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

Citation for the original published paper (version of record):
Olleja, P., Bärgman, J., Lübbe, N. (2022). Can non-crash naturalistic driving data be an alternative to crash data for use in virtual
assessment of the safety performance of automated emergency braking systems?. Journal of Safety Research, 83: 139-151. http://dx.doi.org/10.1016/j.jsr.2022.08.011

(article starts on next page)

Journal of Safety Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

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

Pierluigi Olleja[a,*], Jonas Bärgman[a], Nils Lubbe[b]

[a] Division of Vehicle Safety at the Department of Mechanics and Maritime Sciences, Chalmers University of Technology, 412 96 Göteborg, Sweden
[b] Autoliv Research, Wallentinsvägen 22, 447 83 Vårgårda, Sweden

ARTICLE INFO

Article history:
Received 28 July 2021
Received in revised form 1 April 2022
Accepted 17 August 2022
Available online xxxx

Keywords:
GIDAS
highD
SHRP2
AEB
Crash avoidance

ABSTRACT

Introduction: Developers of in-vehicle safety systems need to have data allowing them to identify traffic safety issues and to estimate the benefit of the systems in the region where it is to be used, before they are deployed on-road. Developers typically want in-depth crash data. However, such data are often not available. There is a need to identify and validate complementary data sources that can complement in-depth crash data, such as Naturalistic Driving Data (NDD). However, few crashes are found in such data. This paper investigates how rear-end crashes that are artificially generated from two different sources of non-crash NDD (highD and SHRP2) compare to rear-end in-depth crash data (GIDAS). *Method:* Crash characteristics and the performance of two conceptual automated emergency braking (AEB) systems were obtained through virtual simulations – simulating the time-series crash data from each data source. *Results:* Results show substantial differences in the estimated impact speeds between the artificially generated crashes based on both sources of NDD, and the in-depth crash data; both with and without AEB systems. Scenario types also differed substantially, where the NDD have many fewer scenarios where the following-vehicle is not following the lead vehicle, but instead catches-up at high speed. However, crashes based on NDD near-crashes show similar pre-crash criticality (time-to-collision) to in-depth crash data. *Conclusions:* If crashes based on near-crashes are to be used in the design and assessment of preventive safety systems, it has to be done with great care, and crashes created purely from small amounts of everyday driving NDD are not of much use in such assessment. *Practical applications:* Researchers and developers of in-vehicle safety systems can use the results from this study: (a) when deciding which data to use for virtual safety assessment of such systems, and (b) to understand the limitations of NDD.

1. Introduction

Traffic crashes are the eighth leading cause of death worldwide; every year 1.35 million people lose their lives in traffic crashes (WHO, 2018). Fatal rear-end crashes accounted for 7.2% of all fatal crashes in the United States (U.S.) in 2017 (NHTSA, 2019).

Safety measures have for many years been developed to address traffic safety. Available safety measures include preventive safety systems, aimed at avoiding or mitigating the consequences of a possible crash before impact, and protective safety systems, aimed at protecting the occupants from the consequences of a crash during impact. One example of an effective protective safety system for rear-end crashes is whiplash protection with energy-controlling structures and optimized headrest designs (Kullgren, Krafft, Lie, & Tingvall, 2007; Kullgren, Stigson, & Krafft, 2013). For preventive safety, Automated Emergency Braking (AEB) has been shown to be an effective preventive safety system, reducing the number and severity of rear-end crashes substantially (Cicchino, 2017; Fildes et al., 2015).

There are different types of AEB algorithms. Early AEB systems only included time-to-collision (TTC; based on the vehicles' relative distance, speeds and accelerations) and the braking response by the driver of the following vehicle (FV) (Brännström, Sjöberg, & Coelingh, 2008), while more mature systems may also consider the FV driver's ability to steer away comfortably (in addition or as an alternative to braking; see Brännström, Coelingh, & Sjöberg, 2010, 2014; Sander, 2018). This consideration reduces false positives, which are activations when either there is no real need for

* Corresponding author.
E-mail addresses: ollejap@chalmers.se (P. Olleja), jonas.bargman@chalmers.se (J. Bärgman), nils.lubbe@autoliv.com (N. Lubbe).

https://doi.org/10.1016/j.jsr.2022.08.011
0022-4375/© 2022 The Authors. Published by Elsevier Ltd.

it or the driver does not feel it is warranted (Bliss & Acton, 2003; Coelingh, Jakobsson, Lind, & Lindman, 2007; Sander & Lubbe, 2016). In the latter case, drivers may still avoid the crash while remaining inside their comfort zone. We have chosen to quantify the comfort zone boundary by selecting an acceptable lateral acceleration for the steering-away maneuvers. The term "comfortable steering" will be used to describe maneuvers that do not take drivers out of their comfort zone, even when performed when braking is no longer an option to avoid a crash. An AEB considering comfort zone boundaries delays the intervention until the driver cannot avoid the crash by steering in a comfortable way (e.g., when the lateral acceleration crosses the driver's comfort zone boundary; see Bärgman, Smith, and Werneke, 2015; Summala, 2007).

To quantify the benefit of safety systems, such as AEB, developers need both assessment methods and data. They need to assess to what extent a specific concept (or even the specific application of a system) will affect safety all through the systems' life cycle (Alvarez et al., 2017). This type of assessment is prospective, predicting the potential safety benefit of a system before data are available from real-world crashes. There are different methods available for the prospective assessment of AEB systems. One increasingly popular method is virtual simulations in computers. This allows for early assessment of a system's potential, and enables fast and iterative improvement of its safety effectiveness. Virtual traffic simulation is one such approach. The movement of traffic participants is modeled with respect to vehicle dynamics, and driver behavior and control, or alternatively, with added automated control (Fahrenkrog et al., 2019; Helmer, 2014). When traffic characteristics, driver, and automation behavior are modeled accurately, one can expect to create an accurate representation of crashes as the outcome of interest; however, such detailed modeling is highly ambitious (Dobberstein et al., 2021). The benefit of automation can be determined by comparing a simulation with only human drivers to simulations with automation, either by resimulating selected critical events (Fahrenkrog et al., 2019; Hallerbach, 2020) or only the crash events generated in the human-driver-only simulations (Tanaka, 2015). Traffic simulation for safety benefit assessment holds the promise of creating an essentially unlimited amount of parameter variations and crash events. However, as driver behavior is complex and modeling them accurately (enough) is difficult (Markkula, 2015), such models "only represent the behavior of real drivers to a certain extent" (Bjorvatn et al., 2021, p. 123) and results are sensitive to the variations of such models (ISO, 2021). Note that the traffic simulation approach is often targeting assessments of higher levels of automation.

Counterfactual or "what-if" simulations is another simulation-based assessment method that has been used extensively to assess advanced driver assistance systems (ADAS; Bärgman, Lisovskaja, Victor, Flannagan, & Dozza, 2015; Davis, Hourdos, Xiong, & Chatterjee, 2011; McLaughlin, Hankey, & Dingus, 2008; Scanlon et al., 2021). These simulations typically assess safety by using pre-crash kinematics from real-world data (Kusano & Gabler, 2012; Lindman & Tivesten, 2006; Sander, 2018; Scanlon et al., 2021), simulating each event with and without an algorithm modeling the preventive safety system under assessment (Kusano & Gabler, 2012; Sander, 2018). The results are typically provided in the form of the proportion of crashes that were avoided with the system, and the impact speed (or injury risk) distribution of the crashes that still occurred after the system was applied. In this way it is possible to virtually compare the original, baseline event with the modified ("what-if") events that include the AEB system.

Data on pre-crash kinematics are needed to perform the counterfactual AEB simulations. Different sources of pre-crash kinematics data include in-depth crash reconstruction, event data recorders (EDRs), and naturalistic driving data (NDD), which are collected either in-vehicle or on-site (i.e., monitoring a specific piece of road; see Krajewski et al., 2018).

In-depth reconstructed crash databases include information not only about the crash, but also about the pre-crash phase (Bakker et al., 2017). Experts can reconstruct the pre-crash kinematics and document many other aspects of the crash, such as the road geometry and other environmental factors—as well as the injuries sustained by the humans involved in the crash (Otte, Krettek, Brunner, & Zwipp, 2003). Typically, however, very little information is available about the pre-crash phase (Schubert, Erbsmehl, & Hannawald, 2013). Also, in-depth crash data with reconstructed pre-crash kinematics are not available in all countries or regions for which safety systems (such as AEB) should be evaluated prospectively. One example of an in-depth crash database is the German In-Depth Accident Study (GIDAS), which started collecting crash data in 1999. Approximately 2,000 crashes from the cities of Hannover and Dresden and their surroundings are added every year (Otte et al., 2003; Liers, 2018). GIDAS crashes are all reconstructed with estimates of crash kinematics and impact speed.

For a subset of crashes in the GIDAS crash database a Pre-Crash Matrix (PCM) is created, which includes the pre-crash kinematics of the vehicles involved up to five seconds before the collision. The crashes are reconstructed using a structured approach (Schubert et al., 2013). As of February 2018, the GIDAS PCM database contained 9,729 crashes (VUFO, 2020). Reconstruction of pre-crash kinematics has also been performed for other in-depth databases, such as the Initiative for the Global Harmonization of Accident Data (IGLAD) (Spitzhüttl, Petzold, & Liers, 2015) and the Road Accident Sampling System India (RASSI) (Shaikh & Sander, 2018).

The pre-crash kinematics data from reconstructed crashes can be used directly in counterfactual simulations (Rosén, 2013; Sander, 2018; Scanlon et al., 2021). Typically, the system under assessment is applied to the pre-crash kinematics and, for each timestep, a threat assessment analysis is performed. The simulation framework always includes a vehicle model, and often a model of the driver. The outcomes of the simulations consist of avoided or (hopefully) crashes with reduced impact speed and, thus, mitigated injury risk. That is, counterfactual simulations can also include collision models, so that in case of a crash the occupants' injury risks can be studied (Sander & Lubbe, 2016, 2018).

As an alternative to using data from in-depth crash investigations, real-world pre-crash kinematics for counterfactual simulations can be extracted from event data recorders. These recorders are already mandatory in new vehicles in several countries (NHTSA, 2006; UNECE, 2019), and more countries are following suit (Šajn, 2019). The event data recorders of today typically record, among other things, the vehicle speed in the few seconds leading up to the crash and the acceleration during the crash. However, the pre-crash data are often recorded at a low frequency (1–5 Hz), so it is often not known exactly when (within the 200 ms to 1 s that the sample frequency provides) the impact occurred, which reduces reconstruction quality (Thomson et al., 2013) and, naturally, impacts simulation validity. Nevertheless, event data recorders are a useful information source when reconstructing crashes: similar to in-depth reconstructed crash databases, they represent real-world crashes and have been used extensively as a basis for counterfactual simulations (Bareiss, Scanlon, Sherony, & Gabler, 2019; Kusano & Gabler, 2012; Scanlon, Kusano, & Gabler, 2015; Scanlon, Page, Sherony, & Gabler, 2016).

Lastly, NDD can also be a source of pre-crash kinematics data for counterfactual simulations. NDD are recorded unobtrusively in real-traffic, and two main types of such data exist: site-based NDD and in-vehicle NDD.

Site-based NDD are collected at one or more specific sites, where, typically, cameras, radars or LIDARs collect data about road-user movements over a time duration from minutes to months (Bock et al., 2020; Krajewski, Moers, Bock, Vater, & Eckstein, 2020; Krajewski et al., 2018; Laureshyn, 2010; Smith, Thome, Blåberg, & Bärgman, 2009). The data are post-processed to produce trajectories and other information, such as speed and acceleration, about the road users captured in the recordings. Most of the data from site-based NDD collections capture normal everyday driving without any critical events; they contain very few crashes (Van Nes, Christoph, Hoedemaeker, & Van Der Horst, 2013). The highD dataset is one example of recent site-based NDD which only includes normal everyday driving data (Krajewski et al., 2018). The data were collected by drones recording video of six stretches of highway in North Rhine-Westphalia, Germany in 2017 and 2018. There were 60 recordings, and the drones recorded 17 minutes per recording, on average. A total of 110,000 vehicle trajectories were recorded, with a typical length (longitudinal road segment) of 420 m. The highD data are freely available for research purposes (Krajewski et al., 2018).

In contrast to site-based NDD, in-vehicle NDD are collected from vehicles instrumented with a data acquisition system that collects vehicle information such as speed and acceleration, driver information such as glance behavior, and data about surrounding traffic (typically using radar and cameras). The largest in-vehicle NDD study to date is SHRP2: data were collected on 3,247 drivers who drove a total of almost 80 million km over a period of three years in the United States. Since SHRP2 collected so much data, more than 1,000 crashes of different severities were recorded (SHRP2 crash severity levels 1–3), as were other critical events (e.g., near-crashes: see Blatt et al., 2015; VTTI, 2020).

NDD have been used to study the safety benefit of, for example, forward collision warning (FCW) and AEB (Bärgman, Boda, & Dozza, 2017; Woodrooffe et al., 2012). When NDD are used in counterfactual simulations of rear-end crashes, the evasive maneuver of the FV in each event can be replaced by an evasive maneuver created by a quantitative driver response model (Bärgman et al., 2017). The main reason for this replacement is that each crash or near-crash is just one instance of the behavior of that driver, which just happened to produce that particular crash or near-crash. If the driver had acted differently, a crash may have been a near-crash or a more severe crash, or a near-crash could have become a crash. Here the underlying mathematical models of driver behavior (glance and response models) are fundamental for exploring the various possibilities (Bärgman et al., 2017). The simplest possible replacement behavior is to assume the driver sleeping. That is, that the driver does not act at all during the crash. This can be considered a worst-case behavior in any particular situation. Another way of using NDD that includes normal driving (and, possibly near-crashes) is to get distributions for stochastic variations, which can be used to both define the exposure to driving scenarios and to vary scenario characteristics. These distributions can then be used in, or together or compared with, virtual traffic simulations. There is research quantifying the relationship between near-crash increase and crash increase (Guo, Klauer, McGill, & Dingus, 2010; Victor et al., 2015), but the same cause-effect behavior is less noticeable when using normal driving data. For these reasons, working with exposure to scenarios from normal driving in relation to crash occurrence needs to be done with caution (Woodrooffe et al., 2012).

The choice of data source (and whether to remove evasive maneuvers from the original event) in a counterfactual simulation is driven by several factors, including what systems are to be assessed (e.g., whether driver behavior is to be evaluated), and whether the data are available. The availability of in-depth crash data with reconstructed pre-crash kinematics and even data recorder data is limited. When a preventive (or protective) safety system is to be developed for a specific market where in-depth reconstructed crashes or event data recorder data are not available, alternatives are needed. One option is to collect NDD and create synthetic crashes based on the structured application of models of driver behavior to non-crashes, as described above (Bärgman et al., 2017).

In this study we investigate the feasibility of using site-based NDD (highD) and in-vehicle NDD (SHRP2) non-crashes to create counterfactually simulated crashes, by comparing the resulting crash characteristics with those from reconstructed in-depth crashes (GIDAS). Comparing highD to GIDAS is comparing two samples of German highway rear-end crashes, hence we believe the comparison gives direct insights in how well the generated NDD crashes represent the reconstructed actual crashes. To study the suitability of generated crashes from the U.S. SHRP2 data, they would ideally be compared to a crash sample of identical sampling criteria, which we did not do, as such time series pre-crash data were not readily available. However, as SHRP2 is by far the most comprehensive NDD in the world to date, and GIDAS-PCM is one of the most commonly used high-quality crash datasets for counterfactual benefit assessment (of, for example, AEB), our comparison aims to study general comparability of data and results from the application of two different AEB system, rather than focusing on regional comparability. We evaluate both the crash avoidance and false positive rates of AEB systems. If the simulated crash characteristics are comparable to those of the reconstructed crashes, in the comparison of German highway rear-end crashes, then it may be feasible to use the more readily available and affordable NDD for early prospective assessments of preventive safety systems. If the U.S. NDD data were comparable to German highway crashes, then results are not sensitive to data choice, suggesting a liberal interpretation on generalization is suitable, at least for the relatively similar driving cultures of the United States and Germany.

The aim of this study can be divided into three parts: first, to compare crash characteristics generated from NDD with real reconstructed crashes; second, to quantify the influence of the data source on a comparison of the practical safety benefits of two AEB algorithms (one basic and one more advanced based on driver comfort zone boundaries, which seeks to reduce false positive activations by accounting for FV steering maneuvers); third, to demonstrate the use of NDD for assessing AEB algorithm false positive rates (and to confirm the hypothesis that the more advanced AEB system has a lower false positive rate).

2. Method

This section describes the data used in the study, the crash generation process, the AEB system application, and the simulation framework. Finally, the analysis steps are outlined.

2.1. Data

Three different sources of data were used in this study: GIDAS (Otte et al., 2003), SHRP2 (Victor et al., 2015; VTTI, 2020) and highD (Krajewski et al., 2018).

2.1.1. GIDAS – PCM

In this study-two subsets of GIDAS data, released in July 2018, were used. The first subset consists of all the highway rear-end crashes for which PCM was available. This subset was used as a reference for the counterfactual simulations assessing AEB. The second subset contained all GIDAS highway rear-end crashes in which the following vehicle (FV) did not perform an evasive braking maneuver prior to impact with a braking lead vehicle (LV), representing crashes as similar as possible to the crashes generated from NDD (where we, in the crash generation, assumed sleeping drivers). This subset includes cases for which PCM data were unavailable. Because impact speed is usually estimated for all GIDAS crashes even when PCM is not available, this subset was used as a reference in the assessment of the differences in impact speed. Within this subset (crashes with no driver evasive maneuver beforehand), PCM was available only in 7 of the 46 crashes. AEB assessment was not performed separately on these seven crashes as the low number of cases would have been too few for a relevant comparison. Note that we filtered out all but highway rear-end crashes from the GIDAS data (both with and without PCM), to maximize the match with the highD data.

The vehicles' relative longitudinal distance and lateral overlap in the pre-crash phase are used to predict the collision path in the AEB implementation. However, PCM does not directly code those values, so they were derived using other metrics in the PCM data. The predicted future path of the FV was generated as an arc with an assumed constant yaw rate for the cases when the FV turned, or as a straight line for the cases when the FV went straight. See Appendix A for a detailed description of the overlap calculations.

2.1.2. highD

The first naturalistic driving dataset used in this study is the highD dataset (Krajewski et al., 2018). It consists of processed drone video recordings of vehicles on German highways. Relevant time-varying parameters such as position, speed, and acceleration of the vehicles were extracted. The criticality of the interaction between each pair of vehicles (FV and LV) was assessed by extracting each LV's lowest acceleration value (its harshest braking maneuver) along with the time it occurred. The time headway between the FV and the LV was noted at that instant in time. Fig. 1 shows the relationship between the minimum acceleration and the time headway for all vehicle pairs in the dataset. In this study, only FV/LV interactions with a minimum LV acceleration of -2 m/s^2 or less and a time headway of five seconds or less were considered potentially critical and used in the crash generation process.

The highD datasets also provided information about the amount of lateral overlap between the LV and the FV. Fig. 2 shows the distribution of the overlaps at the time of minimum acceleration (and time headway extraction; see Fig. 1). The mode of the distribution is at approximately 1.70 m, a reasonable vehicle width in Germany. The distribution in Fig. 2 was obtained by taking the lesser of the left and right overlaps, assuming that the FV can always choose to steer left or right of the LV and that the FV and LV trajectories are always parallel to the road (information about yaw angle of the vehicles were thus not included in the data).

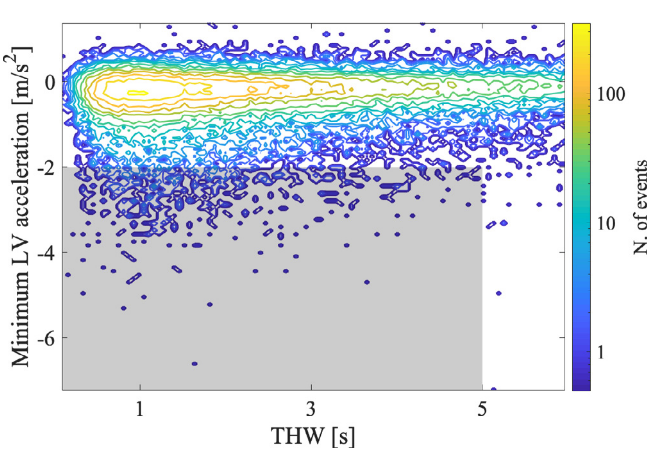

Fig. 1. Contour map of the acceleration and the time headway for all events in the highD dataset. The potential criticality of the scenario increases to the left and down. The grey area contains all the events considered for crash generation.

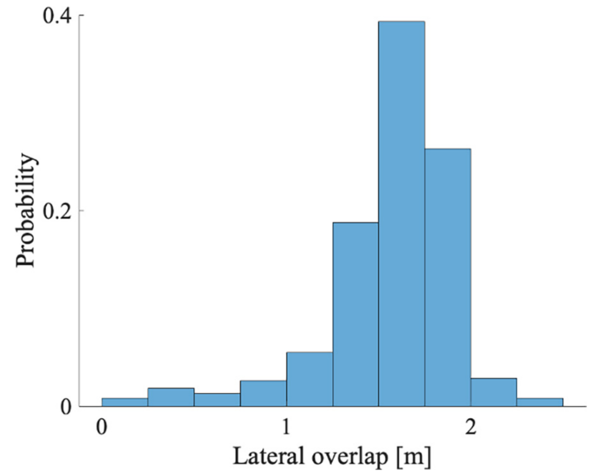

Fig. 2. Distribution of minimum lateral overlap within highD dataset.

2.1.3. SHRP2

The second naturalistic driving dataset used in this study comprises a subset of the SHRP2 naturalistic driving study. The subset originally contained 46 crashes and 211 near-crashes (Victor et al., 2015). In this study only the near-crashes were used. These near-crashes were manually reviewed by expert annotators, after an initial pre-filtering using kinematic or proximity triggers (e.g., longitudinal acceleration; see Hankey et al., 2016 p. 25–26 for details about the near-crash definition used). Of the 211 near-crashes, only 190 had the full kinematics data for both LV and FV vehicles. In 17 cases the crash generation procedure (described in Section 2.2) did not result in a crash, so those were discarded. An additional 42 of the generated crashes were not included, as the LV performed more than one braking maneuver, increasing and decreasing speed multiple times. The number of (near-crash-based) crashes used in the final analysis was thus 131. Note that, in contrast to the closely matched GIDAS PCM and highD data, the U.S. SHRP2 data were not restricted to highway crashes; it included crashes across several road types (such as rural, urban, suburban, and highway).

The comfort-based AEB algorithm (CAEB algorithm, see Section 2.3) requires lateral offset distances to make steering avoidance assessments feasible, but this information was not available in the SHRP2 dataset. Therefore, the offset distribution in the highD data was also used for the SHRP2-generated crashes, assuming parallel trajectories. Multiple simulations were run for each of the original crashes, applying each offset (bin) from the distribution in Fig. 2. The relative probabilities (weights) for each offset (bin) were considered in post-processing and the final results were calculated by weighting the simulation outcomes by their relative probabilities (per bin).

An illustration of the data usage and crash generation process in the study can be found in Fig. 3. For each dataset, only a subset was used for the crash generation and AEB application.

2.2. Crash generation

Crashes were generated from the two naturalistic driving datasets. The original kinematics of the events from highD and SHRP2 were used to define the moment the LV started braking as the moment when the LV reached an acceleration of -1 m/s^2. Decelerations closer to zero were probably caused by the driver's foot lift-

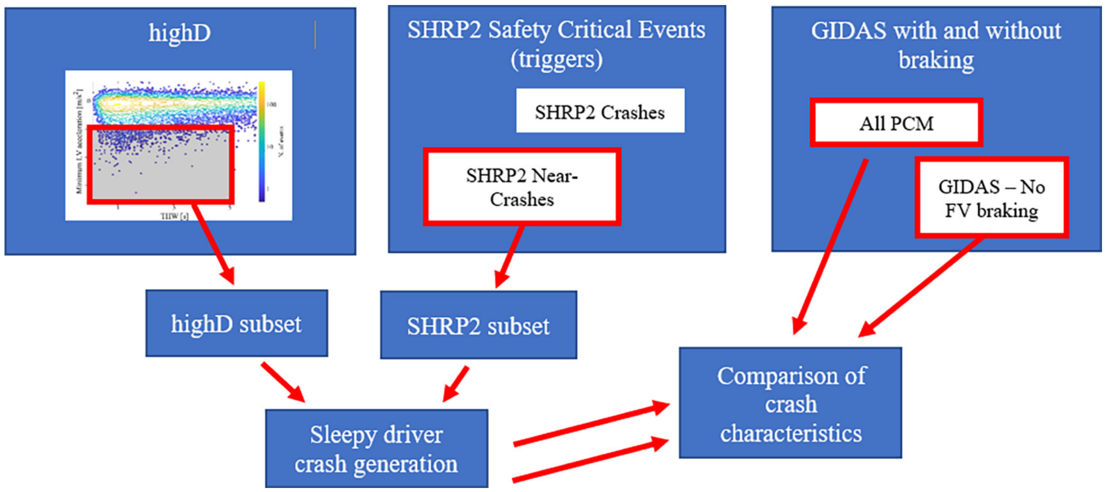

Fig. 3. Visual representation of the data selection process for crash generation and comparison, for highD, SHRP2 and GIDAS databases.

ing slightly from the accelerator pedal, and therefore were not considered further in this work. From the LV's start of braking, the speed of the FV was set to be constant until the crash happened. That is, the FV driver never performed any deceleration in response to the LV deceleration—basically simulating a sleeping driver. Fig. 4 shows this modification process, which was applied for all the highD scenarios and SHRP2 near-crashes used in the study. In many of the selected highD cases, although criticality was established by the LV kinematics, the event ended before the FV could reach the LV because it occurred near the end of the segment of road that each drone covered and recorded. (The segment was only 420 m long, and the LV decelerations happened at different points over this distance). A total of 361 events were excluded from highD dataset as a result. In total, 378 crashes were generated from highD and 131 (all) from SHRP2.

2.3. AEB algorithm descriptions

This work used a reference AEB algorithm (RAEB) and an AEB based on drivers' comfort zone boundaries with respect to lateral acceleration (comfort-based AEB; CAEB). Each algorithm was applied to the crashes from all three datasets in order to compare crash avoidance and mitigation results.

The RAEB, based on the work of Brännström et al. (2008), only considers possible longitudinal avoidance by the system. That is, it does not include the driver's capacity to avoid the crash by comfortable steering, instead identifying the moment when the deceleration required by the system to avoid the crash passes a threshold (the point in time after which the FV would be unable to avoid a crash). The RAEB considers the current speed, acceleration, and relative distance of the FV and LV as well as the maximum braking performance of the FV braking system. The values used to quantify braking performances were the maximum deceleration reachable by the vehicle ($-10\ \mathrm{m/s^2}$), the jerk reachable by the braking system ($-50\ \mathrm{m/s^3}$), and the time delay of the braking system (0.08 s, from Bärgman et al., 2017; Brännström et al., 2008). After the RAEB activation, the FV evasive maneuver (braking) played out according to the values used as input for the braking system limits. If the vehicle was traveling in a straight line, the braking maneuver was simulated in the same direction of travel. If the vehicle was turning (e.g., in a curve), the braking maneuver was simulated with the assumption that the FV traveled at a constant steering angle.

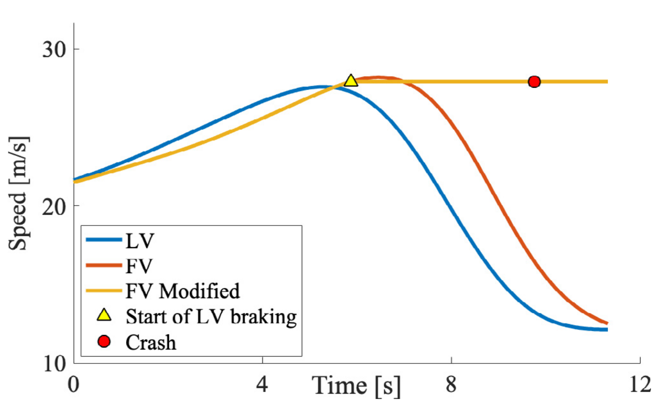

Fig. 4. Conceptual demonstration of the removal of the FV's braking.

The CAEB algorithm was also applied to all crashes in the study. Unlike the RAEB, this algorithm took into account the capability of drivers to avoid crashes by performing a comfortable steering maneuver. Depending on the relative speeds involved in the event, a driver might still comfortably perform an evasive steering maneuver to avoid a crash even when the BAEB may have already triggered (Brännström et al., 2014). Thus the CAEB may eliminate some early interventions (potential false positives). The algorithm also includes parameters of driver comfort limits in terms of lateral acceleration and a basic single-track bicycle model that defined the lateral dynamics of the vehicle. The bicycle model is only a first approximation of a vehicle but it was considered sufficiently accurate for this study.

The CAEB algorithm simulated an S-shaped maneuver by the driver: at the end of its trajectory, the FV is parallel to its position at the beginning and is next to the LV. That is, the FV and the LV are in the same longitudinal position but separated by a lateral safety distance (see Appendix A). The S shape was designed as follows: the angle of the steering wheel was gradually increased, considering the limit for steering wheel speed of 720°/s (Brännström et al., 2014) and the driver comfort limit for lateral acceleration of $5\ \mathrm{m/s^2}$ (Sander, 2018). The latter is determined by the vehicle turning (yaw) rate and the vehicle speed. When the maximum tolerable lateral acceleration (the driver's comfort zone boundary) was reached, the steering angle was kept constant until the FV steered back, ending its trajectory parallel to its position at the start of the maneuver.

A key parameter for the generation of the FV's trajectory was the lateral distance the FV needed to traverse to avoid a crash with the LV, which depends on the lateral overlap (see Appendix A). The simulations of the trajectories included the lateral overlap, and the additional safety distance simulations included run-time estimations of the lateral distance traveled by the FV. To speed up simulations, only the first half of the FV trajectory was simulated: after the FV had traveled half the required lateral distance, the second half of the trajectory was assumed to mirror the first half. Additional vehicle dynamics during lane changing, such as tire slip, were ignored.

The CAEB algorithm was designed to reduce false positive activations and thus activates later than RAEB. Hence, RAEB intervention timing was used as the starting point for the activation decision of CAEB. From this starting point the vehicles were simulated as follows: (a) the LV continued along its original path and (b) the FV was projected along the newly created steering trajectory. If the FV was able to complete the steering trajectory and avoid a crash with the LV, then the AEB intervention was delayed (by one time-step); otherwise the AEB activated. This process was iterated, with the FV getting closer to the LV at each iteration, until the new trajectory resulted in a crash with the LV. Once the steering trajectory resulted in a crash, the AEB was activated and the evasive braking maneuver was applied (simulated).

2.3.1. False positive assessment

The false positive performances of the two AEB algorithms were assessed by applying them to the original events used to generate the crashes for the AEB assessments. Only the potentially critical highD events used to generate crashes (378 events, see Section 2.2) were considered of interest for the false positive assessment; all the other events recorded weak or absent LV braking maneuvers. For the false positive assessment of SHRP2, the original events included only previously selected near-crashes (131 events, see Section 2.1.3). In summary, critical (non-crash) events from the original highD and SHRP2 data (including the original braking behavior of the FV) were simulated with the RAEB and CAEB systems.

3. Results

In this section the results of the study are presented. First, for all datasets, generated crashes are compared to real-world crashes. Second, the results of the AEB algorithms' application to the data are shown, followed by the results from the analysis of false positives.

3.1. Crash comparison

Fig. 5a shows a comparison of the cumulative distributions of the maximum level of deceleration reached by the LV in the pre-crash phase across the three datasets. The SHRP2 and PCM events show harsher braking maneuver s (higher values of deceleration) than the highD events. However, there are more PCM events with relatively low maximum decelerations, in which the LV did not brake or only braked slightly. Note that this also includes LVs that stand still, for example in a traffic jam.

Fig. 5b shows a comparison of the time elapsed from the start of LV braking to the crash across the three datasets. The distributions for SHRP2 and highD consider all the generated crashes, while the PCM distribution only includes crashes in which the LV braked with a deceleration of at least -1 m/s^2 (60 crashes, from Fig. 5a), the same threshold used for the highD crash generation. PCM crashes where the LV applied more than -1 m/s^2 deceleration show a time-to-crash comparable to that of crashes generated from SHRP2 near-crashes. However, there is a substantial difference in the time-to-crash between SHRP2 and PCM on the one hand and highD on the other.

Fig. 5c compares the impact speeds across the three datasets (without AEB applied to the data). As the analyzed crashes are rear-end crashes only, the impact speed was computed as the relative speed between the vehicles, assuming that they were driving parallel to each other. The crashes generated from the naturalistic datasets show an overall substantially lower impact speed compared to the real crashes in PCM. The distribution of crashes in the GIDAS database where the LV was braking but the FV did not brake is also shown. Because not all crashes in GIDAS have been reconstructed into PCM (recall that PCM data were available only for seven of these crashes), simulations were performed on all rear-end PCMs (including those with FV braking), increasing the case count to N = 134. The distributions of the impact speed for all rear-end PCM crashes and the GIDAS no-FV-braking crashes are similar, especially for impact speeds between 10 m/s and 20 m/s, with larger differences in the tails.

Fig. 5d shows the relative FV-LV speed at the point when the LV deceleration reaches -1m/s^2, or, if the LV did not brake, when the FV starts braking or, if also the FV did not brake, when it crashes. The two NDD are very similar in the initial conditions, while the PCM data has much higher initial relative speeds.

Fig. 5e shows the comparison of the lateral overlaps (see Sec 2.1.1) at the time of the crash, indicating that the overlaps were lower for the PCM crashes than for the highD-based crashes. As noted, SHRP2-based crashes did not include information about the lateral overlap, so they were not included in the comparison.

3.2. The influence of data source choice on the comparison of AEB safety performance

The two AEB algorithms were applied to the crashes of all datasets. The RAEB only considers longitudinal kinematics, while the more advanced CAEB aims to decrease early (nuisance) interventions by accounting for the potential of the driver's evasive action (comfortable steering). Fig. 6a shows the cumulative frequencies of impact speed when the RAEB is applied and the crashes are mitigated, but not completely avoided (non-crashes are excluded; remaining crashes are N = 22 for PCM, N = 11 for highD and N = 12 for SHRP2). The remaining crashes all have lower impact speeds than the original crashes, and the crashes generated from highD and SHRP2 have lower impact speeds than the crashes from the PCM. Fig. 6b shows the cumulative frequencies of impact speeds of the mitigated crashes when the CAEB is applied to all three datasets. The impact speeds are higher than those obtained with the RAEB, and fewer original crashes were avoided (N = 34 for PCM, N = 41 for highD and N = 19.9 for SHRP2). (Recall that for SHRP2 crashes, the results include the weighting process described in Section 2.1.3, applying the offset distribution from highD crashes, which is why there are non-integer crash results for SHRP2). Fig. 6c shows the crash avoidance performances (as percentages of original crashes) of the two tested algorithms. As expected, the RAEB avoided more crashes than the CAEB for all the datasets tested.

3.3. False positive analysis

False positives were analysed in the original highD and SHRP2 no-crash events. The RAEB and CAEB algorithms were applied and the results were compared. The application of RAEB to highD resulted in four false positives, all occurring at very low-speed (<3 m/s) events in traffic jams, with FV and LV vehicles closely following each other. CAEB did not avoid any of these four false positive interventions, probably because of the short distances between vehicles (as they were low-speed events), together with the fact that steering is much less effective at low speeds. For SHRP2, the analysis of false positives was first performed on all

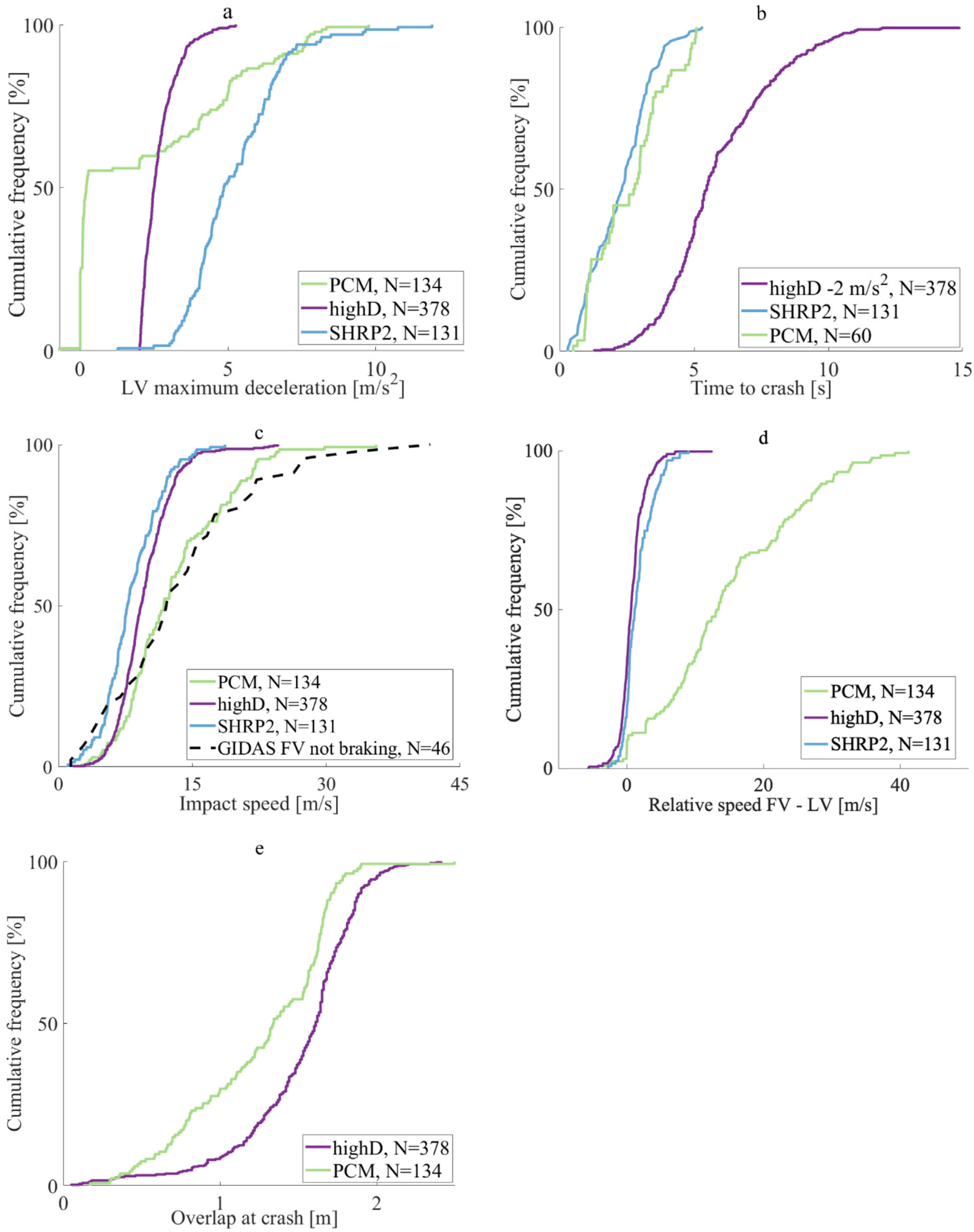

Fig. 5. (a) Cumulative frequency of LV maximum deceleration during the pre-crash phase. (b) Cumulative frequency of the time elapsed from LV braking initiation to the crash (at -2 m/s^2). (c) Cumulative frequency of original impact speed of all datasets. (d) Cumulative frequency of the relative speeds at the start of the event. (e) Cumulative frequency of lateral overlap between LV and FV at the time of crash.

131 events, and then on a subset of these events—those in which at least one of the vehicles reached a speed of 60 km/h (N = 42). This subset was considered more similar to the other datasets in this study. The RAEB application resulted in 28 false positives in 131 events (21.3%); seven occurred in the 42 high-speed events (16.7%). As noted previously, for the application of CAEB to SHRP2 data the simulations used the overlap distribution from the highD data (see Fig. 2). The simulation results were weighted according to the probability of each simulated overlap. The results of this procedure were not necessarily integers. To make this apparent to the

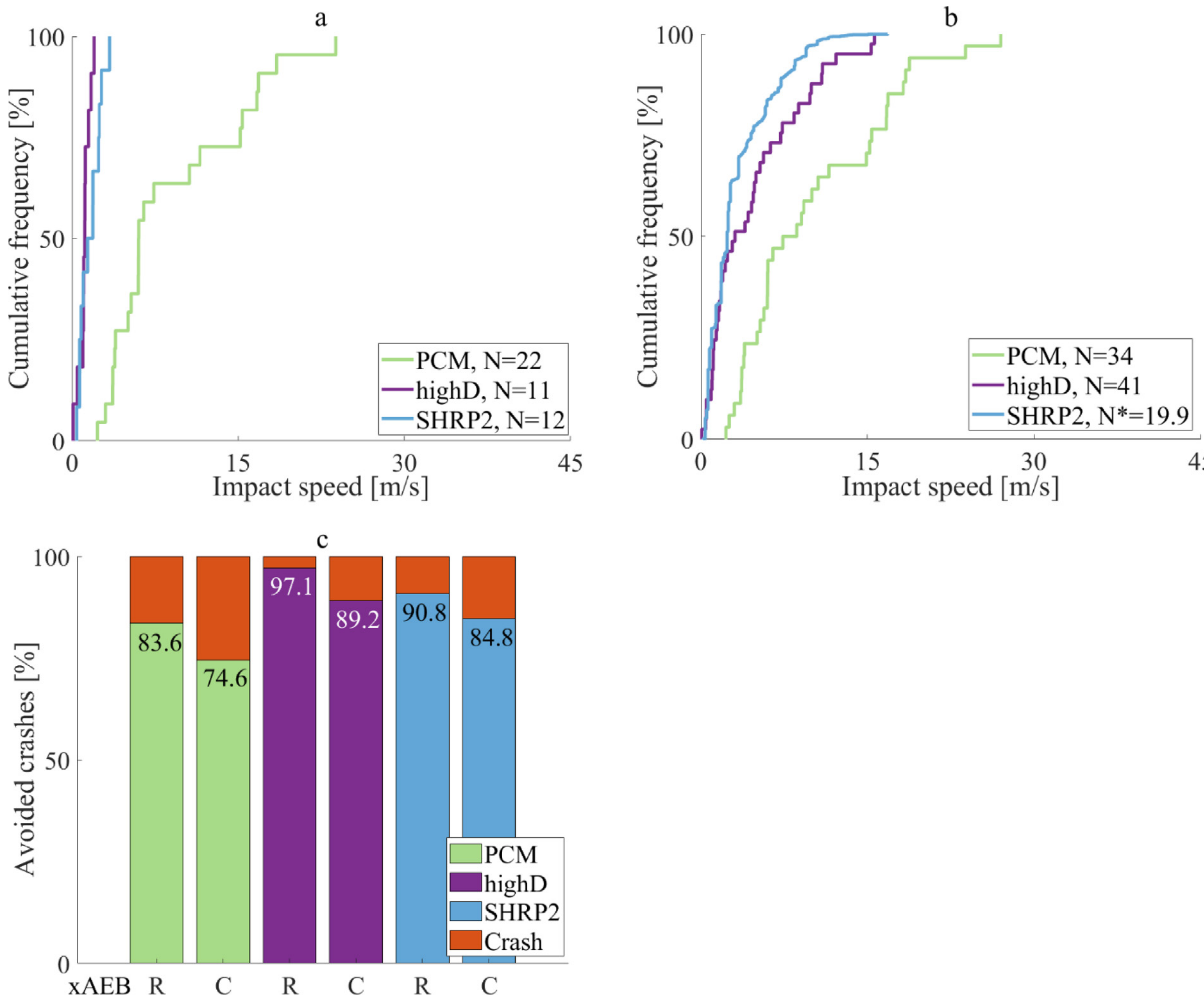

Fig. 6. (a) Cumulative frequency of the impact speed in crashes remaining after basic AEB application. (b) Cumulative frequency of the impact speeds in crashes remaining after advanced AEB application. N* is the theoretical number of avoided crashes resulting from the weighting process of the probabilities of the overlaps. (c) Bar plot showing the percentage of avoided (original) crashes for each dataset and both AEB systems.

reader the false positive counts were, in Table 1, intentionally left with one decimal place. The false positives decreased from 28 to 24.8 (from seven to 5.9 at high speed) when the CAEB was applied. The results are summed up in Table 1.

As expected, when SHRP2 crashes were simulated with small overlaps (see Fig. 2), the CAEB produced fewer false positives than the RAEB. As an example, the CAEB avoided six false positives (two of which were at high speed) at the (Fig. 2) bin with the smallest overlap.

4. Discussion

This study explored the possibility of using crashes generated from non-crash naturalistic driving data (NDD) to complement, or, in some instances even replace, real crashes with time-series kinematics (e.g., from reconstructions) for counterfactual assessment of AEB systems. This possibility would be useful for analysis of safety benefit (and, potentially, system optimization) for countries lacking crash data with time-series kinematics. We have shown that, in general, crashes generated from NDD have substantially lower impact speeds than real crashes (GIDAS PCM), but the pre-crash criticality in crashes generated from U.S. NDD (SHPR2) near-crashes is comparable to the criticality in real German crashes.

Table 1
False positive counts in highD and SHRP2 events for the application of RAEB and CAEB.

	highD		SHRP2	
	RAEB	CAEB	RAEB	CAEB*
Low speed	4	4	21	18.9
High speed	0	0	7	5.9
Total	4	4	28	24.8

* Values for CAEB applied to SHRP2 are weighted according to the overlap probability, possibly resulting in non-integers.

4.1. Comparing crash characteristics across datasets

4.1.1. Crash generation

The events from the two naturalistic datasets used in this study, the highD everyday highway driving and the SHRP2 near-crashes, by definition are not crashes. As an aim of this study was to investigate the suitability of using non-crash NDD to simulate counterfactual crashes, only near-crashes from SHRP2 were included in the study (i.e., not crashes). To generate a crash, the FV braking maneuver had to be removed, so the start of the FV braking had to be defined. In other studies this process was done manually (Bärgman et al., 2017) or computationally by fitting a piecewise linear model to the FV deceleration (Markkula, Engström, Lodin, Bärgman, & Victor, 2016; Svärd, Markkula, Engström, Granum, &

Bärgman, 2017). In this study, however, a different approach was used: an LV deceleration threshold of -1 m/s^2 defined the point in time when the FV speed was set to remain constant, eliminating any further deceleration. That is, the FV driver may or may not have initiated braking at that time. This process can be seen as simulating a substantially "distracted" FV driver or, maybe more accurately (as the duration of eyes-off-road is typically quite long) a sleeping driver, as the driver does not react to the unfolding of the critical event. This approach was used because of the different nature of the datasets analyzed compared to those in previous studies. Removing the evasive maneuver can be considered the worst possible outcome (unless the FV driver accelerates into the crash, which would be even worse — but would be unlikely).

The times needed for the FV to crash into the LV (Fig. 5b) were similar between GIDAS PCM and SHRP2, reflecting the fact that SHPR2 data actually capture critical events—but they were substantially different for highD and SHRP2, reflecting the different origins of the two datasets (everyday driving vs critical events). This difference indicates that highD data are likely not suitable as an artificial source of critical events, neither with respect to timing (criticality) nor impact speed. Some highD events proved to be more safety–critical than others, but a much more extensive data collection (capturing more critical events) is needed to make a highD-like dataset even marginally useful for, for example, AEB assessment. Note that SHRP2 collected over 3,958 driving man-years (the total number of years that data were collected of participants' everyday driving), and still only captured 125 rear-end crashes. Although highD captured many vehicles simultaneously, the amount of data collected (time per vehicle) was several orders of magnitude less than that of SHRP2; further, all data were recorded on straight highways.

In addition to differences in timing and impact speed, the differences in LV decelerations between the datasets (Fig. 5a) is also likely to affect the AEB assessment. In particular, there were many PCM crashes that had low, or no, LV deceleration, while the SHRP2 LV decelerations were much higher compared to highD, and the highD decelerations were quite uniform (and never lower than 1 m/s^2, per definition). The LV decelerations for highD are simply everyday driving decelerations. In order to produce the high impact speeds of the PCM, they are likely not simple car-following events with the LV braking creating the crash; instead, the LV is at a standstill, or driving at a substantially lower speed than the FV, with the FV catching up (or the LV is cutting in front of the FV, at high relative speed). The differences in scenarios can also be seen in Fig. 5d, where 13% of the PCM cases had a relative speed of up to 10 km/h, while 75% of the SHRP2 cases and 89% of the highD cases had low relative speeds (up to 10 km/h). These findings demonstrate that much care should be taken when (if) using crashes generated from near-crashes for virtual safety assessment. Selection criteria (incl. event categorization – specific subtype of the events used) must be at the forefront of the considerations. If the algorithms act differently in the two types of events (e.g., LV braking and FV catch-up), the results are likely to be misleading.

4.1.2. Crash comparison

The comparison of impact speed (Fig. 5c) makes it clear that overall, the GIDAS PCM crashes have much higher impact speeds compared to the generated crashes. Although the PCM crashes were only highway rear-end scenarios, there were sometimes specific circumstances resulting in more critical events, such as the very late appearance of the LV in the FV's path (e.g., due to a cut-in). Therefore, the higher impact speeds were expected, especially when compared to the SHRP2-based crashes, which were mostly lower-speed events (and only a relatively small proportion occurred on highways). That is, the low initial speeds (and possibly short time headway) in the SHRP2-based crashes—and the fact that most were car-following situations (and not FV catch-up situations, with large relative speed differences)—kept them from becoming high impact-speed crashes such as those found in GIDAS PCM data.

What makes the situations less critical in the highD than in the GIDAS PCM is that the LVs typically decelerate less in the former (see Fig. 5a). Further, in the latter, although the LV did not always brake, there were also cases with high LV deceleration (when the LV appeared suddenly in the path of the FV). These cases did not occur in highD: as noted, the deceleration in highD was more uniform.

In summary, the results of this study show that the impact speeds of crashes generated from site-based or in-vehicle NDD near-crashes during normal driving are substantially lower than the impact speeds of real crashes—at least, the real crashes from the GIDAS PCM used here. One reason for the difference is probably that, for an event to be included, the GIDAS PCM required that at least one person be suspected of being injured. If the PCM data also included crashes without personal injury, it may be that the crashes generated from NDD would be similar to the lower tail of the PCM data's impact speed distribution. However, as safety assessment typically prioritizes avoiding human injuries, this observation may not be relevant.

The comparison of lateral overlap at the time of the crash (Fig. 5e) showed similarities between the highD and GIDAS PCM data, but PCM on average had smaller overlaps (0.26 m smaller). The difference in distributions is noticeable for medium overlaps (0.8–1.6 m), while for small overlaps (<0.5 m) the distributions are relatively similar. This difference could be due to an evasive steering maneuver performed by the driver of the FV in the PCM events, resulting in a crash involving only one portion of the vehicle front bumper. In highD the cases with small overlaps are rarer, as the crashes were generated from normal highway driving scenarios, with the vehicles usually driving in the middle of the lane.

The highD overlaps represent normal driving behavior, so we believe that it is reasonable to virtually apply them to NDD near-crashes as part of the crash generation process. However, although the SHRP2 and GIDAS overlaps were similar, the limited event matching between highD and SHRP2 reduces the validity of applying the highD overlaps to SHRP2 data—yet another argument for aiming to use better matched datasets. In fact, the effect on the results if the assumption of similarity of overlaps is violated is not obvious. Actually, it may be counterintuitive: if the overlap in SHRP2 is smaller than that applied from highD (Fig. 2; requiring less steering to avoid a crash), the AEB would likely avoid fewer crashes (since the AEB response would be delayed)—and vice versa (a larger overlap would result in more avoided crashes).

4.2. The influence of data source choice on the comparison of AEB safety performance

4.2.1. Comparing crash characteristics across AEB algorithms

Two different AEB algorithms were applied to the crashes from the three datasets. The first algorithm, the RAEB, performs a run-time threat assessment based exclusively on the longitudinal kinematics of the two vehicles. When the algorithm detects that the FV will soon be unable to brake in time to avoid a crash, it initiates an automated braking maneuver. The second algorithm, the CAEB, goes a step further and accounts for the possibility that the driver will avoid the crash by comfortable steering. Numerous studies consider steering as an opportunity to avoid crashes (Brännström et al., 2010, 2014; Sander, 2018). The purpose of this addition to the AEB algorithm is to try to avoid as many unnecessary interventions as possible, as there could be cases in which the driver is aware of the LV and is planning to perform a late, but comfortable, steering maneuver: for example, when about to overtake.

On the one hand, the number of interventions that are too early to be accepted by the driver could potentially be reduced by considering comfortable steering, but on the other hand these delayed AEB interventions result in more, or more severe, crashes (i.e., increased impact speed; see Fig. 6a,b). That is, if the FV driver does not perform the expected steering maneuver, the vehicle's AEB system may no longer have the time to brake to avoid the collision. This outcome is also shown in Fig. 6c, where the CAEB is less effective at avoiding crashes than the RAEB. Both the fewer avoided crashes and the less reduced impact speed are direct consequences of the delayed AEB intervention. This finding was expected; safety system manufacturers are constantly balancing the potential improvement in user acceptance (Bliss & Acton, 2003; Coelingh et al., 2007) against the decreased effectiveness of the safety system at reducing the number and severity of rear-end crashes.

If crashes generated from near-crashes are to be used for AEB safety assessment, it is important to perform sensitivity analyses on the effects of LV decelerations, and differences in performance between the car-following versus FV-catch-up-to-LV (or cut-in) scenarios, on the preventive systems performance; our analyses show that the differences between the PCM and near-crash generated crashes are large with respect to LV decelerations and scenario type, also for near-crashes. This will likely have different impact on the safety performance assessment, depending on the AEB algorithm. Possibly probability weighting on, for example, scenario type, based on more representative crash databases, may mitigate potential effects of data source differences with respect to safety performance.

4.2.2. False positive analysis

Analysis of false positives is an important part of the assessment of preventive safety systems (Bliss & Acton, 2003; Coelingh et al., 2007; Sander & Lubbe, 2016). With respect to our false positive analysis, it is important to note that when an AEB system was triggered for a near-crash in the SHRP2 data, the situation may not actually have been a false positive, as it actually was quite critical. However, for consistency we decided to consider all AEB triggers as false positives in our analysis and discussion.

There were substantial differences between highD and SHRP2 datasets in the results from the simulation and analysis of false positives. The few highD false positives were all cases in which the FV was following the LV closely at a very low speed, representative of a traffic jam—perhaps not very relevant for human injury prevention. The SHRP2 events, on the other hand, had more false positives, and the dynamics were more safety–critical (at least, the speeds were higher). These differences mirror the near-crash nature of the SHRP2 events compared to the normal highway driving in highD (see Table 1).

The number of false positive activations was smaller with CAEB than RAEB, as logic suggests, although the reduction was substantially less than expected. However, we do show that crashes generated from NDD near-crashes are potentially a good source of data for false positive analysis. In contrast, highD, which contains approximately 17 hours of data (60 recordings of 17 minutes) with approximately 30 vehicles in the image at all times and records FV/LV interactions in normal driving data, appears not to be very useful for understanding even AEB false positive performance – if an AEB system triggered in such events, it would truly be a poorly designed system.

4.3. Limitations and future work

A main limitation of this study is the assumption that FV drivers are not reacting at all to LV braking. That is, we are basically assuming the drivers are sleeping. However, it is possible to virtually add glance behaviors and brake responses to generated events (Bärgman et al., 2017; Bärgman, Lisovskaja, et al., 2015; Lee, Lee, Bärgman, Lee, & Reimer, 2018), using distributions of documented glance behaviors (Morando, Victor, & Dozza, 2019) and driver response models (Markkula et al., 2016; Svärd et al., 2017). This aggregation of knowledge about driver behavior enables the generation of synthetic crashes and counterfactual simulations that take other factors, such as driver distraction, into account (Bärgman, Lisovskaja, et al., 2015; Bärgman & Victor, 2020). However, although driver glance behavior models have been included in previous studies, the approach to apply glances to non-crashes as part of crash generation has not been systematically validated (using, e.g., in-depth crash data). Applying glances to non-crashes in a study similar to this one would bring us one step closer to validating the approach.

In addition to improving the generation of synthetic crashes by adding driver behaviors, investigating different types of crash scenarios could be a next step—expanding from rear-end crashes to, for example, intersection crashes. There are site-based intersection NDD available from drone collections (Bock et al., 2020), similar to the highD data; however, given the results of our study, they would likely be of little use for safety assessment of, for example, AEB. Going forward, the focus must be on identifying and using data sources containing critical events, instead of using small samples of everyday driving. It may, however, still be relevant to use crashes generated also from drone based NDD, such as Bock et al. (2020), for methodological work related to safety assessment, where correct absolute safety benefit estimates are not the main focus.

Further, the simulations themselves are simplifications of reality. For example, the inclusion of more advanced vehicle models and mature AEB algorithms (e.g., in production) would likely improve the generalizability of the simulations. Future research would benefit from working closely with vehicle manufacturers, which have access to both of these. Also, for the RAEB algorithm, all vehicles were assumed to have the same values of jerk and (reachable) deceleration, regardless of the varying performances that different cars can have, and the weather conditions (e.g., rain and ice—braking performance was not tuned to reflect possible changes in road friction coefficient).

It is also important to consider what specific variables are important for specific benefit analysis. This study has shown that the lateral overlap between the LV and FV is important when using rear-end NDD to assess rear-end crashes. Consequently, its precise and accurate recording should be a priority in future data collection – also in critical event NDD recordings.

In this study the scope was to investigate how normal driving and near-crash NDD can (or cannot) be used to generate crashes for counterfactual safety assessment, with focus on AEB (which will be part of all levels of vehicle automation for the foreseeable future). We are not studying how NDD could be used in the process to generate crashes through, for example, traffic simulations. Research and development should continue for both the traffic simulations-based approach (typically targeting higher levels of automation) and for counterfactual safety assessment, with focus on ways of ensuring precise and accurate results, while making sure stakeholders easily can understand assumptions and limitations, and their implications on results and interpretation. Further, as needs now arise to assess the safety of higher levels of vehicle automation, estimating crash rates and exposure "of the future" is crucial. However, the method used in this study – counterfactual simulations (with or without crash generation) – do not attempt to predict the future rates. Instead the crash rates from, for example, crash databases would typically be used (i.e., simply using the original base rates per scenario type). Further development of methods to estimate future crash rates and scenario exposure is needed, where the use of traffic simulations is one path that should be pursued, with focus on validation.

Finally, another main limitation of our analysis is the limited matching of events across datasets. Although the comparability between the highD data and the GIDAS data is high, as both were collected on German highways, the types of scenarios differ substantially: the highD events are almost all car-following, while many of the GIDAS crashes consist of a car catching up to a lead vehicle at high relative speed. The SHRP2 and GIDAS data differ in a different way: the SHRP2 near-crashes rarely occurred on highways and the driving speeds were substantially lower than in GIDAS; there was also a large difference in the proportion of car-following versus catch-up scenarios. Further, the fact that highD is site-based NDD, while SHRP2 and GIDAS data are collected in-vehicle (SHRP2) and crash-based (GIDAS), is likely to affect comparability. Site-based NDD is unlikely to capture crashes that are related to driving context and infrastructure, while continuous in-vehicle and crash-based data collection capture crashes (and for SHRP2, critical events) in all contexts. In the SHRP2 database the events were all safety–critical near-crashes, but only a fraction of the events was at high speed, unlike highD and GIDAS PCM databases, which included mostly high-speed driving events. Also, our use of German crash data (GIDAS) and U.S. in-vehicle NDD (SHRP2) makes it difficult to determine if differences are primarily due to differences between regions, or if they are fundamental data source differences. However, the similarities in criticality indicate that some liberal generalizations (also between regions) on the relative effect of different AEB systems may be possible, in our case likely due to the driving cultures of Germany and the United States being relatively similar. The validity of such liberal generalization is, however, likely much dependent on the driving cultures of the involved regions (see, e.g., the comparison between China and the United States, Bianchi Piccinini, Engström, Bärgman, & Wang, 2017)). Consequently, future studies generating and analyzing more closely matched events, differing only in that some are generated from near-crashes and others from in-depth crash databases, would be most valuable.

5. Conclusion

In-depth crash data with reconstructed pre-crash kinematics can be used to develop both protective and preventive safety systems that are highly effective. Since such data are not available everywhere, alternative data sources are needed to make at least rough estimates of system performance in regions without them, as part of the system development process. This work studied the suitability of using easier-to-obtain non-crash naturalistic driving data (NDD) as a complementary data source for use in virtual assessment of preventive safety systems (specifically AEB).

Results show that virtual AEB assessments based on site-based NDD recordings of everyday driving on highways had neither the criticality nor the impact speed of assessments based on traditional pre-crash kinematics from in-depth reconstructions of crashes. We have consequently shown that site-based NDD that only capture a few tens of hours of normal driving are not suitable for assessing preventive safety performance, crash avoidance, or impact and injury risk reduction.

However, our results also show that the event criticality and the proportion of avoided crashes (but not impact speeds or impact speed reductions) were similar between crashes based on U.S. near-crashes and those based on a traditional German in-depth crash database. Therefore, critical-event near-crash data may be useful to complement in-depth crash data when comparing the safety benefit of different systems. The near crash data also allows an assessment of false-positive activations highlighting differences between systems. However, our results show that data sources that include original crashes, such as in-depth crash data, are still very important and preferred.

With respect to practical applications of our research, the results from our study can be used by system developers and researchers when deciding which data to use for virtual safety assessment (e.g., if NDD is an option). The paper further provides insights into the limitations of NDD for safety assessment, which is important to understand when NDD are considered for use in virtual safety assessment.

Funding

This work was supported by Autoliv Research and the MSc program Master Program of Automotive Engineering at Chalmers University of Technology.

Conflict of interest

Nils Lubbe works at Autoliv Research, located in Vårgårda, Sweden. Autoliv Research is part of Autoliv (www.autoliv.com), a company that develops, manufactures and sells, for example, protective safety systems to car manufacturers. Autoliv is a tier 1 supplier. Results from this study may impact how Autoliv choose to develop their products. Pierluigi Olleja and Jonas Bärgman have no interests to declare.

Acknowledgements

We want to thank Xiaomi Yang and Bo Sui for valuable discussions, and Hanna Jeppsson for her support on understanding the PCM data. Further, we want to thank the Institute for Automotive Engineering RWTH Aachen University, which owns the highD data, for providing access for this research project (the paper is partially based on work performed as part of the main authors' MSc thesis). We also want to thank VTTI, the Transportation Research Board, and the National Academies for providing access to the SHPR2 data. The data was made available under the data licensing agreement SHRP2-DUL-16-A-204. The findings and conclusions of this paper are those of the authors and do not necessarily represent the views of VTTI, the Transportation Research Board or the National Academies.

Appendix A

In the case of an intersection of the predicted FV path with the LV position, the relative longitudinal distance was measured following the FV's predicted path (along either the arc or the line). The lateral overlap is the lateral relative distance between the FV and the LV. That is, if all vehicles were of equal width and directly behind each other (complete overlap), there would be only one overlap value: the vehicle width. That is, the following driver needs to move a full vehicle width to the left (or right) to just barely avoid crashing, in a critical rear-end situation. Overlap values lower than the vehicle width occur when the LV and the FV are travelling at different lateral positions in the same lane. Overlap was computed by measuring the distance of the four vertices of the rectangle representing the LV to the centreline of the path for left (d_{lLV}) and right (d_{rLV}) sides of the LV and adding half the width of the FV (d_{lFV} or d_{rFV}, right and left half of the FV width, respectively, based on whether the steering maneuver is about to take place to the left or to the right of the LV) and an additional safety distance of one metre (e), so that once the FV has completed the lateral movement it has some lateral clearance to the LV, rather than almost touching

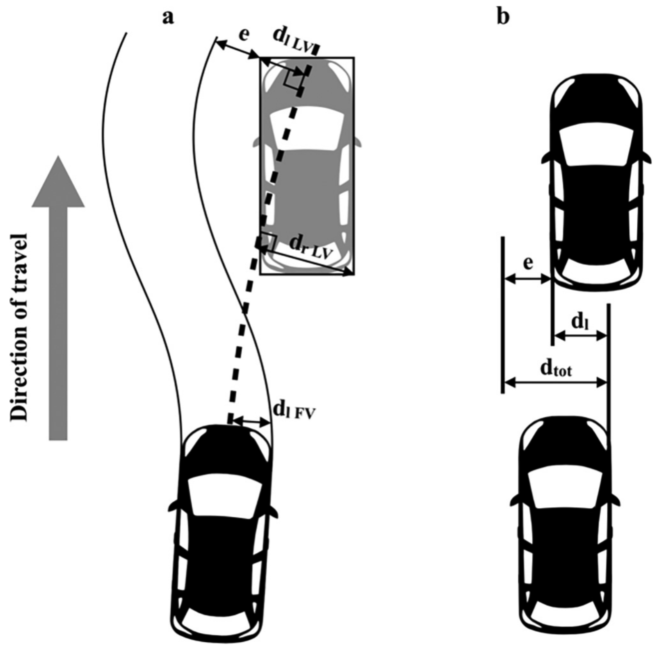

Fig. A1. Overlap measurements between LV and FV. The example shows two scenarios: a) FV steering right towards LV with a curved future path (dotted line); "$d_{l\,LV}$" and "$d_{r\,LV}$" are left and right distances from centreline of the future FV path to LV the furthest corners of the LV space to the left and the right, respectively; comfortable maneuver is chosen to be to the left of the LV (solid lines), moving laterally of "$d_l = d_{l\,LV} + d_{r\,LV} + e$", where "e" is the additional safety distance; b) FV and LV parallel to each other; "d_{tot}" is the total clearance distance required to avoid a collision, comprising overlap "d_l" and "e".

as it passes (see Fig. A1a,b). Fig. A1b shows how the total clearance distance (d_{tot}) was measured: it includes the overlap of the vehicles ($d_{lLV} + d_{lFV}$) and the additional safety distance (e). This procedure assured the availability of all the relevant metrics needed for the AEB application for the PCM crashes.

References

Alvarez, S., Page, Y., Sander, U., Fahrenkrog, F., Helmer, T., Jung, O., Hermitte, T., Düering, M., Döering, S., & Op den Camp, O. (2017). Prospective effectiveness assessment of ADAS and active safety systems via virtual simulation: A review of the current practices. *Proceedings of the 25th international conference on Enhanced Safety of Vehicles (ESV)* http://www-esv.nhtsa.dot.gov/Proceedings/25/25ESV-000346.pdf.

Bakker, J., Jeppsson, H., Hannawald, L., Spitzhüttel, F., Longton, A., & Tomasch, E. (2017). IGLAD - International Harmonized In-Depth Accident Data. In *Proceedings of the 25th international conference on Enhanced Safety of Vehicles (ESV)* (pp. 1–12).

Bareiss, M., Scanlon, J., Sherony, R., & Gabler, H. C. (2019). Crash and injury prevention estimates for intersection driver assistance systems in left turn across path/opposite direction crashes in the United States. *Traffic Injury Prevention, 20*(sup1), S133–S138. https://doi.org/10.1080/15389588.2019.1610945.

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 Behavior, 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 Behavior, 35*, 170–184. https://doi.org/10.1016/j.trf.2015.10.003.

Bärgman, J., & Victor, T. (2020). Holistic assessment of driver assistance systems: How can systems be assessed with respect to how they impact glance behavior and collision avoidance? *IET Intelligent Transport Systems*. https://doi.org/10.1049/iet-its.2018.5550.

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). *L3Pilot deliverable D7.4 – impact evaluation results*. https://l3pilot.eu/fileadmin/user_upload/Downloads/Deliverables/Update_14102021/L3Pilot-SP7-D7.4-Impact_Evaluation_Results-v1.0-for_website.pdf.

Blatt, A., Pierowicz, J., Flanigan, M., Lin, P.-S., Kourtellis, A., Jovanis, P., Jenness, J., Wilaby, M., Campbell, J., Richard, C., Good, D., Czar, N., & Hoover, M. (2015). Naturalistic driving study: Field data collection. In *No. SHRP 2 Report S2-S07-RW-1*. https://doi.org/10.17226/22367.

Bliss, J. P., & Acton, S. A. (2003). Alarm mistrust in automobiles: How collision alarm reliability affects driving. *Applied Ergonomics, 34*(6), 499–509. https://doi.org/10.1016/j.apergo.2003.07.003.

Bock, J., Krajewski, R., Moers, T., Runde, S., Vater, L., & Eckstein, L. (2020). The inD dataset: A drone dataset of naturalistic road user trajectories at german intersections. In 2020 IEEE Intelligent Vehicles Symposium (IV) (pp. 1929-1934).

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

Brännström, M., Coelingh, E., & Sjöberg, J. (2014). Decision-making on when to brake and when to steer to avoid a collision. *International Journal of Vehicle Safety 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.

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*, 142–152. https://doi.org/10.1016/j.aap.2016.11.009.

Coelingh, E., Jakobsson, L., Lind, H., & Lindman, M. (2007). Collision warning with auto brake: a real-life safety perspective. In *Innovations for safety: Opportunities and challenges.*

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

Dobberstein, J., Schmidt, D., Östling, M., Lindman, M., Bálint, A., & Höschele, P. (2021). Final results on detailed crash configurations from collisions expected to remain for automated vehicles, OSCCAR D1.3. www.osccarproject.eu.

Fahrenkrog, F., Wang, L., Platzer, T., Fries, A., Raisch, F., & Kompaß, K. (2019). Prospective safety effectiveness assessment of automated driving functions—From the methods to the results. *26th ESV conference*.

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.

Guo, F., Klauer, S. G., McGill, M. T., & Dingus, T. A. (2010). Evaluating the Relationship Between Near-Crashes and Crashes: Can Near-Crashes Serve as a Surrogate Safety Metric for Crashes? http://www.nhtsa.gov/DOT/NHTSA/NVS/Human%20Factors/Safety%20Problem%20Identification/DOT%20HS%20811%20382.pdf.

Hallerbach, S. (2020). Simulation-based testing of cooperative and automated vehicles. https://oops.uni-oldenburg.de/id/eprint/4649.

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

Helmer, T. (2014). *Development of a methodology for the evaluation of active safety using the example of preventive pedestrian protection*. Springer.

ISO (2021). ISO/TC 22/SC 36/WG 7, ISO/TR 21934-1 Road vehicles—Prospective safety performance assessment of precrash technology by virtual simulation—Part 1: State-of-the-art and general method overview.

Krajewski, R., Bock, J., Kloeker, L., & Eckstein, L. (2018). The highD dataset: A drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems. In *IEEE conference on intelligent transportation systems*. https://doi.org/10.1109/ITSC.2018.8569552.

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

Kullgren, A., Krafft, M., Lie, A., & Tingvall, C. (2007). The effect of whiplash protection systems in real-life crashes and their correlation to consumer crash test programmes. In *Proceedings of the 20th international conference on Enhanced Safety of Vehicles (ESV)*, pp. 1–7. https://www-esv.nhtsa.dot.gov/Proceedings/20/07-0468-O.pdf.

Kullgren, A., Stigson, H., & Krafft, M. (2013). Development of whiplash associated disorders for male and female car occupants in cars launched since the 80s in different impact directions. In *Proceedings of IRCOBI conference* (pp. 51–62).

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.

Laureshyn, A. (2010). *Application of automated video analysis to road user behavior*. Lund University (Doctoral dissertation) (Issue 253).

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*, 29–37. https://doi.org/10.1016/j.aap.2017.10.009.

Lindman, M., & Tivesten, E. (2006). A method for estimating the benefit of autonomous braking systems using traffic accident data. *SAE Technical Papers, 2006*(724). https://doi.org/10.4271/2006-01-0473.

Markkula, G. (2015). Driver behavior models for evaluating automotive active safety: From neural dynamics to vehicle dynamics. https://research.chalmers.se/en/publication/212952.

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.

McLaughlin, S. B., Hankey, J. M., & Dingus, T. A. (2008). A method for evaluating collision avoidance systems using naturalistic driving data. *Accident Analysis and Prevention, 40*(1), 8–16. https://doi.org/10.1016/j.aap.2007.03.016.

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

NHTSA (2006). Event Data Recorders (EDRs).

NHTSA (2019). Traffic safety facts 2017: A compilation of motor vehicle crash data, pp. 24.

Otte, D., Krettek, C., Brunner, H., & Zwipp, H. (2003). Scientific approach and methodology of a new in-depth- investigation study in Germany so called GIDAS. *Proceedings of the 18th international conference on Enhanced Safety of Vehicles (ESV)* http://www-esv.nhtsa.dot.gov/Proceedings/18/18ESV-000161.pdf.

Rosén, E. (2013). Autonomous emergency braking for vulnerable road users. In *Proceedings of IRCOBI conference* (pp. 618–627).

Šajn, N. (2019). *General safety of vehicles and protection of vulnerable road user (Issue April 2019)*. European Parliament | Members' Research Service.

Sander, U. (2018). *Predicting Safety Benefits of Automated Emergency Braking at Intersections Virtual simulations based on real-world accident data*. Chalmers University of Technology https://research.chalmers.se/publication/504728.

Sander, U., & Lubbe, N. (2016). Prediction of accident evolution by diversification of influence factors in computer simulation: Opportunities for driver warnings in intersection accidents. *Aktive Sicherheit Und Automatisiertes Fahren - Methodenentwicklung Im Expertendialog, 29*.

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

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. *Waymo LLC*. https://storage.googleapis.com/waymo-uploads/files/documents/Waymo-Simulated-Driving-Behavior-in-Reconstructed-Collisions.pdf.

Scanlon, J. M., Kusano, K. D., & Gabler, H. C. (2015). Analysis of driver evasive maneuvering prior to intersection crashes using event data recorders. *Traffic Injury Prevention, 16*, S182–S189. https://doi.org/10.1080/15389588.2015.1066500.

Scanlon, J. M., Page, K., Sherony, R., & Gabler, H. C. (2016). Using event data recorders from real-world crashes to evaluate the vehicle detection capability of an intersection advanced driver assistance system. *SAE technical paper, 2016-01–1457*. https://doi.org/10.4271/2016-01-1457.

Schubert, A., Erbsmehl, C., & Hannawald, L. (2013). Standardized pre-crash-scenarios in digital format on the basis of the vufo simulation. 5th International Conference on ESAR "Expert Symposium on Accident Research".

Shaikh, J., & Sander, U. (2018). Creation of pre-crash time-series database for evaluation of active safety systems using RASSI data. *8th international conference on ESAR "Expert Symposium on Accident Research"*.

Smith, K., Thome, A., Blåberg, C., & Bärgman, J. (2009). An invariant may drive the decision to encroach at unsignalized intersections. *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*.

Spitzhüttl, F., Petzold, M., & Liers, H. (2015). PCM from iGLAD database. FAT Forschungsvereinigung Automobiltechnik.

Summala, H. (2007). Towards understanding motivational and emotional factors in driver behavior: Comfort through satisficing. In C. P. Cacciabue (Ed.), *Modelling driver behavior in automotive environments* (pp. 189–207). Springer.

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

Tanaka, S. (2015). Reconstructing accidents by simulation for developing active safety systems. In *Proceedings of the 3rd international symposium on future active safety technology towards zero traffic accidents* (pp. 597–603).

Thomson, R., Sandin, J., Bagdadi, O., Hjort, M., Augusto, B., & Andersson, H. (2013). EDR pre-crash data: Potential for applications in active safety testing. *Proceedings of the 23th international conference on Enhanced Safety of Vehicles (ESV)* https://www-esv.nhtsa.dot.gov/Proceedings/23/files/23ESV-000414.PDF.

UNECE (2019). Event Data Recorder (EDR) GRSG-116 Agenda item 17. April, pp. 3.

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

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. *SHRP 2 report S2–S08A-RW-1*.

VTTI (2020). InSight data access website SHRP2 naturalistic driving study.

Liers, H. (2018) Traffic Accident Research in Germany and the German In-Depth Accident Study (GIDAS). Presented at ACMA-SIAM-VDA Conference On Safer & Sustainable Road Transportation, New Delhi, India. Available at: https://www.vufo.de/wp-content/uploads/2021/02/181120_SIAM_ACMA_VDA_Conference_Liers_GIDAS-Presentation.pdf.

VUFO (2020). PRE-CRASH-MATRIX (PCM) V5.0 FORMAT. https://www.vufo.de/pcm/?lang=en.

WHO (2018). Global status report on road safety 2018. https://www.who.int/publications-detail/global-status-report-on-road-safety-2018.

Woodrooffe, J., Blower, D., Bao, S., Bogard, S., Flannagan, C., Green, P. E., & LeBlanc, D. (2012). Performance characterization and safety effectiveness estimates of forward collision avoidance and mitigation systems for medium/heavy commercial vehicles.

Pierluigi Olleja received the B.Sc. and M.Sc. degree in automotive engineering from Politecnico di Torino, Italy, in 2018 and 2020, respectively. During his M.Sc. he spent one year in Sweden at Chalmers University of Technology in a study exchange programme, where he did his master's thesis at the Division of Vehicle Safety. From 2019 to 2021, he worked at the Division of Vehicle Safety as a project assistant in the Crash Analysis and Prevention group. In July 2021 he started his studies as a PhD candidate. His research focus is counterfactual simulations and driver behaviour modelling.

Jonas Bärgman is an associate professor at Chalmers University of Technology in Gothenburg, Sweden. He started his career in industry at Autoliv Research in 1997, but he has worked as a researcher at Chalmers since 2009. His research focuses on methods for virtual safety assessment of human behavior and in-vehicle technologies in traffic. His work include research into a range of sub-domains needed for virtual assessment, including development computational driver models and crash causation research needed for such development, as well as methodological aspects of integration of driver models in virtual safety assessment.

Nils Lubbe received his PhD degree in Engineering from Chalmers University of Technology. He started his career in vehicle safety engineering at Toyota Motor Europe and is currently Director of Research at Autoliv in Vårgårda, Sweden. His research interests include crash analysis and prevention as well as injury biomechanics.