

HARMONIZED APPROACHES FOR BASELINE CREATION IN PROSPECTIVE SAFETY PERFORMANCE ASSESSMENT OF DRIVING AUTOMATION SYSTEMS

Peter Wimmer

Virtual Vehicle Research GmbH
Austria

Olaf Op den Camp

TNO
The Netherlands

Hendrik Weber

Institute for Automotive Engineering (ika), RWTH Aachen University
Germany

Henri Chajmowicz

LAB
France

Michael Wagner

Continental Autonomous Mobility Germany GmbH
Germany

Jorge Lorente Mallada

Toyota Motor Europe NV/SA
Belgium

Felix Fahrenkrog

BMW AG
Germany

Florian Denk

THI - CARISSMA Institute of Automated Driving
Germany

Paper Number 23-0032

ABSTRACT

In the last years, virtual simulations have become an indispensable tool for safety performance assessment of driving automation systems (DAS) and pre-crash technologies which are part of advanced driver assistance systems (ADAS). Different approaches and tools are used in this domain, making comparison of results of different studies difficult. Therefore, the P.E.A.R.S. (Prospective Effectiveness Assessment of Road Safety) initiative was founded to harmonize methods for prospective safety performance assessment and by this make results of such studies more trustworthy and comparable. One essential pillar of such a harmonization is the establishment of the baseline, the set of data to which the performance of the technology under study is compared to when performing prospective assessments. Various ways have been presented in literature for setting up a baseline. For harmonization, these ways need to be analyzed and categorized so that recommendations can be given on when and how to use a certain baseline approach. The research objective of this paper is first to develop general approaches to establish a baseline based on existing ways and second to identify areas of application for each baseline approach.

Based on existing ways, we defined general approaches for setting up a simulation baseline. These baseline approaches can structure all existing ways based on their characteristics and requirements and impacts on safety performance assessment results. Relevant information for each baseline approach is discussed, such as the used data type(s), data processing steps, applied variations to the original data, application of simulation models, and statistical methods, etc.

We identified three types of baseline approaches: A) Using concrete real-world scenarios without modifications. B) Using modifications of concrete real-world scenarios. Here, real-world scenarios are the basis, but some of the existing measured properties are altered or even new properties are added. C) Creating synthetic cases where more general data such as distributions of relevant parameters (e.g., from collision, road user behavior, traffic data) and mechanisms possibly leading to collisions are used. The paper will provide examples for each baseline approach.

The three approaches can be clearly distinguished and should be able to cover the generation of a baseline for all studies in the field of prospective safety performance assessment. Each of the approaches has its pros and cons, e.g., with respect to their representativeness, and the effort to obtain the required data. Also, the evaluation objective to be addressed needs to be considered when selecting an appropriate baseline approach as it has a strong influence on this selection. The categorization of the three approaches allows for defining common recommendations on when to use which approach.

By the baseline approaches presented, P.E.A.R.S. contributes to the harmonization and acceptance of virtual safety performance assessment of driving automation systems (DAS) and pre-crash technologies. This will greatly enhance trustworthiness, comparability and, transparency of results of prospective safety performance assessments.

INTRODUCTION

Trends in the number of traffic casualties in the EU [1] show that it will be difficult to meet the target of Vision Zero: reducing road fatalities to almost zero by 2050 [2]. At the same time, the road traffic system is changing rapidly due to, e.g., the introduction of new mobility systems such as connected, cooperative, automated driving and new enabling technologies such as artificial intelligence and wireless V2X-communication [3]. Driving Automation Systems (DAS), including vehicle safety features such as advanced driver assistance systems (ADAS), are introduced with the intention to make road mobility services safe for all road users. Moreover, DAS and ADAS are implemented to make road mobility services available, more comfortable, and safe for drivers and passengers.

Authorities are being asked to allow vehicles equipped with new advanced DAS onto public roads. However, an appropriate methodology for approval to deploy these vehicles onto the road is not yet in place. The EC formulated recommendations [4] to ensure that DAS contribute to road safety improvements. This example shows that there is a clear need for a prospective safety assessment framework that is capable to deal with the great challenges and fast developments in technology. To keep a feasible testing effort, an increasing role of virtual testing is foreseen to handle the seemingly infinite number of situations that DAS may end up into during the lifetime of the vehicle. Although DAS and ADAS are complex and the safety assessment procedure can be complicated, its results should be unambiguous, easily understood by experts in the field, and explainable to policymakers and the general public.

The EU Horizon 2020 project HEADSTART [5] defined testing and validation procedures for the safety of Connected, Cooperative, and Automated driving functions for specific use cases. HEADSTART set important requirements for the use of simulations to test and validate DAS. The recently started EU Horizon Europe project V4SAFETY [6] uses the HEADSTART requirements as starting point to develop comprehensive procedures for conducting computer simulations to determine the long-term performance and impact of road safety solutions, from the identification and collection of the relevant input data to the projection of the results to a region of interest (e.g., the EU).

Harmonization of the assessment framework is essential to achieve explainable and comparable results, independent of the specific simulation tool used. Moreover, it is to the benefit of all stakeholders that the developed safety assessment framework not only conforms with European Union [7] and United Nations regulations [8] but with international standards such as ISO and SAE as well. The lack of harmonization of prospective assessment was already recognized in 2012, which led to the establishment of a harmonization group: Prospective Effectiveness Assessment for Road Safety (P.E.A.R.S. [9]). In the last decade, the P.E.A.R.S. harmonization group, currently consisting of 31 organizations from industry, research organizations, and academia, provided input to the ISO working group “Traffic accident analysis methodology”, resulting in the publication of an ISO Technical Report [10]. P.E.A.R.S. is continuing its work in drafting an ISO Technical Specification “Prospective safety performance assessment of pre-crash technology by virtual simulation — Guidelines for application” [11](under development). This document will provide a general description of the process for prospective safety performance assessment of pre-crash technology by virtual simulation.

All stakeholders in road safety indicate the need [5] for a predictive safety assessment framework that allows fast and extensive evaluation of safety solutions, including DAS, for a large variety of relevant traffic scenarios. This is already envisaged by policy makers and consumer associations [7], [12], considered in state-of-the-art research activities worldwide [5], [13] and foreseen to be adopted by industry partners to manage testing efforts [14].

The framework uses predictive virtual simulation (hereafter simply 'simulation') in addition to the already existing physical tests—not only for type approval but for consumer testing as well. In safety assessment of a vehicle function (e.g., a newly developed DAS), the performance of the vehicle with the function in a set of relevant traffic scenarios is compared to the performance of the vehicle without the function in the same set of traffic scenarios. The simulations of the set of scenarios for the vehicle without the function under test is called the *baseline*. The selection of relevant, realistic baseline scenarios is of the utmost importance for the quality of the assessment and the results. However, there is little discussion in literature on how to define such a baseline.

The goal of a predictive effectiveness assessment is to make a reliable prediction of the effect a DAS has on traffic safety. The first step for executing such an assessment lies within formulating the evaluation objective, which defines the overall scope of the study [10]. The overall scope of the assessment can be outlined by a safety solution. Within the automotive industry, many of today’s safety solutions come from DAS that target crash avoidance but can also consider other in-vehicle systems that potentially increase safety. Moreover, safety solutions can also be represented outside the individual vehicle by either changes in infrastructure or policy decisions (e.g. decreased speed limits). The remainder of this paper will focus on the evaluation of DAS, still the methods can be directly transferred to other safety solutions.

An overview of common relevant terms within in the scope of predictive assessments is given in **Table 1**.

*Table 1.
Relevant terms related to prospective safety assessment*

Baseline	set of cases to which the performance of the technology under study is compared to when performing prospective assessments
Case	set of specified conditions used as input for the assessment, generally based on concrete scenarios
Simulation model	a computational model which allows the virtual evaluation of the technology, process, or behavior it represents. A simulation model can also contain other simulation models.
Penetration rate	number of vehicles of a certain type equipped with the activated technology under assessment compared to the total number of vehicles of that type in a certain geographic area over a certain period of time.
Research question	question that a research project is designed to answer
Scenario	description of the traffic, infrastructure, and environmental conditions (for example weather and lighting conditions) for the simulation that consists of a sequence of scenes
Scenario category	selection of scenarios that share one or more characteristics

From the initial outline of the evaluation objective, the objective needs to be defined more precisely, by formulating a research question. Formulating this research question requires the consideration of multiple aspects. The assessment should be executed for a defined collection of scenario categories. Herein, each scenario category represents a selection of scenarios that share one or more characteristics. Limitations to the scope of the assessment should be given - e.g., which weather conditions are considered. Moreover, the metric for the evaluation should be stated. A typical metric is the number of crashes avoided by the applied solution. Surrogate measures which may be considered are metrics describing the criticality of a traffic situation.

An example of a precisely formulated research question is: *What is the safety performance of an AEB (warning + autonomous intervention) at a penetration rate of 100% in rear-end car-to-car crashes while approaching an intersection on urban roads in terms of avoided crashes related to the situation in Europe in 2021?*

From the definition of the evaluation objective, the actual assessment can be executed. For this, it is necessary to create a baseline for the assessment, which describes the scenarios to be analyzed without the technology under assessment. This forms the starting point for the simulation with the technology under assessment (treatment). The baseline needs to be created to match the evaluation objective and scope of the assessment. It should represent all relevant elements of the scenarios that potentially have an influence on the performance of the

safety solution. The baseline scenarios need to be representative for the safety situation of the baseline condition in the comparison. Therefore, it is necessary to derive the cases from data representing the safety situation under comparison. In the latter, individual concrete scenarios will be referred to as cases.

The choice of data and the process of converting the data to cases as input for the prospective assessment has a large influence on the overall result of the assessment. Differences in the data processing for baseline generation may cause two studies to be incomparable, even though safety solution and models are the same. If baseline approaches can be aligned across different studies, the foundation for a comparison of safety solutions is laid. When creating baseline cases, it needs to be considered, which data is available as input.

Various ways have been presented in literature for setting up a baseline (e.g., in [15], [16], [17], [18], [19], [13], [20], [21], [22], [23], [24], [25]). For harmonization, these ways need to be analyzed and categorized so that recommendations can be given on when and how to use a certain baseline approach. The research objective of this paper is first to develop general approaches to establish a baseline based on existing ways and second to identify areas of application for each baseline approach. All authors of this paper are active in the P.E.A.R.S. consortium [6]. The method presented represents an important common ground of members within P.E.A.R.S., which will lay the foundation for any further harmonization of predictive safety assessment methods, also within the scope of the V4SAFETY Project [9].

In the following, a high-level categorization will be presented, that allows comparison of ways how data may be used for baseline generation. The categorization allows collecting the most suitable baseline approach for the intended assessment. The choice of baseline approach depends heavily on the evaluation objective, the data that is available and the safety solution. For some research questions, such as studies regarding systems which only become active immediately before the crash, cases can be constructed which are close to real-world crashes. Depending on the intended comparison defined by the research question, original cases need to be adapted to a certain extent to enable the intended comparison. After an introduction of the different baseline approaches, a recommendation is given, when which approach is most suitable.

METHOD

In order to come up with different baseline approaches, we propose a method which is based on a high-level categorization taking into account the type of input data source used, how the input data source is used and the processing of the data itself. These categories are structured in different layers, as it can be seen in the schematic view of the method which is shown in **Figure 1**.

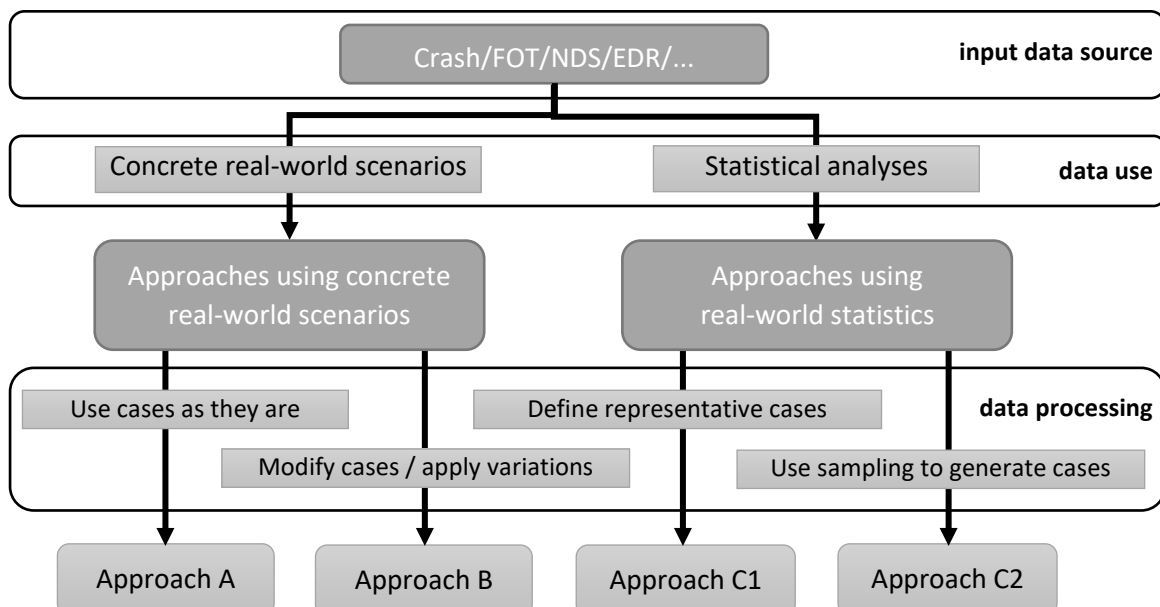


Figure 1. Categorization of the different approaches¹.

¹ FOT: field operational test, NDS: naturalistic driving study, EDR: event data recording

Input data source

Generally, input data is associated with reconstructed real-world crashes or with real-world distributions of pre-crash and normal driving conditions [26]. Such input data is needed to generate the baseline scenarios and it can come from one or various data sources, related to crash data, driving data (e.g., FOT, NDS, EDR), experiment data coming from studies in controlled field (e.g., driving simulator, test track) or other sources. It is necessary to understand the type of information that such data contains:

- Crash data contains crash information, considering that a crash consists of any contact with an object, either moving or fixed, at any speed in which kinetic energy is measurably transferred or dissipated. It includes other vehicles, roadside barriers, objects on or off the roadway, pedestrians, cyclists, or animals [27]. Crash data is generally stored in databases (either as aggregated data or case by case), which contain information from real-world crashes gathered by different means, such as detailed investigation teams, police reports, insurance companies, expert reports, or hospital data. Some examples of crash databases are, GIDAS (German In-Depth Accident Study) [28], NASS [29], CARE [30], ITARDA [31], VOIESUR [32], BAAC [33].
- Driving data contains information of real-world traffic situations, covering the range from nominal driving data up to critical data (i.e., near-crash and in-crash). This data may be relevant to derive real-world distributions that can be used to define synthetic scenarios. There are multiple approaches to collect driving data and depending on the source of the recording, it is possible to differentiate:
 - o Data from in-vehicle data collection: Either using existing components and sensors from production vehicles or adding additional ones, data is recorded from a vehicle-perspective. Some examples are naturalistic driving studies (NDS) or field operational tests (FOT) [34], [35], [36].
 - o Data collected from infrastructure sensors or sensors at a fixed location: Using sensors which are installed under existing or dedicated infrastructure elements, data around a pre-defined area is recorded. Some examples are highway cameras (e.g. focusing on traffic flow measurements), or dedicated cameras at intersections [37]. Recently, the use of drones for driving data collection is also applied [38], providing the flexibility that the sensors do not require to be installed on fixed infrastructure elements, although the data collection approach is similar.
- Experiment data from studies in controlled field encompasses data coming from test track, driving simulator studies or similar experimental data which is gathered in a controlled environment.
- Other: Further data sources may be considered, such as reports, studies, scientific articles which may provide necessary information to build a baseline, as well as complementary sources, such as weather or traffic flow data.

Regardless of the input data type selected, the user performing the prospective safety performance assessment study shall be aware of the details of the data (e.g., quality, data sampling procedure, representativeness), and shall document the used input data sources to ensure transparency of the assessment. This documentation shall contain not only information on the input data used, but also on the selection process of the data, which shall be described as well.

Data use

Once the main input data type has been selected, it is necessary to decide whether to use data directly as it is, or if it is going to be analyzed to derive statistical information from it, such as real-world distributions of relevant parameters. Depending on the selected approach, a distinction can be made between:

- Concrete real-world cases: This approach consists in using the data as it is, so a link exists between the original real-world case and the input data used (e.g., a concrete crash from a crash database is used as input without altering the original recorded case, or a near crash scenario recorded during driving is replayed in the simulation). In some cases, in this step a selection of cases, based on inclusion and/or exclusion criteria may be applied. One example may be to use a group of cases of a crash database [15].
- Results from statistical analysis to determine the baseline: This approach consists in using real-world data as a source for a statistical analysis on the crash/driving data to provide statistics of certain parameters. In a second step, these are used to derive (real-world based) synthetic cases: cases are generated by using distributions of parameter values instead of recorded or reconstructed values and choosing plausible (physically) combinations thereof [24], [13], [21]. In this approach there is only an in-direct link between the synthetic case derived and the original input data.

Data processing

Based on the selected input data and its use, the next step is to consider if any data processing step needs to be applied on the data or not. We distinguish the following options:

For the use of concrete real-world cases:

- Use cases as they are: Collected real-world scenarios are digitized and used as baseline cases. No further modifications or assumptions are applied, besides the ones related to how the data has been collected [15], [17].
- Modify cases / apply variations: The real-world case is the reference, but modifications or variations are applied to it. A distinction is made between:
 - o Modification of real-world cases: The complemented information (e.g. speed) can come from an individual value (e.g., posted speed limit) or from distributions of data (e.g., accident data analysis)[18].
 - o Variation of the original real-world cases: in this case, a variation of parameter values from the real-world case is applied [20], [19].

Although information is added or parameters are varied, there is a strong link with the original real-world cases.

For the usage of statistical data to derive synthetic cases, we distinguish two possibilities:

- Define representative cases: In this case, statistical data from various sources is considered to build a limited set of representative cases [21], [22].
- Use sampling to generate cases: The approach consists in generating a (usually) large number of cases following a sampling scheme, in order to cover the whole range of relevant cases [24], [13].

In the above-mentioned approaches, no direct link exists between the derived synthetic cases and the original data.

As shown in Figure 1, depending on the input data source used, the use of this data and its processing, we identified four different baseline approaches: approach A, approach B, approach C1 and approach C2. A detailed description of the approaches follows in the next section “Results”.

RESULTS

As an output of the previously described method, three main types of baseline creation approaches can be distinguished. A graphical overview of the different approaches is given in **Figure 2**, a detailed description follows below.

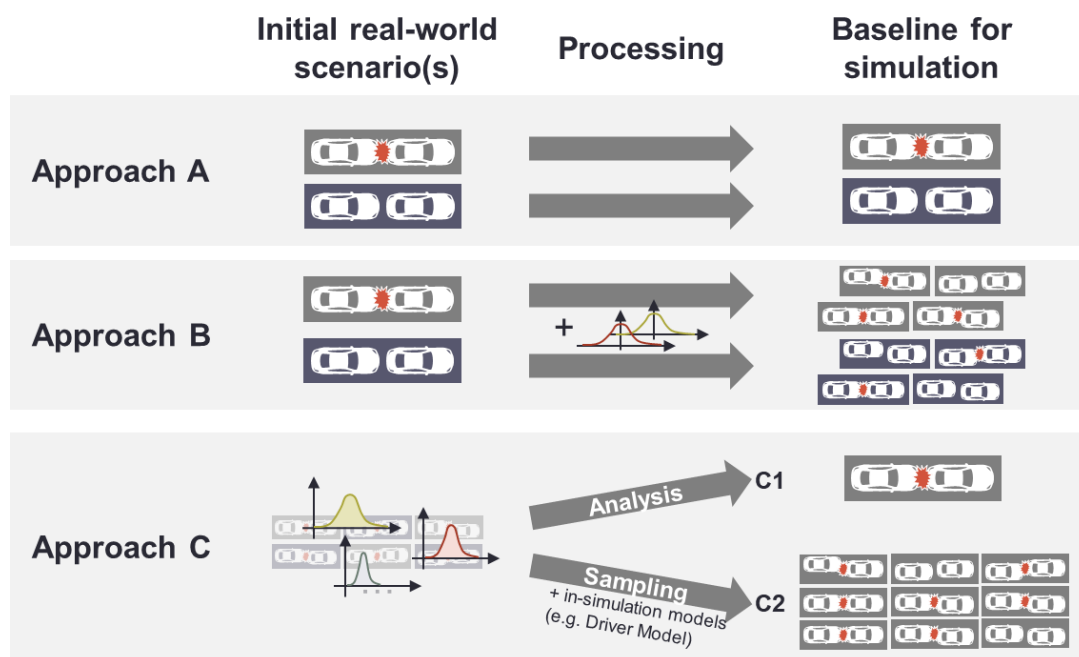


Figure 2. The three baseline approaches, image taken from [11](under development).

Approach A: Digitized real-world scenarios without modification

In this approach, individual real-world scenarios are digitized and used as baseline cases without altering them. The data sources are usually databases consisting of recorded driving data or reconstructed crashes. From these data sources, time series of the dynamic elements (vehicle under test (VuT), surrounding traffic) as well as positions of relevant infrastructural elements and information such as weather, lighting conditions etc. are used to set up each test case, replicating the real-world situation. Although the digitized cases and the possibly resulting crash consequences could in principle be taken directly as baseline results, the authors strongly recommend performing the following steps during safety performance assessment:

- Use the time-series data for the dynamic elements as input for re-simulations in the to be used simulation tool. The results of these re-simulations are the baseline results.
- For the treatment, repeat the baseline re-simulations but with the technology under assessment present. The results are treatment simulation results.
- Derive the differences between both results in a case-by-case analysis as basis for the safety performance assessment.

By this, it is ensured that the differences in the safety performance only result from the influence of technology under assessment and these differences are not artifacts caused by using different methods for obtaining baseline and treatment results.

Approach B: Modification or variation of real-world scenarios

In this approach, real-world scenarios are the basis as well, but modifications to the original data for building the required baseline are made in order to alter existing properties or even to add new ones, e.g., to be able to use an older or less complete dataset and to modify it towards the current-state-of-the-art.

Data sources are usually databases consisting of recorded driving data or reconstructed crashes. In contrast to the previous approach, in this approach the data from the database is enriched by:

- Adding parameters to compensate for unavailable information: The reason for the addition can be missing information from the original real-world case that is needed to define the baseline (e.g., missing speed information). Another aim of this approach is to update the existing data to a specific state of technology. This can be done by using additional in-simulation models (e.g., driver models or technology models such as ABS) and re-simulating the original scenario. In this way, an updated version of an original real-world scenario is generated.
- Modifying existing parameters to compensate for uncertain information: This can be achieved by, for example, adding variations of known parameters to create multiple variants of one single real-world scenario.

In this approach, the number of baseline cases is equal to or higher than the number of real-world scenarios considered in the set-up. The number of treatment simulations is the same as the number of baseline simulations.

Approach C: Baseline consisting of synthetic cases

In this approach statistical information of real-world data is used instead of individual real-world scenarios. This information is used to determine distributions of traffic relevant parameters such as speed, time headway, braking behavior etc. With these distributions synthetic cases can be set up which represent what happens in the real world. These synthetic cases consist of trajectories of all traffic participants of interest for the specific case.

Two variations in the application of approach C are distinguished:

- Approach C1: Here the scenario statistics are analyzed to generate a limited set of test cases that are representative for the scenarios that the function under test will encounter in the real world.
- Approach C2: Here the statistical information can be used to set up a usually large number of cases that cover not only the most typical situations but also the rare combinations of parameters (large test space). The generation of cases for the second variant of this approach could be done using sampling techniques or by using a simulation including models that describe human behavior, such as driver models or pedestrian behavior models. The chosen sampling method will determine the number of test cases that result from this approach. Another distinction of this second variant is that here not necessarily the same cases need to be re-simulated in the treatment simulations. Even the number of simulations for baseline and treatment simulation can be different as long as it is ensured that the number of simulations is large enough to ensure a stable result.

Although C1 and C2 differ, they are grouped in approach C as the main difference is only the number of cases, not the methodology. Approaches A, B, and C are distinguished by the use of a different methodology.

Examples for the different baseline approaches

Existing, published ways of setting up a baseline can be attributed to one of these approaches. Following are descriptions of examples for each of the approaches.

Approach A

- In [15], an AEB-Pedestrian was evaluated by virtual simulation of accidents selected from the French accident database, VOIESUR[32]. The database gathers all fatal accidents and a sample of 5% of all injury accidents that occurred in France in 2011, proportionally distributed over the whole French mainland territory. The database is weighted to represent accident severity, involved user type, conflict type and location (by proxy) as described in the national census BAAC[33]. The accident subset used for virtual simulation consists of accidents in which a pedestrian (any age) was hit by the front of a passenger car and pre-crash trajectory / impact speed was available or could be reconstructed from e.g., projection distances contained in reports. Exclusions were: car in loss of control situation, side swipes, pedestrians lying on the road prior to impact and suicides. The final weighted sample consists of 197 fatally injured, 1863 severely injured and 3103 slightly injured pedestrians with injury severity scale of the Police (not AIS).
- SIMPATO (Safety IMPact Assessment Tool)[16]: The SIMPATO tool was used in the EU-funded project “interactIVe” that developed active safety systems for multiple conflict types. The SIMPATO tool focuses on those conflict types that were addressed by most of the interactIVe systems, namely rear-end and run-off conflicts. For the rear-end conflicts, 364 real-world crashes of the GIDAS database have been analyzed. The simulations have been conducted for systems that warned the driver and/or reacted by means of braking or evasive maneuver. The models for the systems’ reaction were derived from the interactIVe track tests. For the run-off road conflict, 150 GIDAS accidents were considered. Here, the interactIVe system reaction was always a steering maneuver.

Approach B

- Reference [18] shows an example for adding missing information according to Approach B. The main input data type used is real-world crash data, from a police reports accident database in Germany. The data contains information on the accident conflict situation, collision configuration, geo-coordinates of accident, participants involved as well as injury level of each participant, among other variables. The database does not contain time series information such as trajectories and speed. Trajectory information is added based on the description of the police report. Speed profile information is added based on a statistical analysis of an in-depth accident database (GIDAS), based on participant maneuver, accident location, participant type and injury level sustained. Driving speed, collision speed and deceleration value are extracted to define the speed profile (mean values are assigned for each parameter). Based on the added information, the simulation files can be created. Plausibility checks are done to confirm that the collision is realistic, and that the collision configuration is as reported in the police report.
- An example for adding missing information and creating variations is given in [19]. The main input data type used is real-world crash data, from a police reports accident database in Germany. The data contains information on the accident conflict situation, collision configuration, geo-coordinates of accidents, participants involved as well as injury level of each participant, among other variables. The database does not contain time series information such as trajectories and speed. Trajectory information is added based on the description of the police report. Speed profile information is added based on a statistical analysis of another accident database (GIDAS). The statistical analysis provides 3 variations of 3 parameters (driving speed, collision speed and deceleration value). Considering the 2 participants involved in the accident, a maximum of 729 variations can be generated per each accident. Only the simulation files that confirm a collision exists are considered as plausible data.
- In the L3Pilot safety impact assessment [13] the DAS developed in the project were assessed. The project covered different types of DAS, namely a motorway, an urban and parking DAS. For the motorway DAS, two of the baseline approaches were used. Approach B was used in the counterfactual simulations of rear-end and cut-in conflicts. The real-world cases came from the dataset involving crashes with Volvos (VCTAD [39]), crashes from the Traffic Accident Scenario Community (TASC [18]) database and critical situations from the SHRP2 database[36]. By means of the critical driving scenarios the false positive behavior of the DAS was assessed.

Approach C1

- The CATS project (2014 - 2016) [21] provided a proposal for a test matrix towards Euro NCAP for the testing and safety rating of Autonomous Emergency Braking Systems onboard passenger cars that are capable to avoid or mitigate collisions with cyclists. By studying car-to-cyclist accidents in the EU, obtained from data of France, Germany, Italy, the Netherlands, Sweden, and the United Kingdom, the

five most common scenarios for accidents between passenger cars and cyclists were selected. These scenarios describe the trajectories and maneuvers of cyclists and cars for several seconds up to the moment of impact. Next step was to construct test scenarios for the three most dominant accident scenarios. An in-depth study into the accidents was conducted to determine the most relevant parameters and the most relevant ranges of these parameters. Additional input was collected from observation studies that were conducted on specific locations in the Netherlands where many interactions (without collisions) between cyclists and passenger cars were observed. These studies revealed the influence on the cyclist and vehicle speed in an approach of an intersection in the presence of a strong view-blocking obstruction. Based on the accidentology and test parameter studies described above, the set of baseline tests has been proposed. Relevant and realistic test cases were provided based on statistical analysis of thousands of accidents.

Approach C2

- L3Pilot safety impact assessment [13]: Next to the counterfactual simulation described above, which applies baseline approach B, L3Pilot used also approach C2 to assess the safety impact of automated driving. The C2 approach had been applied in typical crash scenario as well as scenarios that pose a challenge for the DAS. For the motorway the following scenarios were considered: lane change conflict, conflict with VRU, minimum risk maneuver, wrong activation, end of lane, obstacle in the lane, lower speed limit and passing a motorway entrance. The number of analyzed cases per scenario varied depending on the considered infrastructure and traffic parameters. Overall, more than 25 000 cases were simulated. The C2 approach was also applied for the urban DAS. Here, all scenarios were generated with a stochastic sampling approach using copulas, which was presented in [40]. Input to the generation of the simulation cases were different sources including accident data, traffic data and data from L3Pilot pilot studies.
- In [23], the effectiveness of a pedestrian protection system implemented in a vehicle was studied. The analyzed scenario was a pedestrian crossing a street unauthorized at an unprotected location. To this end, warning, automated emergency braking and a combination of both were evaluated for varying parametrizations of the algorithm. The approach C2 was employed since both the traffic on the street as well as the pedestrian crossing the street were simulated. The goal of the simulation was to replicate the risk in the described scenario as precisely as possible. Hence, not every simulation resulted in an accident. In order to resolve statistically significant differences in virtual accident numbers, 18 million crossings were simulated in the baseline and 100 million crossings were simulated for the treatment.
- In [25], the effect of a simplified automated driving function with / without external information by an infrastructure-based LiDAR sensor is analyzed. The specific scenario that is studied is a right-turn scenario: cars are turning right and have to yield to straight going cyclists. Thereby, occlusion due to parked cars and a construction site was present. The approach C2 is used in this publication since the authors replicate the crash causation mechanisms in the described scenario and use traffic simulation to create scenarios. In order to resolve an effect size of 10%, 200 million cyclist crossings were simulated in the baseline and for each of the 3 different levels of treatment.

DISCUSSION AND LIMITATIONS

Each of the presented baseline approaches has its advantages and disadvantages, e.g., with respect to the power to generalize to the overall population (“representativeness”) or the demand on the required data. In the following these aspects are discussed.

Approach A has the advantage, that it is based on real-world driving or crash scenarios. Compared to the other approaches, it is relatively easy to derive the baseline from these unaltered scenarios. Basically, the baseline cases can be used directly without any change. The approach might require a conversion of the real-world cases into the required format. However, the scenario should not be changed in terms of trajectories of the involved traffic participants. In the simulation, it needs to be ensured that the traffic participants follow the original real-world trajectory. This is typically done via a so-called trajectory following model. The approach does not need any complex driver behavior models.

The question how likely it is that reconstructed characteristics of an investigated accident will happen again in a comparable manner, can be argued. But the investigated cases represent realistic crash configurations. In this context, we distinguish between cases that are measured (typically NDS or FOT data) and cases that are derived from reconstruction (typically in detailed accident databases). In the latter case the quality and number of available variables of the reconstruction defines how well the case represents the reality and its applicability to

investigate safety effects. It is especially difficult to reconstruct the cognitive state of the involved drivers. However, the effect of an ADAS system is often sensitive to these variables.

An often-encountered issue is that there are not enough real-world cases to derive statistically meaningful results. In contrast to the approaches B and C, the number of cases cannot be increased above the number of cases in the used database. This issue is in particular relevant when simulating cases for which crash data is used. In case driving scenarios are derived from NDS or FOT, this issue might be less relevant. However, NDS and FOT data contain typically only a very low number of crashes – if they include crashes at all. This need to be accounted when doing the assessment.

For approach A, it must be considered that the covered time frame of cases – in particular for crash cases – is typically limited (e.g., cases in GIDAS PCM database cover up to 5 s). Therefore, it must be checked whether the effect of the technology under assessment can be evaluated appropriately in this time frame. If the time period of the case is insufficient, one can switch to another approach (e.g., B or C2). Approach B allows to make changes to the original scenarios, which allows also to extend the scenario. However, this extension in approach B will be limited to a couple of seconds since the link to the original case needs still to be given. Approach C2 offers in terms of simulation time more freedom, since typically only the start conditions need to be defined. However, crash mechanisms from the real-world must be still represented correctly in the baseline.

Approach B combines the realism of approach A with the flexibility to adapt the case to specific needs. The needs could have different facets. This could be for instance to consider in the baseline additional technologies to make the cases more representative for today's traffic, to complete missing parameters of the used databases, to enlarge the test space in safety assessment or to consider variations to account for possible shortcomings in the reconstruction of cases. Due to the close relation with approach A, most of the advantages and disadvantages apply also for this approach. For instance, also for approach B the number of suitable real-world cases might be quite low. By variation, the number of simulated cases can be increased. However, the representativity would not be changed, since the number of original baseline cases would stay the same.

The possibility of variation offers quite some degrees of freedom for the assessment. Therefore, approach B can be applied for many different evaluation objectives. In this sense it can be seen as an evolution of approach A. However, it must be considered that the degrees of freedom come with the challenge of ensuring the cases resulting from variations are plausible and representative. For instance, if the driver reaction is changed from the original baseline case, it must be checked how probable this variation is under the traffic conditions in the baseline case to assure adequate weighting in the statistical analysis of the data. The variation of the driver behavior would also mean that the trajectory following model is not sufficient any longer and a more sophisticated model is required.

Baseline approaches C1 and C2 are also quite different from the approaches A and B. Since they rely only on distributions sampled from real-world data and the understanding of crash causation mechanisms and not the concrete real-world data time series, the number of baseline cases can be chosen arbitrarily. Both C approaches (C1 and C2) mainly differ in the number of considered cases for the analysis. While C1 investigates a very limited number of cases, C2 assesses typically a quite high number of cases. Dealing with a high number of cases is typically less a challenge in a virtual assessment than in real-world testing. Thus, it is for C2 much easier to reach a sufficient number of cases for a statistically sound comparison between baseline and treatment. This leads for the virtual assessment often to the choice of C2, since it allows to cover a large scenario space. Nevertheless, there are few evaluation objectives, in which C1 is the choice for the assessment. One example is the round-robin simulation of P.E.A.R.S. [9], in which the difference between several simulation tools in the same simulated scenario is investigated. But it must be considered that this study did not investigate the safety performance of technology. Other evaluation objectives in which C1 would be useful are comparisons of virtual simulation with real-world tests, for which a high number of tests would increase the effort heavily, or in case the simulations of cases are very heavy on computational effort.

The sampling from parameter distributions – if done in a sufficient manner and resulting in much more cases than available in the real-world data source – allows also for a wide coverage of the scenario space ranging from crash via critical scenarios to normal driving. In general, all baseline approaches can cover critical scenarios that both did and did not lead to collision. Collision cases aim at investigating of true positive behavior (cases which required an activation by the technology and in which the technology became active) and false negative behavior (cases which required an activation, but the technology did not become active). By means of non-collision cases false positive (cases which did not require an activation, but in which the technology became active) as well as true negative behavior (cases which did not require an activation and the technology was not activated) can be analyzed. As discussed, exemplary in [23], the rate of false positive activations influences the effectiveness of ADAS systems, since a high number of unnecessary warnings lead to the deactivation of the

system by the driver, which clearly undermines the intended positive effects of the system. With respect to the baseline approaches, in the approach C1 and C2 non-collision case can be derived even from databases that contain only collision cases, while approach A and B required the consideration of data source that contains non-collision case. Furthermore, for C2 the non-collision cases are derived implicitly by the number of considered cases and use generation process. For the other baseline approaches A, B and C1, the decision about consideration of non-collision cases in the analysis needs to be made explicitly (e.g. by choosing the real-world case or in the generation).

For C1 and C2 the duration of the simulated cases can also be chosen, which provides an advantage for investigations of a technology that intervenes into the driving dynamics of the vehicles for longer time periods. Surrounding traffic participants can also be considered in baseline approach C1 and C2, although it must be noted that each traffic participant increases the effort and complexity of the simulation. This is in particular of relevance for more complex technology, like e.g. automated driving: the reaction of the technology might depend on the surrounding traffic and vice versa, the flow of the surrounding traffic might depend on the reaction of the technology [41]. But also for ADAS this could be of relevance, e.g. when checking for secondary effects like does the AEB braking lead to more rear-end collisions with the following traffic. For approach B the consideration of surrounding traffic as one variation parameter is also feasible but increases the requirements for the simulation models that represent the traffic in this approach quite heavily. For approach A this is not feasible.

The approaches C1 and in particular C2 offers a very high degree of freedom. However, these opportunities do not come without challenges. For this approach it is vital to understand the mechanisms leading to a crash and the underlying parameter and distributions to ensure that the simulated case represent the real-world cases sufficiently. This requires a deep understanding of the used data as well as of the simulation models, especially whether they reproduce the underlying crash causation mechanisms. A key model for this approach is the driver behavior model, since it decides how critical a case is going to be and what the resulting crashes variables are. A simple trajectory following model is not enough for this approach. Rather a sophisticated driver model is required here. Thus, the big question when using the approach results from the fact that there is no direct link to original real-world cases: does the simulation produce realistic cases? To answer this, an increased effort for the validation and verification of the simulation models and the scenario is required compared to the other approaches. Moreover, this approach requires a high amount of several input data for generating the input data's distributions, which establish the link to the real-world scenarios.

The evaluation objective to be addressed needs to be considered when selecting an appropriate baseline approach as it has a strong influence on this selection. **Table 2** provides exemplary research questions in which the authors would apply a certain approach. The common theme is car-to-car rear-end collisions in an urban environment. The choice of one research question does not mean that no other approach would be suitable. But other approaches were not the preferred option by the authors.

*Table 2.
Exemplary possible research questions for the different baseline approaches*

Baseline Approach	Exemplary Research Question	Rationale
A	What is the safety performance of an AEB (warning + autonomous intervention) at a penetration rate of 100% in car-to-car crashes on urban roads in terms of MAIS 2+ injuries related to the situation in Germany from 2015 - 2017 as represented in GIDAS PCM?	Approach A is preferred since due to the specific naming of the country, time frame and database to be used. The question implies rather to investigate the performance in particular cases than in the general safety performance of an AEB.
B	What is the safety performance of an AEB (warning + autonomous intervention) at a penetration rate of 100% in car-to-car crashes on urban roads in terms of MAIS 2+ injuries related to the situation in Germany from 2010 - 2017 as represented in GIDAS PCM while considering only ESC-	Approach B is preferred since a direct link to certain crash data is desired (GIDAS-PCM) similar as in the previous example. However, here an altering of the baseline cases is required for those cases in which the vehicle was not equipped with ESC. Now the vehicle needs to be equipped with AEB. The second aspect that hints towards approach B is the consideration of different road frictions. This

	equipped vehicles and considering different road friction?	could also be easier achieved in B by varying the baseline than in approach A, since the different road frictions will not be presented equally, i.e., it is likely that the database contains too few cases with low road frictions.
C1	What is the safety performance of a VRU AEB (warning + autonomous intervention) in car-to-car crashes as defined in the Euro NCAP protocol?	This research question requires only a few simulations, and consequently approach C1 has been chosen. However, the main task is rather to get to the representative crash cases. This step has been done by experts of Euro NCAP when defining the test protocols.
C2	What is the safety performance of an AEB (warning + autonomous intervention) at a penetration rate of 100% in rear-end car to car crashes while approaching an intersection on urban roads in terms of avoided crashes related to the situation in Europe in 2017?	For this research question, Approach C2 is recommended as an insufficient number of real-world crashes is likely to exist in databases. For some countries a reasonable number of crashes in databases might exist. However, for other countries this is not the case. Distributions that are describing the general traffic behavior can be used to generate such conflict cases for different countries.

CONCLUSIONS

This paper presents a methodological analysis of different ways to set up a baseline for prospective safety performance assessment. We found three main elements of the set-up process: input data source selection, data use, and data processing. We distinguish three main approaches for setting up a baseline depending on the choices taken in each of these elements. These approaches should cover any baseline set-up process, some examples from literature are presented in the results section.

The various ways presented in literature for setting up a baseline for prospective safety performance assessment can be attributed to one of these approaches with this methodology. This will help to understand what has been done in past studies and increases comparability and trustworthiness of past and future studies in this field.

Moreover, the paper discusses the advantages and disadvantages of the different approaches as well as the dependency of the approach selection on the evaluation objective of a safety performance assessment study, the data that is available and the safety solution. This will help the readers in the selection of a suitable baseline approach for future studies.

With this work, the authors and P.E.A.R.S. as a whole contribute to the harmonization and acceptance of virtual safety performance assessments of DAS and ADAS. This will greatly enhance trustworthiness, comparability and, transparency of results of such assessments. Furthermore, these baseline set-up approaches will be part of [11] (under development).

REFERENCES

- [1] European Commission. 2021a. *European Road Safety Observatory Annual statistical report on road safety in the EU 2021*.
- [2] European Commission. 2020b. *Sustainable and Smart Mobility Strategy – putting European transport on track for the future*.
- [3] Piao, J., & McDonald, M. 2008c. "Advanced driver assistance systems from autonomous to cooperative approach". *Transport Reviews*, 28(5), 659–684. <https://doi.org/10.1080/01441640801987825>
- [4] Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659). 2020d. *Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility*.
- [5] Sakura, SIP-adus, & HEADSTART. 2021e. *Towards the harmonization of safety assessment methods of automated driving*. https://www.headstart-project.eu/wp-content/uploads/2021/12/211209_HEADSTARTSAKURA_WP_Final-draft.pdf

- [6] V4SAFETY. 2022f. *Vehicles and VRU Virtual eValuation of Road Safety*. EU Horizon Europe Funded Project. <https://www.linkedin.com/company/v4safety/>
- [7] European Commission. 2022g. "Commission Implementing Regulation (EU) 2022/1426, uniform procedures and technical specifications for the type-approval of the automated driving system (ADS) of fully automated vehicles". *Official Journal of the European Union*, L221, 1–64.
- [8] United Nations. 2021h. *Regulation No 157, uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping Systems [2021/389]*.
- [9] Wimmer, P., Düring, M., Chajmowicz, H., Granum, F., King, J., Kolk, H., Op Den Camp, O., Scognamiglio, P., & Wagner, M. 2019i. "Toward harmonizing prospective effectiveness assessment for road safety: Comparing tools in standard test case simulations". *Traffic Injury Prevention*, 20(Supplement 1: Peer-Reviewed Journal for the 26th International Technical Conference on the Enhanced Safety of Vehicles (ESV)), 139–145. <https://doi.org/10.1080/15389588.2019.1616086>
- [10] ISO. 2021j. *ISO/TR 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>
- [11] ISO. n.d.-k. *ISO/AWI TS 21934-2: Road vehicles — Prospective safety performance assessment of pre-crash technology by virtual simulation — Part 2: Guidelines for application (under development)*. <https://www.iso.org/standard/81790.html>
- [12] Euro NCAP. 2017l. *Euro NCAP 2030 Roadmap*. <https://cdn.euroncap.com/media/74190/euro-ncap-vision-2030.pdf>
- [13] Bjorvatn, A., Page, Y., Fahrenkrog, F., Weber, H., Aittoniemi, E., Heum, P., Lehtonen, E., Silla, A., & Bärghman, J. 2021m. *L3Pilot 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
- [14] Harnarayan, S. S. M.-M. P. T. 2021n. "Simulation Based Virtual Testing for Safety of ADAS Algorithms". In SAE International (Ed.), *SAE WCX Digital Summit*.
- [15] Saadé, J., Chajmowicz, H., & Cuny, S. 2019o. "Prospective Evaluation of the Effectiveness of Autonomous Emergency Braking Systems in Increasing Pedestrian Road Safety in France.". *IRCOBI*. <http://www.ircobi.org/wordpress/downloads/irc19/pdf-files/37.pdf>
- [16] Van Noort, M., Bakri, T., Fahrenkrog, F., & Dobberstein, J. 2015p. "SIMPATO-The safety impact assessment tool of interactive". *IEEE Intelligent Transportation Systems Magazine*, 7(1), 80–90. <https://doi.org/10.1109/MITS.2014.2340054>
- [17] Hasegawa, T., Takahashi, H., & Udaka, S. 2017q. "Clarification of Priority Factors for Reducing Traffic Accident of Fatalities in the U.S. and Benefit Estimation of AEB System for Oncoming Vehicles". *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. <https://www-esv.nhtsa.dot.gov/Proceedings/25/25ESV-000171.pdf>
- [18] Urban, M., Erbsmehl, C., Lorente Mallada, J., Puente Guillen, P., & Tanigushi, S. 2020r. "A Methodology for Building Simulation Files from Police Recorded Accident Data (for ADAS Effectiveness Assessment)". *FISITA World Congress*. <https://www.fisita.com/library/fisita-world-congress/2020/f2020-pif-039>
- [19] Urban, M., Erbsmehl, C., Landgraf, T., Pohle, M., Lorente Mallada, J., Puente Guillen, P., & Tanigushi, S. 2020s. "Parameterization of Standard Test Scenarios of Automated Vehicles Using Accident Simulation Data". *FISITA World Congress*. <https://www.fisita.com/library/fisita-world-congress/2020/f2020-pif-038>
- [20] Leledakis, A., Lindman, M., Östh, J., Wågström, L., Davidsson, J., & Jakobsson, L. 2021t. "A method for predicting crash configurations using counterfactual simulations and real-world data". *Accident Analysis & Prevention*, 150, 105932. <https://doi.org/10.1016/J.AAP.2020.105932>
- [21] Op den Camp, O., van Montfort, S., Uittenbogaard, J., & Welten, J. 2017u. "Cyclist Target and Test Setup for Evaluation of Cyclist-Autonomous Emergency Braking". *International Journal of Automotive Technology*, 18(6), 1085–1097. <https://doi.org/10.1007/s12239>

- [22] Stoll, J., Schneider, A., Wisch, M., Seiniger, P., & Schaller, T. 2016v. *PROSPECT D3.1-The addressed VRU scenarios within PROSPECT and associated test catalogue*. <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5a90e757e&appId=PPGMS>
- [23] Helmer, T. 2014w. *Development of a methodology for the evaluation of active safety using the example of preventive pedestrian protection* [TU Berlin]. <https://depositonce.tu-berlin.de/handle/11303/4270>
- [24] Wang, L., Fahrenkrog, F., Vogt, T., Jung, O., & Kates, R. 2017x. "Prospective safety assessment of highly automated driving functions using stochastic traffic simulation". *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, 1–13. <https://www-esv.nhtsa.dot.gov/Proceedings/25/25ESV-000224.pdf>
- [25] Denk, F., Brunner, P., Huber, W., Margreiter, M., Bogenberger, K., & Kates, R. 2022y. "Assessment of traffic safety interventions using virtual randomized controlled trials: potential of connected and automated driving including V2X for collision reduction at urban intersections". *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 1183–1190. <https://doi.org/10.1109/ITSC55140.2022.9921764>
- [26] Alvarez, S., Page, Y., Sander, U., Fahrenkrog, F., Helmer, T., Jung, O., Hermitte, T., Duering, M., Doering, S., & Op den Camp, O. 2017z. "Prospective Effectiveness Assessment of ADAS and Active Safety Systems via Virtual Simulation: A Review of the Current Practices". *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. <https://www-esv.nhtsa.dot.gov/Proceedings/25/25ESV-000346.pdf>
- [27] Guo, F., Klauer, S., McGill, M., & Dingus, T. 2010aa. *Evaluating the Relationship Between Near-Crashes and Crashes: Can Near-Crashes Serve as a Surrogate Safety Metric for Crashes?* https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/dot_hs_811_382.pdf
- [28] GIDAS. n.d.-ab. <https://www.gidas.org/start-en.html>
- [29] NHTSA. n.d.-ac. NASS. <https://www.nhtsa.gov/crash-data-systems/national-automotive-sampling-system>
- [30] European Commission. n.d.-ad. CARE. https://road-safety.transport.ec.europa.eu/statistics-and-analysis/methodology-and-research/care-database_en
- [31] ITARDA. n.d.-ae. ITARDA. <https://www.itarda.or.jp/english/>
- [32] Lesire, P., Léopold, F., Chauvel, C., Hervé, V., & Cuny, S. 2015af. "Implication of Children in Road Accidents in France in 2011". *24th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*.
- [33] Amoros, E., Lardy, A., Wu, D., Viallon, V., & Martin, J.-L. 2015ag. *Projet VOIESUR - Livrable L3 - Méthodologie : redressement et extrapolation, projet VOIESUR - ANR*. <https://hal.archives-ouvertes.fr/hal-01212490>
- [34] Innamaa, S., Aittoniemi, E., Bjorvatn, A., Fahrenkrog, F., Gwehenberger, J., Lehtonen, E., Louw, T., Malin, F., Penttinen, M., Schindhelm, R., Silla, A., Weber, H., Borrack, M., Lillo, L. Di, Merat, N., Metz, B., Page, Y., Shi, E., & Sintonen, H. 2020ah. *L3Pilot D3.4-Evaluation Plan*. https://l3pilot.eu/fileadmin/user_upload/Downloads/Deliverables/Update_07102021/L3Pilot-SP3-D3.4-Evaluation_plan-v1.0_for_website.pdf
- [35] Eenik, R., Barnard, Y., Baumann, M., Augros, X., & Utesch, F. 2014ai. "UDRIVE: the European naturalistic driving study". *Transport Research Arena*.
- [36] Hedlund, J. 2015aj. "SHRP 2 naturalistic driving study: The first data analysis studies". *Journal of Safety Research*, 54, 3–4.
- [37] Nakamura, H., Muslim, H., Kato, R., Préfontaine-Watanabe, S., Nakamura, H., Kaneko, H., Imanaga, H., Antona-Makoshi, J., Kitajima, S., Uchida, N., Kitahara, E., Ozawa, K., & Taniguchi, S. 2022ak. "Defining Reasonably Foreseeable Parameter Ranges Using Real-World Traffic Data for Scenario-Based Safety Assessment of Automated Vehicles". *IEEE Access*, 10, 37743–37760. <https://doi.org/10.1109/ACCESS.2022.3162601>

- [38] Krajewski, R., Bock, J., Kloeker, L., & Eckstein, L. 2018al. "The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems". In IEEE (Ed.), *21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 2118–2125). <https://doi.org/10.1109/ITSC.2018.8569552>
- [39] Isaksson-Hellman, I., & Norin, H. 2005am. "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://pubmed.ncbi.nlm.nih.gov/16179140/>
- [40] Rösener, C. 2020an. *A Traffic-based Method for Safety Impact Assessment of Road Vehicle Automation* [RWTH Aachen University]. <https://doi.org/10.18154/RWTH-2020-08950>
- [41] Helmer, T., Kompaß, K., Wang, L., Kühbeck, T., & Kates, R. 2017ao. "Safety Performance Assessment of Assisted and Automated Driving in Traffic: Simulation as Knowledge Synthesis". In *Automated Driving* (pp. 473–494). Springer International Publishing. https://doi.org/10.1007/978-3-319-31895-0_20