to contact. *IEEE Intelligent Vehicles Symposium*, 37–42.

Das, S., Dutta, A., & Sun, X. (2020). Patterns of rainy weather crashes: Applying rules mining. *Journal of Transportation Safety and Security*, *12*(9). https://doi.org/10.1080/19439962.2019.1572681

Davidian, M., & Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association*, *82*(400). https://doi.org/10.1080/01621459.1987.10478543

de Gelder, E., & Paardekooper, J. P. (2017). Assessment of Automated Driving Systems using real-life scenarios. *IEEE Intelligent Vehicles Symposium, Proceedings*, *Iv*, 589–594. https://doi.org/10.1109/IVS.2017.7995782

de Heide, R., & Grünwald, P. D. (2021). Why optional stopping can be a problem for Bayesians. In *Psychonomic Bulletin and Review* (Vol. 28, Issue 3). https://doi.org/10.3758/s13423-020-01803-x

Dean, M. E., Lubbe, N., Fredriksson, R., Sternlund, S., & Gabler, H. C. (2023). Assessing the applicability of impact speed injury risk curves based on US data to defining safe speeds in the US and Sweden. *Accident Analysis and Prevention*, *190*. https://doi.org/10.1016/j.aap.2023.107151

Deng, B., Wang, H., Chen, J., Wang, X., & Chen, X. (2013). Traffic accidents in shanghai - general statistics and in-depth analysis. *In Proceedings of the 23rd International Technical Conference on the Enhanced Safety of Vehicles*, *Figure 4*, 27–30.

Ding, C., Bohman, K., Zhang, W., Li, Y., & Zhao, X. (2016). *Rear seated occupants in near-side impacts-study based on Chinese accident databases*. https://www.researchgate.net/publication/315695253

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(10). https://doi.org/10.1073/pnas.1513271113

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., & Knipling, R. R. (2006). *The 100-car naturalistic driving study phase II – results of the 100-car field experiment (No. DOT HS 810 593)*.

Doecke, S. D., Baldock, M. R. J., Kloeden, C. N., & Dutschke, J. K. (2020). Impact speed and the risk of serious injury in vehicle crashes. *Accident Analysis and Prevention*, *144*. https://doi.org/10.1016/j.aap.2020.105629

Donnelly, B. R., Galganski, R. A., Lawrence, R. D., & Blatt, A. (2001). Crash reconstruction and injury-mechanism analysis using event data recorder technology. *Annual Proceedings. Association for the Advancement of Automotive Medicine. Association for the Advancement of Automotive Medicine*, *45*, 331–351. https://europepmc.org/article/med/12214359

Dozza, M., Aderum, T., Kovaceva, J., Rasch, A., Thalya, P., & Knauss, A. (2020). *MICA - modelling interaction between cyclists and automobiles: public report.* https://trid.trb.org/View/1875996

Duan, J., Li, R., Hou, L., Wang, W., Li, G., Li, S. E., Cheng, B., & Gao, H. (2017). Driver braking behavior analysis to improve autonomous emergency braking systems in typical Chinese vehicle-bicycle conflicts. *Accident Analysis and Prevention*, *108*. https://doi.org/10.1016/j.aap.2017.08.022

Eclipse Foundation. (2024). *Stochastic Cognitive Model (SCM).* https://projects.eclipse.org/projects/automotive.scm

Edwards, M., Nathanson, A., Carroll, J., Wisch, M., Zander, O., & Lubbe, N. (2015). Assessment of integrated pedestrian protection systems with autonomous emergency braking (AEB) and passive safety components. *Traffic Injury Prevention*, *16*. https://doi.org/10.1080/15389588.2014.1003154

Efron, B. (2007). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, *7*(1). https://doi.org/10.1214/aos/1176344552

Eichberger, A., Tomasch, E., Hirschberg, W., & Steffan, H. (2011). Potentials of Active Safety and Driver Assistance Systems. *ATZ Worldwide EMagazine*, *113*, 56–63. https://doi.org/10.1365/s38311-011-0079-3

Elvik, R., & Mysen, A. (1999). *Incomplete Accident Reporting Meta-Analysis of Studies Made in 13 Countries*. https://doi.org/10.3141/1665-18

Elvira, V., Martino, L., Luengo, D., & Bugallo, M. F. (2019). Generalized multiple importance sampling. *Statistical Science*, *34*(1). https://doi.org/10.1214/18-STS668

Esenturk, E., Khastgir, S., Wallace, A., & Jennings, P. (2021). Analyzing real-world accidents for test scenario generation for automated vehicles. *IEEE Intelligent Vehicles Symposium, Proceedings*, *2021-July*. https://doi.org/10.1109/IV48863.2021.9576007

Euro NCAP. (2022). *European new car assessment programme assessment protocol – safety assist collision avoidance*. http://www.euroncap.com/en

Euro NCAP. (2023). *European new car assessment programme- test protocol – AEB car-to-car systems*. https://www.euroncap.com/media/79864/euro-ncap-aeb-c2c-test-protocol-v43.pdf

European Commission. (2022). *Vehicles and VRU virtual eValuation of road safety*. https://cordis.europa.eu/project/id/101075068

Fahrenkrog, F. , Das, A. , Sander, D. , Bärgman, J. , Urban, M. , Pohl, M. , & Glasmacher, C. (2024). *Prospective Safety Assessment Framework – Instruction, Deliverable D2.1 of the Horizon Europe project V4SAFETY.* https://v4safetyproject.eu/outputs

FHWA SHRP2 Implementation Team. (2024). *A Brief Look at the History of SHRP2*. https://shrp2.transportation.org/pages/History-of-SHRP2.aspx

Fildes, B., Keall, M., Bos, N., Lie, A., Page, Y., Pastor, C., Pennisi, L., Rizzi, M., Thomas, P., & Tingvall, C. (2015). Effectiveness of low speed autonomous emergency braking in real-world rear-end crashes. *Accident Analysis and Prevention*, *81*, 24–29. https://doi.org/10.1016/j.aap.2015.03.029

Finesso, R., Spessa, E., & Venditti, M. (2016). An unsupervised machine-learning technique for the definition of a rule-based control strategy in a complex HEV. *SAE International Journal of Alternative Powertrains*, *5*(2). https://doi.org/10.4271/2016-01-1243

Fries, A., & Fahrenkrog, F. (2021). Validation and verification of the stochastic cognitive driver model. *ACIMobility Summit*. https://aci-mobility.de/fileadmin/user_upload/ACImobility/Papers/acimobility_2021_fries_bmw.pdf

Fries, A., Fahrenkrog, F., Donauer, K., Mai, M., & Raisch, F. (2022). Driver behavior model for the safety assessment of automated driving. *IEEE Intelligent Vehicles Symposium, Proceedings*, *2022-June*(Iv), 1669–1674. https://doi.org/10.1109/IV51971.2022.9827404

Funke, J., Srinivasan, G., Ranganathan, R., & Burgett, A. (2011). Safety impact methodology (SIM): application and results of the Advanced Crash Avoidance Technologies (ACAT) Program. *Proceedings of the 22nd International Technical Conference on the Enhanced Safety of Vehicles*, 11–0367.

Gabler, H. C., Hampton, C. E., & Hinch, J. (2004). Crash severity: A comparison of event data recorder measurements with accident reconstruction estimates. *SAE Technical Papers*. https://doi.org/10.4271/2004-01-1194

Golzar, J., & Noor, S. (2022). Simple Random Sampling. *IJELS*, *1*(2), 78–82.

Griffin, W., Haworth, N., & Twisk, D. (2020). Patterns in perceived crash risk among male and female drivers with and without substantial cycling experience. *Transportation Research Part F: Traffic Psychology and Behaviour*, *69*. https://doi.org/10.1016/j.trf.2019.12.013

Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, *37*(3). https://doi.org/10.1002/rob.21918

Guo, F., Klauer, S. G., Hankey, J. M., & Dingus, T. A. (2010). Near crashes as crash surrogate for naturalistic Driving Studies. *Transportation Research Record: Journal of the Transportation Research Board*, *2147*(1). https://doi.org/10.3141/2147-09

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

Haus, S. H., Sherony, R., & Gabler, H. C. (2019). Estimated benefit of automated emergency braking systems for vehicle–pedestrian crashes in the United States. *Traffic Injury Prevention*, *20*(sup1), S171–S176. https://doi.org/10.1080/15389588.2019.1602729

Hautzinger, H., Pfeiffer, M., & Schmidt, J. (2004). Expansion of GIDAS Sample Data to the Regional Level: Statistical Methodology and Practical Experiences. *1st International Conference on ESAR "Expert Symposium on Accident Research."*

He, X. (2024). *Driver's perceived risk in relation to automated vehicle behaviour: Evaluation and mitigation of perceived risk through simulator studies, computational models, and user interface design* [Delft University of Technology].

https://research.tudelft.nl/en/publications/drivers-perceived-risk-in-relation-to-automated-vehicle-behaviour

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

Horowitz, M. (2014). 1.1 Computing's energy problem (and what we can do about it). *Digest of Technical Papers - IEEE International Solid-State Circuits Conference*, *57*. https://doi.org/10.1109/ISSCC.2014.6757323

Hugemann, W., & Nickel, N. (2003). Longitudinal and lateral accelerations in normal day driving. *6th International Conference of The Institute of Traffic Accident Investigators*.

Hughes, B. P., Newstead, S., Anund, A., Shu, C. C., & Falkmer, T. (2015). A review of models relevant to road safety. *Accident Analysis and Prevention*, *74*, 250–270. https://doi.org/10.1016/j.aap.2014.06.003

Ingle, S., & Phute, M. (2016). Tesla Autopilot : Semi Autonomous Driving, an Uptick for Future Autonomy. *International Research Journal of Engineering and Technology*, *3*(9), 369–372.

Isaksson-Hellman, I., & Norin, H. (2005). How thirty years of focused safety development has influenced injury outcome in Volvo cars. *Annual Proceedings - Association for the Advancement of Automotive Medicine*, *49*, 63–77. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3217440/

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

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

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

Jeong, E., & Oh, C. (2017). Evaluating the effectiveness of active vehicle safety systems. *Accident Analysis and Prevention*, *100*. https://doi.org/10.1016/j.aap.2017.01.015

Jeppsson, H., & Lubbe, N. (2020). Simulating automated emergency braking with and without torricelli vacuum emergency braking for cyclists: effect of brake deceleration and sensor field-of-view on accidents, injuries and fatalities. *Accident Analysis and Prevention*, *142*. https://doi.org/10.1016/j.aap.2020.105538

Jermakian, J. S. (2011). Crash avoidance potential of four passenger vehicle technologies. *Accident Analysis and Prevention*, *43*(3), 732–740. https://doi.org/10.1016/j.aap.2010.10.020

Jesenski, S., Tiemann, N., Stellet, J. E., & Zöllner, J. M. (2020). Scalable Generation of Statistical Evidence for the Safety of Automated Vehicles by the Use of Importance Sampling. *2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020*. https://doi.org/10.1109/ITSC45102.2020.9294503

Jing, L., Tian, K., & Huang, J. Z. (2015). Stratified feature sampling method for ensemble clustering of high dimensional data. *Pattern Recognition*, *48*(11). https://doi.org/10.1016/j.patcog.2015.05.006

Junaid, S., Bo, S., & Nils, L. (2024). Creation of Pre-Crash Time-Series Data using Chinese In-Depth Crash Database from SHUFO. *IRCOBI Conference*.

Kauffmann, N., Fahrenkrog, F., Drees, L., & Raisch, F. (2022). Positive risk balance: a comprehensive framework to ensure vehicle safety. *Ethics and Information Technology*, *24*(1). https://doi.org/10.1007/s10676-022-09625-2

Kerrigan, D., Hullman, J., & Bertini, E. (2021). A survey of domain knowledge elicitation in applied machine learning. *Multimodal Technologies and Interaction*, *5*(12). https://doi.org/10.3390/mti5120073

Khastgir, S., Dhadyalla, G., Birrell, S., Redmond, S., Addinall, R., & Jennings, P. (2017). Test Scenario Generation for Driving Simulators Using Constrained Randomization Technique. *SAE Technical Papers*, *2017-March*(March). https://doi.org/10.4271/2017-01-1672

Kibalama, D., Tulpule, P., & Chen, B. S. (2022). AV/ADAS Safety-Critical Testing Scenario Generation from Vehicle Crash Data. *SAE Technical Papers*, *2022*. https://doi.org/10.4271/2022-01-0104

Kiefer, R. J., Cassar, M. T., Flannagan, C. A., LeBlanc, D. J., Palmer, M. D., Deering, R. K., & Shulman, M. A. (2003). *Forward collision warning requirements project: refining the camp crash alert timing approach by examining "last-second" braking and lane change maneuvers under various kinematic conditions.* (No. DOT HS 809

574) United States. Department of Transportation. National Highway Traffic Safety Administration.

Kim, J., & Mahmassani, H. S. (2011). Correlated Parameters in Driving Behavior Models. *Transportation Research Record: Journal of the Transportation Research Board*, *2249*(1). https://doi.org/10.3141/2249-09

Kim, S., He, X., van Egmond, R., & Happee, R. (2024). Designing user interfaces for partially automated Vehicles: Effects of information and modality on trust and acceptance. *Transportation Research Part F: Traffic Psychology and Behaviour*, *103*, 404–419. https://doi.org/10.1016/j.trf.2024.02.009

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The Impact of Driver Inattention On Near Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data (No. DOT HS 810 594)*. https://doi.org/10.21949/1530253

Knipling, R. R., Mironer, M., Hendricks, D. L., Tijeripa, L., Everson, J., Allen, J. C., & Wilson, C. (1993). *Assessment of IVHS countermeasures for collision avoidance: rear-end crashes* (Issue May). (No. DOT-HS-807-995). United States. Joint Program Office for Intelligent Transportation Systems.

Kolekar, S., de Winter, J., & Abbink, D. (2020a). Human-like driving behaviour emerges from a risk-based driver model. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-18353-4

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

Kovaceva, J., Bärgman, J., & Dozza, M. (2022). On the importance of driver models for the development and assessment of active safety: A new collision warning system to make overtaking cyclists safer. *Accident Analysis and Prevention*, *165*(November). https://doi.org/10.1016/j.aap.2021.106513

Kruber, F., Wurst, J., Botsch, M., & Chakraborty, S. (2023). Unsupervised Random Forest Learning for Traffic Scenario Categorization. In *Machine Learning and Optimization Techniques for Automotive Cyber-Physical Systems*. https://doi.org/10.1007/978-3-031-28016-0_20

Kruschke, J. K. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan, second edition. In *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, Second Edition*. https://doi.org/10.1016/B978-0-12-405888-0.09999-2

Kumar Akinapalli, P., Pawar, D. S., & Dia, H. (2023). Classification of motorized two-wheeler riders' acceleration and deceleration behavior through short-term naturalistic riding study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *96*. https://doi.org/10.1016/j.trf.2023.06.008

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

Kusano, K. D., & Gabler, H. C. (2014). Comprehensive target populations for current active safety systems using national crash databases. *Traffic Injury Prevention*, *15*(7), 753–761. https://doi.org/10.1080/15389588.2013.871003

Kusano, K. D., Scanlon, J. M., Chen, Y.-H., McMurry, T. L., Chen, R., Gode, T., & Victor, T. (2024). Comparison of Waymo Rider-Only Crash Data to Human Benchmarks at 7.1 Million Miles. *Traffic Injury Prevention*, *25*(1), S66–S77. https://doi.org/10.1080/15389588.2024.2380786

Kyriakidis, M., Happee, R., & Winter, J. C. F. De. (2015). Public opinion on automated driving : Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Psychology and Behaviour*, *32*, 127–140. https://doi.org/10.1016/j.trf.2015.04.014

Lee, J. D., & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1). https://doi.org/10.1518/hfes.46.1.50.30392

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 and Prevention*, *110*(April), 29–37. https://doi.org/10.1016/j.aap.2017.10.009

Lemmen, P., Fagerlind, H., Unselt, T., Rodarius, C., Infantes, E., & van der Zweep, C. (2012). Assessment of integrated vehicle safety systems for improved vehicle safety. *Procedia - Social and Behavioral Sciences*, *48*. https://doi.org/10.1016/j.sbspro.2012.06.1138

Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7). https://doi.org/10.1111/oik.05985

Li, Y., Liu, F., Xing, L., He, Y., Dong, C., Yuan, C., Chen, J., & Tong, L. (2023). Data generation for connected and automated vehicle tests using deep learning models. *Accident Analysis and Prevention*, *190*. https://doi.org/10.1016/j.aap.2023.107192

Li, Z., Gong, J., Zhang, Z., Lu, C., Knoop, V. L., & Wang, M. (2024). Interactive Behavior Modeling for Vulnerable Road Users with Risk-Taking Styles in Urban Scenarios: A Heterogeneous Graph Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*, *25*(8), 8538–8555. https://doi.org/10.1109/TITS.2024.3399481

Likmeta, A., Metelli, A. M., Tirinzoni, A., Giol, R., Restelli, M., & Romano, D. (2020). Combining reinforcement learning with rule-based controllers for transparent and general decision-making in autonomous driving. *Robotics and Autonomous Systems*, *131*. https://doi.org/10.1016/j.robot.2020.103568

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

Lubbe, N. (2017). Brake reactions of distracted drivers to pedestrian Forward Collision Warning systems. *Journal of Safety Research*, *61*. https://doi.org/10.1016/j.jsr.2017.02.002

Lubbe, N., & Davidsson, J. (2015). Drivers' comfort boundaries in pedestrian crossings: A study in driver braking characteristics as a function of pedestrian walking speed. *Safety Science*, *75*. https://doi.org/10.1016/j.ssci.2015.01.019

Ma, Y., Sun, C., Chen, J., Cao, D., & Xiong, L. (2022). Verification and Validation Methods for Decision-Making and Planning of Automated Vehicles: A Review. In *IEEE Transactions on Intelligent Vehicles* (Vol. 7, Issue 3). https://doi.org/10.1109/TIV.2022.3196396

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

Matt, M. (2022). *Confused drivers think they have 'self-driving' cars. That's dangerous, an insurance group warns.* Retrieved on February 10th, 2023 from: https://edition.cnn.com/2022/10/11/business/iihs-autopilot-supercruise-propilot/index.html

Mattas, K., Albano, G., Donà, 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 and Prevention*, *174*(February). https://doi.org/10.1016/j.aap.2022.106743

Menzel, T., Bagschik, G., & Maurer, A. M. (2018). Scenarios for development, test and validation of automated vehicles. *IEEE Intelligent Vehicles Symposium, Proceedings*, *2018-June*. https://doi.org/10.1109/IVS.2018.8500406

Moerbeek, M. (2021). Bayesian updating: increasing sample size during the course of a study. *BMC Medical Research Methodology*, *21*(1). https://doi.org/10.1186/s12874-021-01334-6

Mohd Ishanuddin, N., Rusli, N. E., Sukadarin, E. H., Abdul Aziz, H., Widia, M., Fauzan, N. S., Zakaria, J., Osman, H., Ahmad, A. A., Mohd Jawi, Z., Mohd Nawi, N. S., Roslin, E. N., & Yassierli. (2022). Preliminary study on drivers satisfaction and continuance intention to use automatic emergency braking in malaysia. *Lecture Notes in Mechanical Engineering*. https://doi.org/10.1007/978-981-16-4115-2_27

Mohd Ishanuddin, N., Sukadarin, E. H., Mohd Nawi, N. S., Widia, M., Ab Rashid, A. A., Abdul Aziz, H., Zakaria, J., Fauzan, N. S., Osman, H., Roslin, E. N., Jawi, Z. M., & Yassierli. (2021). Consumers' perception of automatic emergency braking (AEB): theoretical model and construct development. *Journal of the Society of Automotive Engineers Malaysia*, *5*(2). https://doi.org/10.56381/jsaem.v5i2.165

Molnar, L. J., Ryan, L. H., Pradhan, A. K., Eby, D. W., St. Louis, R. M., & Zakrajsek, J. S. (2018). Understanding trust and acceptance of automated vehicles: An exploratory simulator study of transfer of control between automated and manual driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*. https://doi.org/10.1016/j.trf.2018.06.004

Montanari, F., Ren, H., & Djanatliev, A. (2021). Scenario Detection in Unlabeled Real Driving Data with a Rule-Based State Machine Supported by a Recurrent Neural Network. *IEEE Vehicular Technology Conference*, *2021-April*. https://doi.org/10.1109/VTC2021-Spring51267.2021.9449032

Moon, S., & Yi, K. (2008). Human driving data-based design of a vehicle adaptive cruise control algorithm. *Vehicle System Dynamics*, *46*(8). https://doi.org/10.1080/00423110701576130

Morales-Alvarez, W., Sipele, O., Léberon, R., Tadjine, H. H., & Olaverri-Monreal, C. (2020). Automated driving: A literature review of the take over request in conditional automation. In *Electronics (Switzerland)* (Vol. 9, Issue 12). https://doi.org/10.3390/electronics9122087

Nalic, D., Mihalj, T., Baumler, M., Lehmann, M., Eichberger, A., & Bernsteiner, S. (2020). Scenario based testing of automated driving systems: a literature survey. *Paper Presented at FISITA Web Congress 2020, Virtuell, Czech Republic.*

National Bureau of Statistics of China. (2023). *China statistical yearbook : basic statistics on traffic accidents*. https://www.stats.gov.cn/sj/ndsj/2023/indexeh.htm

NHTSA. (2015). *Critical reasons for crashes investigated in the national motor vehicle crash causation survey (DOT HS 812 115)* (Issue August). https://www.govinfo.gov/app/details/GOVPUB-TD8-PURL-gpo172313

O'Kelly, M., Duchi, J., Sinha, A., Namkoong, H., & Tedrake, R. (2018). Scalable end-to-end autonomous vehicle testing via rare-event simulation. *Advances in Neural Information Processing Systems*, *2018-December*.

Olleja, P. (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*.

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

Östh, J., Pipkorn, B., Forsberg, J., & Iraeus, J. (2022). Numerical Reproducibility of Human Body Model Crash Simulations. *Proceedings of Science*, *395*.

Otte, D., Jänsch, M., & Haasper, C. (2012). Injury protection and accident causation parameters for vulnerable road users based on German In-Depth Accident Study GIDAS. *Accident Analysis and Prevention*, *44*(1). https://doi.org/10.1016/j.aap.2010.12.006

Owen, A., & Zhou, Y. (2000). Safe and Effective Importance Sampling. *Journal of the American Statistical Association*, *95*(449). https://doi.org/10.1080/01621459.2000.10473909

Park, J., Byon, E., Ko, Y. M., & Shashaani, S. (2024). Strata Design for Variance Reduction in Stochastic Simulation. *Technometrics*. https://doi.org/10.1080/00401706.2024.2416411

Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB). https://doi.org/10.1016/j.trf.2014.04.009

Petridou, E., & Moustaki, M. (2000). Human factors in the causation of road traffic crashes. *European Journal of Epidemiology*, *16*(9), 819–826. https://doi.org/10.1023/A:1007649804201

Pipkorn, B., Östh, J., Brynskog, E., Larsson, E., Rydqvist, L., Iraeus, J., Perez-Rapela, D., & Jakobsson, L. (2022). Validation of the SAFER Human Body Model Kinematics in Far-Side Impacts. *Proceedings of Science*, *395*.

Pisharam, P. P., Lubbe, N., & Davidsson, J. (2022). Estimated lives saved by recently implemented vehicle safety standards in india: implications and future safety needs. *Proceedings of Science*, *395*.

Ponn, T., Fratzke, D., Gnandt, C., & Lienkamp, M. (2019). Towards certification of autonomous driving: Systematic test case generation for a comprehensive but economically-feasible assessment of lane keeping assist algorithms. *VEHITS 2019 - Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems*. https://doi.org/10.5220/0007678603330342

Prati, G., Pietrantoni, L., & Fraboni, F. (2017). Using data mining techniques to predict the severity of bicycle crashes. *Accident Analysis and Prevention*, *101*. https://doi.org/10.1016/j.aap.2017.01.008

Puente Guillen, P., & Gohl, I. (2019). Forward collision warning based on a driver model to increase drivers' acceptance. *Traffic Injury Prevention*, *20*(sup1). https://doi.org/10.1080/15389588.2019.1623397

Reagan, I. J., Cicchino, J. B., Kerfoot, L. B., & Weast, R. A. (2018). Crash avoidance and driver assistance technologies – Are they used? *Transportation Research Part F: Traffic Psychology and Behaviour*, *52*, 176–190. https://doi.org/10.1016/j.trf.2017.11.015

Riedmaier, S., Ponn, T., Ludwig, D., Schick, B., & Diermeyer, F. (2020). Survey on scenario-based safety assessment of automated vehicles. *IEEE Access*, *8*, 87456–87477. https://doi.org/10.1109/ACCESS.2020.2993730

Rödel, C., Stadler, S., Meschtscherjakov, A., & Tscheligi, M. (2014). Towards autonomous cars: The effect of autonomy levels on Acceptance and User Experience. *AutomotiveUI 2014 - 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, in Cooperation with ACM SIGCHI - Proceedings*. https://doi.org/10.1145/2667317.2667330

Roesener, C., Sauerbier, J., Zlocki, A., Fahrenkrog, F., Wang, L., Várhelyi, A., Breunig, S., Mejuto, P., -Galician, C. T. A. G., Tango, F., & Lanati, J. (2017). A Comprehensive Evaluation Approach for Highly Automated Driving. *25th Enhanced Safety of Vehicles 2017*.

Rosén, E. (2013). Autonomous emergency braking for vulnerable road users. *2013 IRCOBI Conference Proceedings - International Research Council on the Biomechanics of Injury (IRCOBI)*, 618–627.

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin and Review*, *21*(2). https://doi.org/10.3758/s13423-014-0595-4

Ruscio, D., Ciceri, M. R., & Biassoni, F. (2015). How does a collision warning system shape driver's brake response time? the influence of expectancy and automation complacency on real-life emergency braking. *Accident Analysis and Prevention*, *77*. https://doi.org/10.1016/j.aap.2015.01.018

Ryan, T. P. (2013). Sample Size Determination and Power. In *Sample Size Determination and Power*. https://doi.org/10.1002/9781118439241

Sadia, F., & Hossain, S. S. (2014). Contrast of bayesian and classical sample size determination. *Journal of Modern Applied Statistical Methods*, *13*(2), 420–431. https://doi.org/10.22237/jmasm/1414815720

SAE. (2021). *SAE J3016: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems*. SAE J3016. https://www.sae.org/standards/content/j3016_202104/

Sander, U. (2018). *Predicting safety benefits of automated emergency braking at intersections virtual simulations based on real-world accident data* [Doctoral thesis, Doctoral dissertation, Chalmers University of Technology]. https://research.chalmers.se/en/publication/504728

Sander, U., Ek, P. , Sander, D. , Skudlarek, H. , Breunig, S. , Bärgman, J. , Menzel, T. , Glasmacher, C. , Urban, M. , Erbsmehl, C. T. , Chajmowicz, H. , Davidse, R. , Nabavi Niaki, M. , Schermers, G. , Lorente Mallada, J. , Op den Camp, O. , Charoniti, E. , Klunder, G. , Meocci, M. , … Hay, J. (2024). *Data Sources for Baseline Generation – Overview, Grading, and Recommendations. Deliverable D4.1 of the Horizon Europe project V4SAFETY.* https://v4safetyproject.eu/media/pages/outputs/45637bc1ba-1714029598/d4.1-data-sources-for-baseline-generation-v4safety.pdf

Sander, U., & Lubbe, N. (2018). Market penetration of intersection AEB: Characterizing avoided and residual straight crossing path accidents. *Accident Analysis and Prevention*, *115*(February), 178–188. https://doi.org/10.1016/j.aap.2018.03.025

Sander, U., Lubbe, N., & Pietzsch, S. (2019). Intersection AEB implementation strategies for left turn across path crashes. *Traffic Injury Prevention*, *20*(sup1), S119–S125. https://doi.org/10.1080/15389588.2019.1602728

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 and Prevention*, *163*. https://doi.org/10.1016/j.aap.2021.106454

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

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin and Review*, *25*(1). https://doi.org/10.3758/s13423-017-1230-y

Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. In *Journal of the Indian Society of Agricultural Statistics* (Vol. 5, Issue 1194).

Sharma, D., Stern, S., Brophy, J., & Choi, E.-H. (2007). An Overview of NHTSA's Crash Reconstruction Software WinSMASH. *20th International Technical Conference on the Enhanced Safety of Vehicles Conference*.

Son, J., Yoo, H., Kim, S., & Sohn, K. (2015). Real-time illumination invariant lane detection for lane departure warning system. *Expert Systems with Applications*, *42*(4). https://doi.org/10.1016/j.eswa.2014.10.024

Song, Y., Chitturi, M. V, & Noyce, D. A. (2022). Intersection two-vehicle crash scenario specification for automated vehicle safety evaluation using sequence analysis and Bayesian networks. *Accident Analysis & Prevention*, *176*, 106814. https://doi.org/10.1016/j.aap.2022.106814

Stark, C., Medrano-Berumen, C., & Akbas, M. I. (2020). Generation of autonomous vehicle validation scenarios using crash data. *Conference Proceedings - IEEE SOUTHEASTCON*, *2020-March*. https://doi.org/10.1109/SoutheastCon44009.2020.9249662

Stark, L., Düring, M., Schoenawa, S., Maschke, J. E., & Do, C. M. (2019). Quantifying vision zero: crash avoidance in rural and motorway accident scenarios by combination of ACC, AEB, and LKS projected to German accident occurrence. *Traffic Injury Prevention*, *20*(sup1). https://doi.org/10.1080/15389588.2019.1605167

Stefan, A. M., Schönbrodt, F. D., Evans, N. J., & Wagenmakers, E. J. (2022). Efficiency in sequential testing: Comparing the sequential probability ratio test and the sequential Bayes factor test. *Behavior Research Methods*, *54*(6). https://doi.org/10.3758/s13428-021-01754-8

Stutts, J. C., Reinfurt, D. W., Staplin, L., & Rodgman, E. A. (2001). The role of driver distraction in Traffic crashes. *AAA Foundation for Traffic Safety*, *May*.

Tesla. (2022). *Full self-friving hardware on all cars*. https://www.tesla.com/autopilot

Thacker, B. H., Doebling, S. W., Hemez, F. M., Anderson, M. C., Pepin, J. E., & Rodriguez, E. a. (2004). Concepts of Model Verification and Validation. *Concepts of Model Verification and Validation*.

Tokdar, S. T., & Kass, R. E. (2010). Importance sampling: A review. In *Wiley Interdisciplinary Reviews: Computational Statistics* (Vol. 2, Issue 1). https://doi.org/10.1002/wics.56

Tran, D., Do, H. M., Sheng, W., Bai, H., & Chowdhary, G. (2018). Real-time detection of distracted driving based on deep learning. *IET Intelligent Transport Systems*, *12*(10). https://doi.org/10.1049/iet-its.2018.5172

Tsoi, K. K., Mulder, M., & Abbink, D. A. (2010). Balancing safety and support: Changing lanes with a haptic lane-keeping support system. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*. https://doi.org/10.1109/ICSMC.2010.5642414

Ulbrich, S., Menzel, T., Reschka, A., Schuldt, F., & Maurer, M. (2015). Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, *2015-October*. https://doi.org/10.1109/ITSC.2015.164

UNECE. (2021). *UN Regulation No. 157. Uniform provisions concerning the approval of vehicles with regard to Automated Lane Keeping Systems.* https://unece.org/sites/default/files/2021-03/R157e.pdf

Viano, D. C., & Parenteau, C. S. (2010). Injury risks in frontal crashes by delta V and body region with focus on head injuries in low-speed collisions. *Traffic Injury Prevention*, *11*(4). https://doi.org/10.1080/15389581003751623

Viano, D. C., & Ridella, S. (1996). Crash causation: A case study of fatal accident circumstances and configurations. *SAE Technical Papers*. https://doi.org/10.4271/960458

Victor, T., Dozza, M., Bärgman, J., Boda, C. N., Engström, J., Flannagan, C., Lee, J. D., & Markkula, G. (2015). Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk. In *SHRP 2 Report S2-S08A-RW-1*. https://doi.org/doi:10.17226/22297

Wang, C., & Stamatiadis, N. (2014). Evaluation of a simulation-based surrogate safety metric. *Accident Analysis and Prevention*, *71*. https://doi.org/10.1016/j.aap.2014.05.004

Wang, C., Xie, Y., Huang, H., & Liu, P. (2021). A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. *Accident Analysis and Prevention*, *157*. https://doi.org/10.1016/j.aap.2021.106157

Wang, M., Liao, Y., Lyckvi, S. L., & Chen, F. (2020). How drivers respond to visual vs. auditory information in advisory traffic information systems. *Behaviour and Information Technology*, *39*(12). https://doi.org/10.1080/0144929X.2019.1667439

Wang, M. S., Jeong, N. T., Kim, K. S., Choi, S. B., Yang, S. M., You, S. H., Lee, J. H., & Suh, M. W. (2016). Drowsy behavior detection based on driving information.

*International Journal of Automotive Technology*, *17*(1). https://doi.org/10.1007/s12239-016-0016-y

Wang, X., Liu, Q., Guo, F., Fang, S., Xu, X., & Chen, X. (2022). Causation analysis of crashes and near crashes using naturalistic driving data. *Accident Analysis and Prevention*, *177*(August), 106821. https://doi.org/10.1016/j.aap.2022.106821

Wang, X., Peng, H., & Zhao, D. (2021). Combining reachability analysis and importance sampling for accelerated evaluation of highway automated vehicles at pedestrian crossing. *ASME Letters in Dynamic Systems and Control*, *1*(1). https://doi.org/10.1115/1.4046610

Waymo. (2022). *The world's first autonomous ride-hailing service*. Retrieved on January 12th, 2023 from. https://waymo.com/

Wege, C., Will, S., & Victor, T. (2013). Eye movement and brake reactions to real world brake-capacity forward collision warnings - A naturalistic driving study. *Accident Analysis and Prevention*, *58*. https://doi.org/10.1016/j.aap.2012.09.013

Wei, C., Romano, R., Merat, N., Wang, Y., Hu, C., Taghavifar, H., Hajiseyedjavadi, F., & Boer, E. R. (2019). Risk-based autonomous vehicle motion control with considering human driver's behaviour. *Transportation Research Part C: Emerging Technologies*, *107*, 1–14. https://doi.org/10.1016/j.trc.2019.08.003

Wei, T., Kang, K., Zhu, T., & Liu, H. (2022). The effect of driver's response features on safety effectiveness of autonomous emergency braking. *SAE Technical Paper*, *01*(7131). https://doi.org/https://doi.org/10.4271/2022-01-7131

Winner, H., Lemmer, K., Form, T., & Mazzega, J. (2019). PEGASUS—first steps for the safe introduction of automated driving. In *Lecture Notes in Mobility*. https://doi.org/10.1007/978-3-319-94896-6_16

World Health Organization. (2017). *Powered two- and three-wheeler safety: a road safety manual for decision-makers and practitioners*.

World Health Organization. (2019). *Global status report on road safety: The Southeast Asia story*. https://www.unescap.org/sites/default/files/Road Safety Data Management - WHO Global Status Report on Road Safety - WHO_0.pdf

World Health Organization. (2023). *Global status report on road safety*. https://www.who.int/publications/i/item/9789240086517

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* [Thesis For The Degree of Licentiate Of Engineering. Chalmers University of Technology]. https://research.chalmers.se/publication/542271/file/542271_Fulltext.pdf

Wulfe, B., Hartong-Redden, R., Chintakindi, S., Kodali, A., Choi, S. C. T., & Kochenderfer, M. J. (2018). Real-time prediction of intermediate-horizon automotive collision risk. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, *2*.

Xia, Q., Duan, J., Gao, F., Chen, T., & Yang, C. (2017). Automatic Generation Method of Test Scenario for ADAS Based on Complexity. *SAE Technical Papers*, *Part F129886*(September). https://doi.org/10.4271/2017-01-1992

Xu, Y., Zou, Y., & Sun, J. (2018). Accelerated testing for automated vehicles safety evaluation in cut-in scenarios based on importance sampling, genetic algorithm and simulation applications. *Journal of Intelligent and Connected Vehicles*, *1*(1). https://doi.org/10.1108/JICV-01-2018-0002

Yates, F., & Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society: Series B (Methodological)*, *15*(2). https://doi.org/10.1111/j.2517-6161.1953.tb00140.x

Zhang, G., Yau, K. K. W., Zhang, X., & Li, Y. (2016). Traffic accidents involving fatigue driving and their extent of casualties. *Accident Analysis and Prevention*, *87*. https://doi.org/10.1016/j.aap.2015.10.033

Zhang, J., Shu, Y., & Yu, H. (2021). Human-Machine Interaction for Autonomous Vehicles: A Review. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12774 LNCS*. https://doi.org/10.1007/978-3-030-77626-8_13

Zhang, T., Tao, D., Qu, X., Zhang, X., Zeng, J., Zhu, H., & Zhu, H. (2020). Automated vehicle acceptance in China: Social influence and initial trust are key determinants. *Transportation Research Part C: Emerging Technologies*, *112*. https://doi.org/10.1016/j.trc.2020.01.027

Zhang, X., Jin, X., Qi, W., & Guo, Y. (2008). Vehicle crash accident reconstruction based on the analysis 3D deformation of the auto-body. *Advances in Engineering Software*, *39*(6). https://doi.org/10.1016/j.advengsoft.2007.05.002

Zhao, D., Lam, H., Peng, H., Bao, S., LeBlanc, D. J., Nobukawa, K., & Pan, C. S. (2017). Accelerated Evaluation of Automated Vehicles Safety in Lane-Change Scenarios Based on Importance Sampling Techniques. *IEEE Transactions on Intelligent Transportation Systems*, *18*(3). https://doi.org/10.1109/TITS.2016.2582208

Zhao, Di., Huang, X., Peng, H., Lam, H., & Leblanc, D. J. (2018). Accelerated Evaluation of Automated Vehicles in Car-Following Maneuvers. *IEEE Transactions on Intelligent Transportation Systems*, *19*(3). https://doi.org/10.1109/TITS.2017.2701846

Zhao, J., Zhao, W., Deng, B., Wang, Z., Zhang, F., Zheng, W., Cao, W., Nan, J., Lian, Y., & Burke, A. F. (2024). Autonomous driving system: A comprehensive survey. In *Expert Systems with Applications* (Vol. 242). https://doi.org/10.1016/j.eswa.2023.122836

Zhao, M., Wang, H., Chen, J., Xu, X., & He, Y. (2017). Method to optimize key parameters and effectiveness evaluation of the AEB system based on rear-end collision accidents. *SAE International Journal of Passenger Cars - Electronic and Electrical Systems*, *10*(2). https://doi.org/10.4271/2017-01-0112