ORIGINAL PAPER

Positive risk balance: a comprehensive framework to ensure vehicle safety

Nina Kauffmann[1] · Felix Fahrenkrog[1] · Ludwig Drees[1] · Florian Raisch[1]

Abstract

The introduction of automated vehicles promises an increase in traffic safety. Prior to its launch proof of the anticipated reduction in the sense of a positive risk balance compared with human driving performance is required from various stakeholders such as the European Union Commission, the German Ethic Commission, and the ISO TR 4804. To meet this requirement and to generate acceptance by the public and the regulatory authorities, a qualitative Risk- Benefit framework has been defined. This framework is based on literature research on approaches applied in other disciplines. This report depicts the framework, adapted from the pharmaceutical sector called PROACT-URL which serves as a structured procedure to demonstrate a positive risk balance in an understandable and transparent manner. The qualitative framework needs to be turned in quantitative methods once it should be applied. Therefore, two steps of the framework are discussed in more detail: First, the definition of adequate development thresholds that are required at an early stage of the development. Second the simulation-based assessment to prove the positive risk balance prior to the market introduction.

Keywords Positive risk balance · Safety threshold · PrOACT-URL · Safety assessment by simulation · Accident statistics · Safety performance

Introduction

The introduction of automated driving systems (ADS) promises a variety of improvements with regard to passenger comfort, traffic flow, emission reduction and road safety (Kyriakidis et al., 2015; Payre et al., 2014; Rödel et al., 2014). Current crash statistics indicate human error contributes to 93% of road collisions, which originate in temporary distraction, decreased reaction time, limited perception of the environment or emotional reactions in traffic situations (Petridou & Moustaki, 2000). The introduction of ADS has great potential to rule out human errors.

Vehicles with level 3 automation (SAE, 2016) offer on selected road types the opportunity to take over the driving task, allowing passengers to take their hands off the wheel and engage in so-called non-driving related tasks such as eating and reading (Hancock et al., 1999; NHTSA, 2013). Taking over the driving task by the system represents a major challenge. This is related to those aspects in which attentive drivers have their strengths, such as situational awareness and prediction of future outcomes (Endsley, 1995) as well as recognizing and interpreting communication patterns (Färber, 2015).

These aspects of the ADS concern various aspects of machine perception. The ADF's main challenges are the perception (Dietmayer, 2016) and the human factors (HF) aspects which range from the system interaction with the driver when it comes to take-over from and to the driver, designing a take-over request (TOR) (Merat & De Waard, 2014) or the system`s reaction if the driver fails to respond and executes a minimal risk maneuver (MRM).

In 2016, the German government established an Ethics Commission to address legal and ethical issues in automated driving. In its 2017 published report, it requests to prove that an ADS will cause fewer collisions compared to the human driver, i.e. a positive risk balance (PoRiBa) (Fabio et al., 2017). The importance of decreasing the accident risk by means of ADS was also emphasized in the report *Ethics of Connected and Automated Vehicles* published by the Commission Expert Group of the European Union in 2020 (Bonnefon et al., 2020).

✉ Nina Kauffmann
 Nina.Kauffmann@bmw.de

[1] BMW Group, Knorrstraße 147, 80788 Munich, Germany

An initial response to these requirements by industry represents the White Paper *Safety first for automated driving* (Wood et al., 2019). Here, the authors propose safety-by-design and verification-and-validation (V&V) methods of ADS of SAE level 3 and level 4 (SAE, 2016) to demonstrate a PoRiBa. In 2020, the UNECE released the regulation No. 157 for automated lane keeping system (ALKS) to drive automated up to 60 kph. Regarding safety, it states that the driver as well as the surrounding traffic participants should not be exposed to a higher risk with the introduction of ADS. This aspect is highlighted by the requirement to compare the ADS to a competent and careful driver, which should be the baseline for not explicitly scenarios addressed by the ALKS (UNECE, 2020). In the same year, ISO published the technical report 4804, which describes steps for the development and validation of ADS based on fundamental safety principles derived from globally applicable publications (ISO TR 4804, 2020).

This document aims as a supplement to existing standards and publications to provide a technical implementation to achieve a PoRiBa of ADS throughout the entire development process.

Definition of PoRiBa

The starting point of the PoRiBa framework is the 2nd requirement of the above-mentioned German Ethics Commission: "The objective is to reduce the level of harm until it is completely prevented. The licensing of automated systems is not justifiable unless it promises to produce at least a diminution in harm compared with human driving, in other words a positive balance of risks."(Fabio et al., 2017, p.10). For the practical application of this requirement, it needs to be operationalized. This led to the development of framework that should not only provide the result of a PoRiBa, but also show how the result was achieved. The objective is to create acceptance by society and authorities.

Based on the publication of the Blumenthal et al. (2020), the PoRiBa can be divided into three categories.

A. Safety as a measurement Safety as a measurement refers to the quantitative and qualitative methods used to demonstrate the PoRiBa. Blumenthal et al. (2020) distinguishes between lagging and leading measures. Lagging measures are for instance current crash statistics that give clues to current human driving performance, and leading measures are indicators such as the analysis of driving behavior before a collision has occurred.

B. Safety as a threshold Safety as a threshold can be expressed in a qualitative (e.g. the risk acceptance criteria ALARP (as low as reasonably practicable) or in a quantitative manner (e.g. Minimum Endogenous Mortality). An overview of risk acceptance principles is given by Kron (2004). Setting thresholds is necessary for the development of automated vehicles to demonstrate a PoRiBa.

C. Safety as a process Safety as a process refers to the establishment of a safety culture in the company. Standards and processes are defined in such a way that the information-, product- and reporting obligation of the company as well as the proof of the standards (e.g. ISO 26262, ISO 21448) can be guaranteed.

It is obvious that it would not be enough to determine the risk balance at the end of the development and hope for the best. The PoRiBa frame work needs rather to cover the entire development process starting from the concept to the release of the product (see Fig. 1). Therefore, in the following section the principles of the framework are explained (chapter 3). Afterwards the consideration of the PoRiBa for two stages are described (concept phase in "Quantitative risk balance" and its final prove in the release phase in "Assessing the safety performance of an ADS"). The development and 'in operation' phases are not explained, since in those phases the consideration of PoRiBa is very much related to safety procedures like functional safety, safety of the intended functionality, cybersecurity, and field observation.

Fig. 1 Definition of a positive risk balance (PoRiBa) (blue: relevant aspect in this paper)

Development of a framework for PoRiBa

To define the PoRiBa framework two steps are necessary. First, to define requirements of the framework. Here, a detailed analysis of the entire report of the German Ethics Committee (Fabio et al., 2017) was conducted. Based on this analysis, the framework shall consider and cover the following aspects:

- public's demand for transparent information about new technologies and their use,
- the many unknown factors in the approval process of automated systems,
- combining quantitative and qualitative methods,
- aspects regarding the monetary cost of incidents shall not be included.
- the complexity of the decisions.

The second step was literature review on similar frameworks in other fields and branches, like railway transportation (DIN EN 50129, 2019; DIN EN 50126, 2018) or the aviation industry (International Civil Aviation Organization, 2008), which have already been established and accepted by regulatory authorities. To find the best possible model, we compared the identified frameworks with the previously defined requirements and exclusion criteria. A schematic overview of the procedure is given in Fig. 2.

Many of these frameworks were not suitable because they do not meet the criteria defined above. For example, some frameworks are only applicable in their defined context and therefore cannot be applied to our decision-making process. These include the TR Strab Brandschutz

(2014) in the area of structural fire protection in traffic facilities and the DIN EN ISO 14971 (2020) for medical devices. Other frameworks show deficiencies in the transparency of the decision making (FDA, 2018) or consider the monetary cost of incidents in risk analysis (Hazardous Substances: REACH Authorization, Medical devices: DIN EN ISO 14971, 2020). While automated systems have been approved in the aviation industry for many years, the nature of automation is not at all comparable to automotive automation. In airplanes, the pilot will continue to monitor the automation, i.e. autopilot systems. In automobiles, the driver is expected to take the place of the "passenger" in the long term and will thus be removed from the control-feedback loop (Banks et al., 2019). Besides, most people are not pilots and rarely fly. By car, on the other hand, most people drive regularly. Due to that, problems are clearly more obvious, which makes the demands on the transparency of the process much higher when introducing ADS. Therefore, the model of the aviation industry does not fully meet the requirements.

High consistency of the defined criteria could be detected with the requirements for the approval of pharmaceuticals in Europe. Similarities regarding the magnitude of uncertainties and risks were identified in the pharma and automobile sectors. Both deal with a limited amount of data in advance. Conventional pharmaceutical approaches generally include a stepwise implementation including preclinical animal studies at inception which are followed by studies on human subjects in several phases. The automotive industry evaluates relevant traffic scenarios based on driving simulator studies or real traffic scenarios with human test subjects. After launching new drugs or in the automotive sector, a new function, field observation plays a crucial role. In both industries
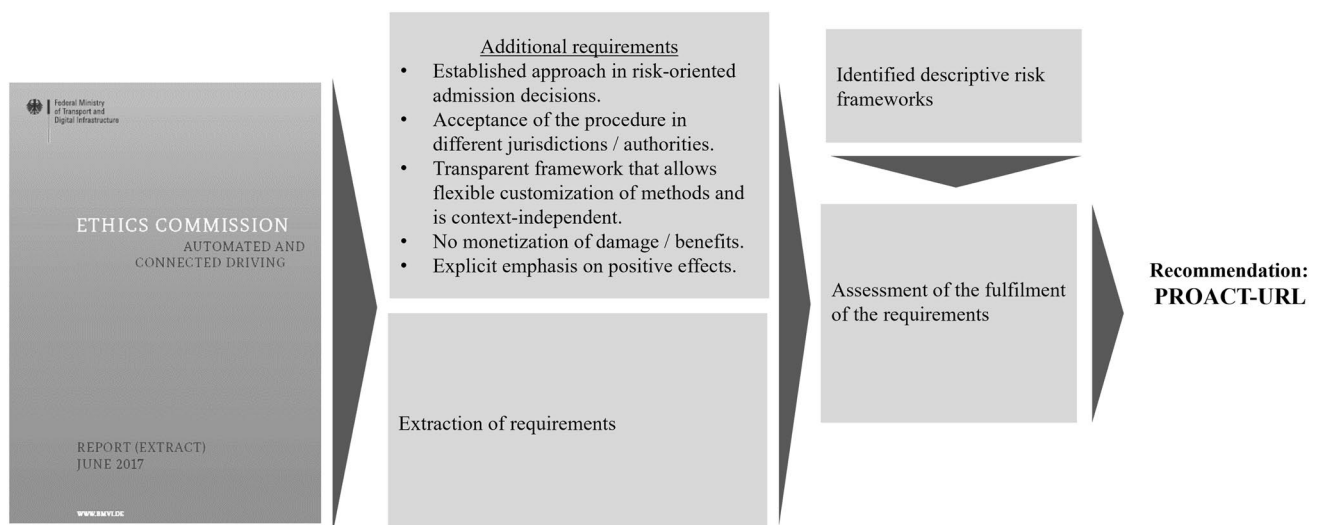

Fig. 2 Procedure of identifying framework

customers are consulted by using different data collecting methods. The automotive industry uses for example field operational tests, customer studies or if necessary, accident research or crash investigation. In the pharmaceutical branch doctors, patients or pharmacies are obliged to report the occurrence of rare side effects to the European Medicines Agency (EMA). Both industries also need to inform their relevant stakeholders about favorable and unfavorable effects. This information is accessible in the patient information leaflet of drugs and concerning the automotive sector in the user manual of a function. Due to the strong similarity of requirements, corresponding approaches applied in the pharma sector were investigated further.

The basic approach to drug approval is described in more detail in a research project funded by the EMA. Based on an exemplary initial approval and subsequent withdrawal of approval of a drug, a systematic procedure for the creation and maintenance of a risk balance sheet was developed as a benchmark for the approval. Using the initial letters of the individual process steps, this approach is referred to as PrOACT-URL (EMA, n.d.).

Risk balance method according to PrOACT-URL

Within the European Union, there exist several empirical verification strategies for determining the positive effect of medicinal products. The PrOACT-URL method was developed with the aim of creating a modern drug monitoring and approval system that is both patient-centered and relevant as well as accountable from a societal perspective. The second objective of PROACT-URL is to strengthen the monitoring of the risk–benefit balance of medicines in Europe. To achieve this overall goal, PrOACT-URL was designed as a comprehensive and integrated framework that aims to develop and validate tools and methods. The PrOACT-URL procedure provides eight stages of risk–benefit assessment. The following Fig. 3 shows the different process steps and their description (EMA, n.d.; Hunink et al., 2001):

Procedure for risk balancing of highly automated driving

Then, the PrOACT-URL framework was adapted for the automotive industry. The procedure of a PoRiBa is depicted as a control loop. The presented loop shows a possible adaption of the framework by additionally extending the process with a product observation step, because product observation after start of production (SOP) is a key element for the PoRiBa of ADS over lifetime. The individual steps can be assigned to the manufacturer of the vehicle on the one hand and to the approval authority on the other, as shown in Fig. 4.

Problem	Description of the problem and the setting (e.g. Involved stakeholders, target group, type of problem, outcome, affected time period)
Objectives	Specification of objectives Criteria definition for target evaluation
Alternatives	Elaboration of available options (pre and post approval)
Consequences	Definition of impact (e.g. severity, incidence or magnitude) for the individual options
Trade-offs	Evaluation of the options favorable and unfavorable effects
Uncertainty	Elaboration and assessment of quantitative and qualitative uncertainties with regard to individual framework steps Evaluation of associated impact on risk balance
Risk Tolerance	Determination of the stakeholders (e.g. the regulatory authority) risk attidtue towards the product
Linked decisions	Consistent decision making with regard to past or comparable decisions as well as potential future decisions

Fig. 3 Process steps of PrOACT URL (EMA, n.d.)

Fig. 4 Basic concept of the risk balancing procedure for the automotive industry
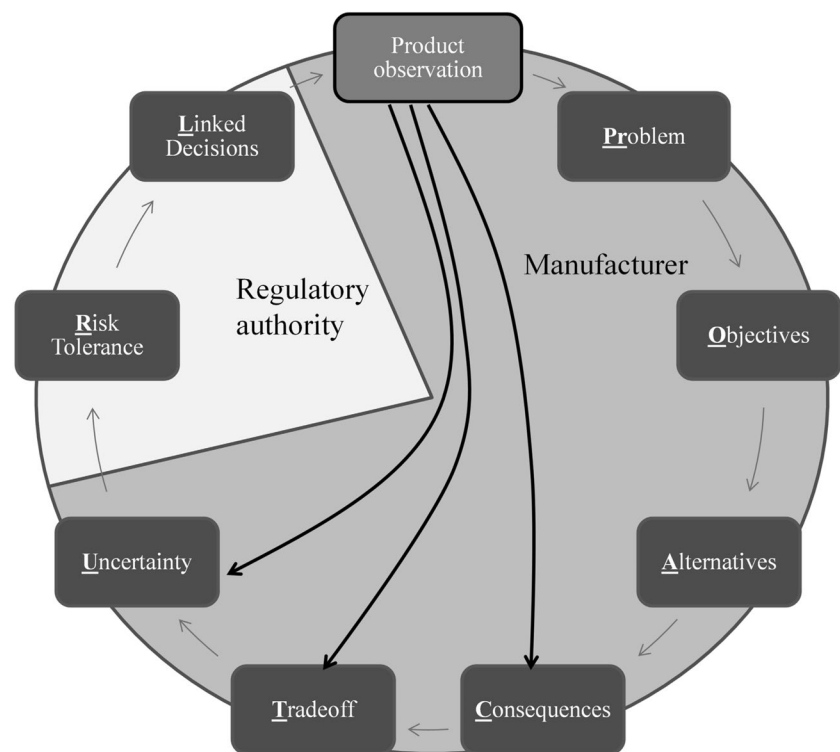
A new aspect of the procedure for the automotive industry is the control loop concept for dealing with uncertainties in vehicle functions. This gives the possibility of subsequent improvement after the vehicle has been placed on the market as part of the release and approval decision.

Problem Here a new driving function is to be introduced, which aims at reducing the number of collisions. This step consists in a comprehensive description of the operational design domain (ODD), which includes a description of the vehicle in operation, for example a description of road types, environmental conditions, and other constraints (NHTSA, 2016), collisions on which a system can have a direct or indirect impact (is referred to functional field of application (FFoA)) and a description of the affected target group.

Objective Presentation of objectives that indicate the overall purposes to be achieved and the development of criteria against which alternatives can be evaluated.

1. the ADS yield improved safety performance compared with the human driver
2. the risks avoided by ADS exceed the risks caused by ADS.

The objectives also include the description of favorable and unfavorable effects.

Alternatives Display of absence of ADS in terms of safety performance considering associated favorable and unfavorable effects.

Consequences Assessment of the impact of the different options for instance concerning the ADS function in comparison with human driving performance.

Trade Off Assessment of the balance between favorable and unfavorable effects by different stakeholders, for instance through the establishment of a company internal committee formed by representatives of different departments and expertise incorporating diverse perspectives.

Uncertainty Reporting uncertainties of qualitative and quantitative types at every step of the process. For example, this includes the assessment of the quality of crash data or accuracy of simulation model. The balance between favorable and unfavorable effects due to uncertainty is also considered at every step as well as the extent to which the benefit-risk balance is reduced by including all sources of uncertainty, to provide a benefit-risk balance, and the reasons for the reduction.

It is now possible to make a release recommendation or to identify improvements that are necessary, e.g., in the function, the verification strategy, or the argumentation chain. A transparent and comprehensive documentation of an initial risk balance is handed over to the authorities.

Risk tolerance and linked decision concern the regulatory authorities

They consider the consistency of this decision with comparable decisions taken in the past and assess whether taking this decision could impact future decisions either favorably or unfavorably (e.g., would it set a precedent or make similar decisions in the future easier or more difficult) (Krumbach & Schnieder, 2019).

Quantitative Risk Balance

In accordance with the traditional V-model as it is defined in ISO 26262 (ISO26262, 2018) the implementation of the PoRiBa at system level is described in the following. Chapter 4 gives the definition of safety thresholds as a basis for the development of ADS as it is required at an early development stage (see step *objective* of PrOACT-URL).

Safety metric

The PoRiBa compares the safety performance SP of automated vehicles SP_{AV} with the safety performance of the human driver SP_{HD}. Hence, a safety metric and a corresponding threshold value is required. In terms of a PoRiBa, the following equation must hold.

$$SP_{AV} > SP_{HD} \tag{1}$$

But what is a suitable safety metric to measure the safety performance? We propose to use the average distance d between two collisions as a safety metric. In general, the safety performance can be then calculated as follows:

$$SP = \frac{m}{n_{Collision}} \tag{2}$$

m denotes the annual mileage and $n_{Collision}$ refers to the annual number of crashes, e.g. of a human driver. Other metrics, such as crashes per hour (crash rates) can be derived accordingly based on d. For example, the crash rate results from the inverse of d, i.e. $1/d$.

In the following, a method is described to quantify the safety performance of human drivers SP_{HD} based on accident data. Evaluating the national accident statistics, for example Destatis for Germany (Destatis, 2019) reveals the occurrence of various types of collisions (e.g. single truck collisions, collisions between passenger vehicles or between passenger vehicles and pedestrians). Our focus is on automated passenger cars, we propose to count collisions only, if a passenger car is involved (e.g. passenger car vs truck, single collision of a passenger car). Collisions, where no passenger car was

involved are not relevant to this analysis. Therefore, it is important to only include collisions that correspond to the type of ADS and vehicle in question. Likewise, if other vehicles are considered, e.g. automated trucks, only collisions where a truck was involved shall be counted.

However, one may argue that in some cases, a passenger car is involved in a collision without being at fault, for example, at the end of a traffic jam. A common crash scenario is a rear-end collision with a truck. In these kinds of crashes, even an automated vehicle barely has any chance to avoid the collision if the automated vehicle was rear-ended by someone else. Similarly, as mentioned above, such crashes are independent regarding automation. In this sense, it may not be reasonable to include these types of collisions ($n_{Collision}$). Although the rear-end example may be very intuitive about being at fault (the vehicle that rear-ended the automated vehicle), the reality is different. Today, courts evaluate in many cases who of the involved parties is at fault, although the police report filed a main culprit (Destatis, 2021). To solve that issue for the safety metric, it has been decided to consider all collisions on which a system can have an impact, the FFoA independently of the question of who is at fault. Thus, the assumption is made that the likelihood of being involved in this type of collision is independent of the automation (i.e., manual or automation).

However, $n_{Collision}$ does not only depend on the type of vehicle involved. Developing an automated vehicle requires a description of the ODD. In this example, an ADS that is capable to operate on highways with a maximum velocity of 130 kph is considered. For simplicity, other constraints such as weather are neglected. Accordingly, the question arises if $n_{Collision}$ refers to collisions, where a passenger car was involved on a highway with a maximum speed of 130 kph. Here, $n_{Collision}$ would refer to the ODD. But how to deal with collisions, where a passenger car was involved but outside the ODD? For example, single car crashes on a highway, where the collision speed of the vehicle was at 180 kph?

Again, as the high-speed collision from above was outside the ODD, one may argue that this collision should not contribute to $n_{Collision}$. However, as the considered ADS is only capable of driving 130 kph, such high-speed collisions are eventually prevented (positive effect of an ADS). This example shows that it is not sufficient to focus only on collisions in the ODD. Thus, it is necessary to evaluate if the ADS has an impact on other crash types outside the ODD. Hence, the ODD is extended and is called FFoA. The precise definition/description of the FFoA is typically not limited by possible criteria, but by information that data sources can provide. Today accident databases either are limited in their granularity or in the number of cases. Solving this issue would allow a more precise identification of relevant collisions.

The following Fig. 5 shows the relationship between the ODD, FFoA and all possible traffic situations. The traffic
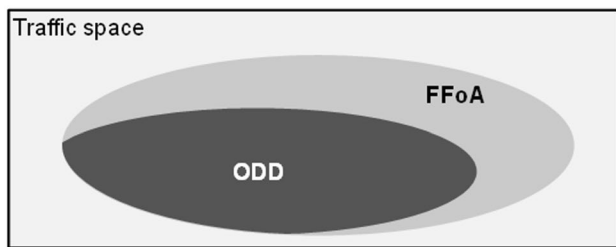
Fig. 5 Relationship between ODD and FFoA

Table 1 Number of collisions with involvement of a passenger car on German motorways (FFoA) (Destatis, 2015, Destatis, 2017, Destatis, 2018, Destatis, 2019, Destatis, 2021)

Severity	Number of collisions with passenger car per year				
	2015	2016	2017	2018	2019
Fatal	287	250	236	248	230
Severe	3495	3753	3627	35,473	3457
Slightly	14,342	15,258	14,973	14,518	14,308
Property damage	129,966	140,626	152,572	133,684	136,851

Table 2 Estimated mileage of passenger vehicles per year (Bast, 2014)

Driven mileage	2014	2015	2016	2017	2018	2019
Motorway (all vehicles) [10^9 km]	230.6	237.6	243.5	246.4	249.4	252.1
Rel. change [%]		3.0	2.5	1.2	1.2	1.1
Estimated motorway (passenger vehicles) [10^9 km]	164.5	169.5	173.7	175.8	177.9	179.9

Table 3 Driven Distance between two collisions of passenger cars in Germany per year

	2015	2016	2017	2018	2019
Motorway [Mio. km]	591	695	745	717	782

of severity refer to the maximum severity of all participants involved in the crash. For example, if a passenger vehicle collides with a pedestrian who is getting killed, the accident is counted as fatal, although the driver of the vehicle may not be injured at all.

Based on the equation above, the mileage of passenger vehicles m on highways is required. Depending on the country, the annual mileage may not be provided for every year. For example, in Germany, the mileage of passenger vehicles on a highway is estimated only every several years. Specifically, 2014 is currently the latest year for which mileage for passenger vehicles on highways is available (Bast, 2017). However, the annual mileage of all types of vehicles (including trucks, motorbikes, etc.) on German highways is available for every year. Accordingly, the relative change of the overall mileage $\Delta m_{i+1,AllVehicles}$ is used to extrapolate the mileage as follows

$$m_{i+1} = m_i \cdot (1 + \Delta m_{i+1,AllVehicles}) \tag{3}$$

where m_i denotes the mileage of passenger cars in a specific year.

Table 2 shows the resulting mileage for all vehicle types, the relative change per year and the calculated mileages for passenger vehicles on German highways using Eq. (3).

Having the number of collisions and the annual mileage available, we can calculate the average distance between two collisions for the individual severities (Eq. 2). Table 3 shows the average distance between two fatal collisions over the years.

As fatal collisions are extremely rare events, one can see that the distance varies between 591 Mio. km and 782 Mio. km. These variations induce difficulties determining a valid baseline. Considering two different companies developing the same ADS but starting in different years, the requirement of their system would be significantly different, although developing the same ADS. Therefore, we propose to take the

space contains other collisions of other road types, while in our example, FFoA and ODD are limited to highway collisions.

In general, the ODD is always a subset of FFoA, which itself is always a subset of the whole traffic space except for fully automated vehicles of SAE Level 5.

To conclude this section, the question remains open, what type of human driver is considered to quantify SP_{HD}. In the literature, different proposals are made if safety performance should be based on an attentive driver or an average driver (Blumenthal et al., 2020). Finally, the question may be answered by the availability of such data, especially if SP_{HD} is based on crash statistics.

Safety performance based on crash statistics

In this section national accident data and national data on mileage is used to quantify SP, see Eq-2, based on the considerations above. The following Table 1 shows the number of collisions $n_{Collision}$ where a passenger car was involved on German highways (Destatis, 2015, Destatis, 2017, Destatis, 2018, Destatis, 2019, Destatis, 2021). The different levels

average, for example, of the last five years ($n_{year} = 5$). This provides a realistic estimate of today's safety performance while averaging over seasonal effects.

$$SP_{HD} = \frac{\sum_{i=1}^{n_{year}} SP_{HD,i}}{n_{year}} \tag{4}$$

The discussion above highlights the difficulty of extracting the right collisions to quantify the safety performance of a human driver. Using the average value is based on all drivers, i.e. ranging from the attentive driver up to the inattentive and not experienced driver. Furthermore, the outcome of a collision with respect to injury severity is influenced by active safety measures or infrastructure (Bundesministerium für Verkehr und digitale Infrastruktur, 2015). For example, guardrails prevent vehicles to run off the road and collide with trees, which would generally result in severe injuries. Accordingly, speaking of the safety performance of a human driver might be misleading. The numbers above reflect rather the safety performance of today's traffic including also other stakeholders. Furthermore, the ODD may not be activated at specific environmental conditions (e.g. heavy snowfall). Following the logic from above, collisions under those conditions and being one of the contributing factors should be neglected as the ADS isn`t capable prevent those collisions. However, national statistics does not provide sufficient details allowing the extraction of respective information. This underlines uncertainties in crash statistics and requires the consideration of safety factors.

Safety factors and safety thresholds

As discussed above, the average distance between two fatal collisions exhibits large variations. Although, an average value can be obtained, it is still unknown, what the true value of today's traffic safety performance is. Assuming a Gaussian distribution to model the uncertainties in the safety performance, the real value may not be the obtained average value from above. This means that if an ADS system is designed to simply lie above the average value (here distance between two collisions), the risk balance might not be fulfilled. Consequently, we propose to set the threshold higher than the estimated average value. Specifically, we propose to consider a safety margin comprising of two safety factors. The first safety factor considers a target threshold two times the standard deviation σ (safety factor 1) above the average value. Accordingly, the target safety performance SP_{Target} is calculated as:

$$SP_{Target} = SP_{HD} + 2\sigma \tag{5}$$

In addition, when developing ADS, engineers face a lot of uncertainties with respect to sensor performance, occurrence frequency of specific situations (e.g. obstacles on the road), etc. In some cases, only expert judgment is available. This means, that an additional safety factor (2nd safety factor) is required based on the uncertainty with respect to design decisions of the ADS. In the pharmaceutical sector, a safety factor of 10 is applied to account for uncertainties in data quality concerning the dosage of a drug for a human being (CDER, 2005). However, if more and more data and information is available, the uncertainty is reduced. This would allow a reduction of the 2nd safety factor accordingly. However, the value SP_{Target} must never be exceeded. In general, such uncertainties are not limited to ADS. The inclusion of safety factors is an established concept and are sometimes requested by regulations (CDER, 2005). Figure 6 also highlights that a residual risk remains despite using different safety factors. Society, industry, and ultimately everyone must answer the question, what is a suitable and acceptable residual risk.
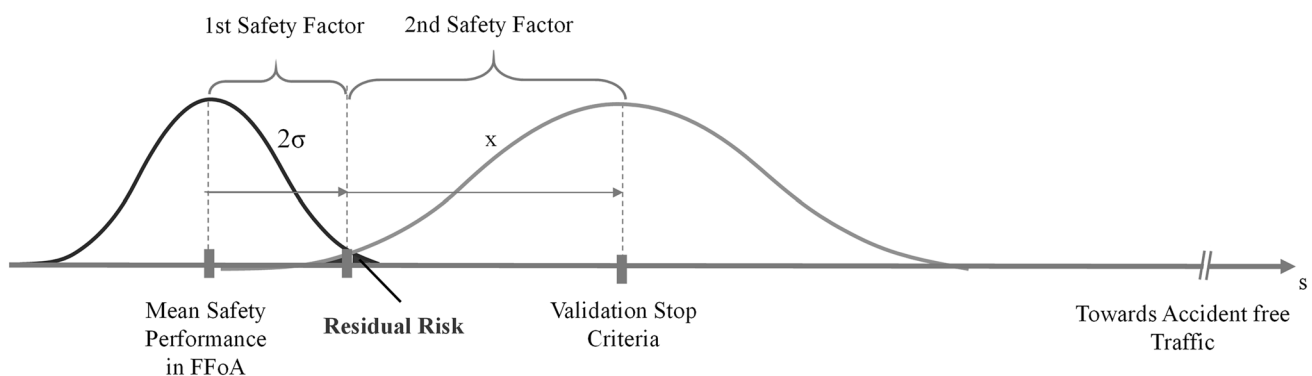

Fig. 6 Safety factors and residual risk

Assessing the safety performance of an ADS

On system level there is the question of how the PoRiBa can be proven once the system is approved for operation on public roads. The European Commission requests for such an assessment in its ethic report the establishment of an objective baseline represented by non-automated vehicles, the application of coherent metrics of road safety as well as new methods for continuously monitoring ADS safety (Bonnefon et al., 2020). However, all relevant ethical reports remain rather on a general level without describing detailed approaches for answering this question. Therefore, in the following approaches for the safety performance assessment to prove a positive risk balance of ADS are discussed.

Challenges related to the safety assessment of automate driving

Before discussing different approaches, it is important to discuss the challenges related to the safety performance assessment with respect to traffic safety. Three main challenges are identified for the assessment of ADS:

Challenge A. Timing of assessment The traditional approach is to prove the system's safety performance by means of statistical analysis of accident data (Unselt, 2004; HLDI, 2021; Spicer et al., 2018). For this purpose, different sources of accident data can be applied, such as (police) reported accident data (Famer, 2004; Knoll et al. 2006) (Unselt et al., 2004), insurance data (HLDI, 2021) or emergency call data (Spicer et al., 2018). However, the German Ethics Commission stated that the introduction of ADS depends on proving of the system's PoRiBa (Fabio et al., 2017). This poses a challenge to the traditional assessment approaches since historical accident data will not be available at this stage.

Furthermore, it is expected that ADS data on this matter with sufficient sample size will not be available in near future. To support the argument a small example is calculated (see Table 4). It is presumed that a car manufacturer sells on average 5000 automated vehicles per year. The annual mileage on the motorway of each vehicle with active ADS is 5000 km. To consider the fact that vehicles are constantly sold over the year, in the current year only half of the total amount is considered for the calculation. The vehicles that encountered a crash with sever or fatal injuries are removed from the fleet. The risk of having a collision is equal to the current given risk of manual driving e.g. the risk of a collision in German is used (see chapter 4). According to the prediction in the given example in five years approx. 280 collisions including 33 collisions

with injuries are likely to occur. These numbers indicate that the empirical validation of ADS is rather a challenge as already been discussion by Winner et al. (2015) and Zhao et al. (2017). This leads to the conclusion that apart from the traditional retrospective assessment method further approaches are required. These approaches are called prospective safety assessment approaches (Page et al., 2015) and are discussed in chapter 5.2.

Challenge B. Comprehensive assessment: To comply with the recommendations in the EU's Ethic report (Bonnefon et al., 2020) of a fair assessment, a comprehensive view on the safety performance is required. Different frameworks for impact assessments have been defined that group the potential effects into different categories. Smith et al. (2017) take a quite broad approach considering apart from safety and vehicle operations effects like network efficiency, land use and public health (Innamaa et al., 2018; Smith et al., 2017). A similar direction is taken by Milakis et al (2017), which structures the impact of ADS depending in three levels, in which the first level covers rather the direct impacts and the higher levels more indirect effects. However, these frameworks are rather intended to serve as a qualitative assessment than a quantitative assessment. A widely deployed framework for the quantitative safety impact assessment is the nine safety mechanism approaches (Draskóczy et al., 1998). This approach considers a technology's direct and indirect effects as well as exposure effects. The approach has been applied in several European research projects (Innamaa et al., 2020; Larsson et al., 2012; Malone et al., 2008). One limitation of its application is the absence of a detailed calculation instructions for the individual mechanism, which often leads to the use of expert opinions instead of data-based calculations—in particular for the indirect and exposure effects. Although these effects contribute to the overall safety impact performance, the focus in the assessment is typically placed on direct safety effects of a technology. According to the L3Pilot project, a technology can affect traffic safety directly in three possible ways (Metz et al., 2019):

(1) *Reduction of risk in a scenario* This corresponds to the intended effect for each technology that aims at improving traffic safety. However, it must be noted that an improvement can only be achieved in such situations, in which the human driver does not perform well, i.e. today's collisions.

(2) *Not affecting the risk in a scenario* This category typically includes scenarios that are outside of the vehicles ODD and therefore are not directly affected by the ADS.

(3) *Causing potentially new risks* This category covers scenarios, in which ADS potentially performs at least as well as the human driver. During the development it is essential to reduce the occurrence of these scenarios to

a technical minimum by ensuring the functional safety (ISO, 26262, 2018), the safety of the intended functionality (ISO, 21448, 2019) as well as the cybersecurity (UNECE, 2020). An often-discussed example for this category is the MRM (Innamaa et al., 2020).

For a comprehensive analysis it is indispensable that the assessment covers scenarios of all three categories, although the identification of scenarios for the last category represents the greatest challenge since they might not even be known to the developer (ISO, 21448, 2019). In this context, "corner cases" (i.e. scenarios that are very rare but have a significant societal impact due to its potential negative consequences) need to be considered.

Challenge C. Definition of a baseline: An objective baseline is a key aspect for the safety performance assessment (Bonnefon et al., 2020). The P.E.A.R.S. initiative (an open consortium to harmonise the prospective effectiveness assessment of active safety systems by simulation) formulated five different aspects that need to be defined prior to the assessment (Page et al., 2015, P.E.A.R.S., 2021):

- technology under assessment (incl. penetration rate of technology under assessment as well as other technologies that are considered in the baseline),
- scenario and metric,
- considered (environmental, infrastructure etc.) limitations,
- considered region and time horizon of the projection,
- envisioned level of confidence in relation to the objective of the research question.

In particular, the definition of the penetration rate of other technologies in the baseline is a key aspect in the quantification of the ADS safety performance. The reason is that today's safety oriented advanced driver assistance systems (ADAS), like autonomous emergency braking (AEB) systems, address similar crash scenarios as ADS. The effectiveness of ADAS has been proven in multiple studies (Isaksson-Hellman et al., 2012, Spicer, 2018). Already today, human driver's benefit from these systems. This needs to be recognized when defining the baseline for the assessment. Consequently, this leads to the challenge that a reasonable quantification of the existing ADAS penetration rate in the market is indispensable for the assessment. L3Pilot, for instance, defined a market penetration of 7.5% for the AEB systems (Innamaa et al., 2020).

Prospective safety performance assessment approaches

Overview on approaches

Different approaches have been defined to conduct a prospective safety assessment. There are rather simple approaches

such as identifying the field of application by means of accident data analysis. This approach has been applied for instance by Kocherscheid (2004), Dryselius (1990) and Sternlund (2017). Although this approach is rather simple to implement, its outcome is strongly affected by quality and level of detail of the analysed accident data and thus often provides only a rough estimate, since the actions of the new technology are not considered in detail. Furthermore, due to the analysis of historical accident data and the constant changes in the traffic environment, it is questionable whether *challenge C* can be addressed adequately with this approach.

Studies in driving simulators or on test tracks can be used to identify the technology's safety performance. This approach is described for instance by Breuer (2015). The clear advantage of this approach is that it can analyse the interaction between driver and technology. The major drawback is that typically only very limited number of situations can be analysed due to resource limitations. In the light of ADS that are applicable in various scenarios, this approach allows only to investigate few situations in detail (e.g. takeover situations). However, it will not allow for a comprehensive assessment as mentioned in *challenge B*.

A third approach to investigate the safety performance of a technology is by means of a field operational test (FOT). This approach has been applied in several projects (e.g. Kessler et al., 2012; Sayer et al., 2010). The advantage of the approach is that it allows to investigate technology`s effect in a real environment. On the other hand, collisions remain rare events. Thus, detecting statistically relevant number of collisions is rather difficult with reasonable effort. This often leads to use surrogate measures, like e.g. critical driving situations (Benmimoun et al., 2011; Dingus et al., 2006). These measures can provide an indication on the safety performance of a technology. However, they do not allow for a direct link to the safety performance in terms of reduction of collisions. It is rather necessary to conduct further analysis as described by e.g. Najm et al. (2000) and Najm et al. (2006). Furthermore, this approach can only be applied at a very late stage in the development since the technology must be mature enough to operate on public roads safely. Thus, *challenge A* is a serious challenge.

A fourth approach is a simulation-based assessment approach. Simulation can be implemented in different ways: Hardware-in-the-Loop, Model-in-the-Loop or Software-in-the Loop. An overview can be found in Hakuli and Krug (2015). The focus in the following is on the latter. Simulation can be applied already from an early stage on in the development (*challenge A*). They allow investigating a high number of scenarios with reasonable efforts (*challenge B*) and can be set up in a controlled way (*challenge C*). However, there remains the question regarding the transferability of conclusions to the real-world. The answer to this question highly depends on the evaluation scope and the quality of the

implementation (P.E.A.R.S., 2021). To prove the correctness of simulation validation and verification activities are essential to this approach. Despite this aspect, virtual simulation remains the most promising for a comprehensive assessment of ADS's safety performance. Therefore, these approaches will be discussed in more detail in the following.

Simulation based approaches

A simulation tool consists of several models that represent the environment, the vehicle including technologies as well as the driver during the course of simulation. Depending on the assessment scope and the applied simulation approach, requirements regarding model fidelity vary. Overall, P.E.A.R.S. defines four approaches for deriving the simulated baseline case (P.E.A.R.S. 2021, ISO 21934):

A: Direct usage of real-world cases (i.e. reconstructed crash data or field data) without any changes). This approach has been applied for instance in interactIVe (van Noort, 2015).

B: Usage of real-world cases plus varying the initial values by means of distribution. Examples for the application of this approach are Sander et al. (2018) and Fahrenkrog (2016).

C1: Deriving scenario mechanism and distribution from real world case and selecting a low number of representative cases. This approach has been applied for instance in Euro NCAP for passive safety tests (Ellway et al., 2019) or for active safety technology by the CATS Project (op den Camp, 2016).

C2: Deriving scenario mechanisms and distributions from real world cases and applying a sampling in order to get multiple cases. Examples for the application of this approach are Helmer (2014) and Fahrenkrog et al. (2019).

Each of the simulation approaches has advantages and disadvantages depending on the available data and scope of evaluation. For the assessment of ADS, technology's fundamental mechanism must be considered. ADS is constantly operating and in contrast to a safety oriented ADAS not only in case of critical situations. Therefore, it is not sufficient to cover only a couple of seconds in the simulation. It is rather important to extend the simulation period to allow the ADS to enter the relevant situation with realistic speed and distance to other traffic participants. Regarding the usage of possible data sources this aspect sets limits for those approaches that have a direct link to the real-world cases (see approach A. and B.), since reconstructed accident data cover typically only couple of seconds. In the German Pre-Crash Matrix (PCM) database, the period of a scenario is only 5 s (Spitzhüttl et al., 2015). Typically, the application of field data from naturalistic driving studies or FOTs does not have such limitations, since in these studies time series data are logged permanently (Kessler et al., 2012; Sayer et al., 2010). On the other hand, in this dataset only a limited number of collisions is recorded. In the light of comprehensive assessment, this is not ideal either. Approach C1 is also not applicable since with the limited number of cases a comprehensive analysis of the situation space is not possible.

For these reasons we expect that the approach C2, which does not directly rely on the real-world cases but establishes the link to them via distributions, is the most appropriate approach for the prospective safety performance assessment of ADS. In this approach only the starting conditions of the traffic participants are sampled from input distributions (Fahrenkrog et al., 2019; Helmer, 2014), which can be derived from accident data as well as field data depending on the scenario in question. In general, there is a need for extending in-depth database covering traffic and crash scenario. From such extension not only C2 but all approaches would benefit. In sense, data driven activities like SHRP2 (National Academies of Sciences 2014), PEGASUS (2019) or ADScene (Arnoux, 2021) contribute to better assessments.

However, the challenge of approach C2 is that in contrast to approach A and B no predefined trajectories are available. Therefore, the movement of the traffic participants resulting from the starting conditions needs to be derived during the simulation. Typically, a driver behavior model handles this task. This leads to high requirements for the driver behavior model since it needs to cover "everyday driving" as well as the human reaction in critical situations in a realistic manner. For both types of situations, models have been developed. "*Everyday driving*" models that cover the behavior in non-critical traffic situation are for instance the Wiedemann model (Wiedemann, 1974), the intelligent driver model (Treiber et al., 2000) or Sumo (Alvarez Lopez et al., 2018). Models covering the behavior in critical situation are often tailored to certain scenarios. Examples are the tau-theory model (Lee, 1976), Japanese model in the UN ECE ALKS regulation (UN ECE ALKS, 2020) or the model described by Rösener (2020). Further developments start to cover also both types of models. Examples of such driver behavior models are the model described by Mai (2017) and the stochastic cognitive model (Witt, Kompaß, et al., 2019; Witt, Wang, et al., 2019). Despite the recent developments, further improvements in this area are required to achieve a realistic modelling of human driving behavior to ensure a correct baseline for the assessment.

Examples for the application for a simulation assessment of the ADS's traffic safety effects can be found Sinha et al. (2020) and Bjorvatn et al. (2021), which itself include an overview about different studies. In general, the reported effects of ADSs differs quite a lot among the studies depending on the analysed system, traffic situation, conflict type and penetration rates. This underlines the need for

detailed investigations in the safety effects for individual implementations.

Exemplary application of simulation

To exemplarily demonstrate the principles of approach C2 for the prospective safety assessment by simulation as described above, it is applied for exemplary ADS in two scenarios. The first scenario covers rear-end conflicts with the predecessor, in which the ADS is expected to minimize the risk. The predecessor is either a slow vehicle, a decelerating vehicle or slow vehicle that as well starts to decelerate at random point of time. The second scenario is a MRM scenario that could lead to potentially additional risk. The MRM manoeuvre is activated at a random point of time in the simulation.

The exemplary ADS is a SAE Level 3 automated driving system which is designed for motorway drives up to speed of 130 kph. It must be noted that the exemplary ADS do not allow for any conclusions on a real implementation. The function is capable of performing automated lane changes. The minimum risk manoeuvre is executed by a constant deceleration with 2 m/s² up to stand still. During the MRM the ADS does not change the lane. The MRM is activated when neither the driver nor the vehicle is in charge of controlling the vehicle and corresponds to a fallback level intending to resolve such situations which are likely to result in collisions. Such situations are likely to result in a collision. The MRM is only executed if the driver does not respond to a TOR within a certain time period. The MRM stops once the driver takes over. This means that stand still

will only be reached in case the driver does not respond at all (Fig. 7).

Both scenarios have been simulated for the baseline condition with the human driver model and for the treatment condition with the ADS for the ego-vehicle. The applied driver behavior model is the SCM model (Witt, Kompaß, et al., 2019; Witt, Wang, et al., 2019). Multiple runs have been simulated, in which next to the kinematic starting condition also the infrastructural (i.e. speed limit, number of lanes) and traffic conditions (i.e. traffic volumes) have been varied. For the rear-end scenario the speed limit (unlimited) and the number of lanes (3 lane) is less relevant. Therefore, the focus is purely on the variation of the traffic volume. Covered traffic volumes are 100, 250, 500, 750, 1000 and 1500 vehicles per hour and lane. For the MRM the motorway has been simulated with 2 and 3 lanes as well as the speed limits 80 kph, 100 kph, 120 kph, 130 kph and unlimited. The simulated traffic volumes are the same as for the rear-end scenario.

The simulation has been conducted by means of open-source simulation tool openPASS (openPASS, 2021) on a standard desktop computer. The results are given in Table 4.

This example is intended to highlight the capabilities of assessing the effect of automated driving in different scenarios but does not provide a complete assessment. It must also be noted that the calculation of the effects in the scenarios, will not be enough to calculate the safety impact. In addition, the scenarios' frequencies need to be considered. For the crash-related scenarios, national and international statistics shall be considered as described in chapter 4. For those scenarios, in which the ADS potentially increase the risk, accident data are typically not an appropriated source
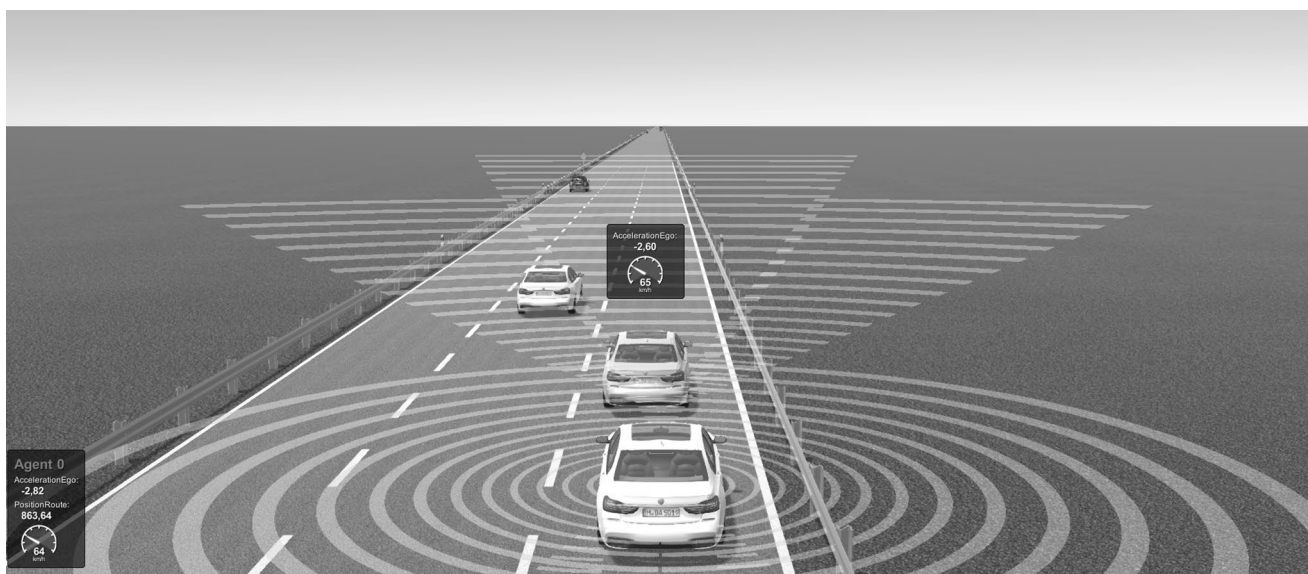

Fig. 7 Visualization of a rear-end conflict scenario (slower and decelerating predecessor) for an ADS simulated with openPASS

Table 4 Prediction of cumulative number of collisions of motorway ADS for years after market introduction for the given scenario based on the calculated human drivers' collision risk

Scenario	Number of simulations baseline [–]	Crash rate baseline [–] (%)	Number of simulations treatment [–]	Crash rate treatment [–] (%)	Relative change of crash rate [–] (%)
Rear-end	594	33.3	598	19.6	− 41.1
MRM	929	3.9	1574	5.8	+ 50.8

for describing the frequency. Here, in case of infrastructure related scenarios the frequency should be determined by means of infrastructure data (e.g., spatial frequency of passing a motorway entrance) or in case of system related scenario—as for the MRM—based on system information. A combination of different data sources could be required to determine this frequency.

Conclusion

This paper provides a methodological approach of how to deal with the requirements of a Positive Risk Balance defined by regulatory authorities. The qualitative framework PrOACT- URL, adapted from the pharmaceutical sector, represents an adequate approach. The next step consists in further refining the quantitative methods as a part of the PoRiBa. These include aspects such as the way in which the results are presented to the authorities as well as the examination of methods that allow for a potential offsetting of the results from the individual disciplines such as safety of intended functionality, cybersecurity, functional safety as well as prospective safety performance assessment. Another challenge constitutes the application of the safety metric to other functional use cases (for example in traffic jam assist) as well as other markets as it requires the availability of data with sufficient quality.

Proofing the PoRiBa for an ADS requires at the beginning of its deployment prospective safety assessment methods. Among different possible methods and tools, the virtual simulation will play a crucial role. Therefore, the compliance of the virtual assessment with the real world needs to be proven by means of the validation and verification of the simulation and its sub-models. The required fidelity of the models as well as the required accuracy of the overall simulation is still an open issue that needs further discussion involving different stakeholders. Another important aspect related to the virtual assessment is the availability of data to set the parameters or to validate the simulation. Here, large differences in terms of amount and quality can be recognized between different countries. Since data builds the

foundation for a solid assessment, joint efforts are required from all stakeholders (industry, academia & research as well as on governmental side) involved. In this context, learning from previous assessments and developments is a valuable contribution to the method. Therefore, a feedback-loop shall be considered once the ADS has been launched and the next generation of ADS is developed.

References

Alvarez Lopez, P., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., & Wießner, E. (2018). Microscopic traffic simulation using SUMO. IEEE Intelligent Transportation Systems Conference (ITSC), IEEE. *The 21st IEEE International Conference on Intelligent Transportation System*s, 4.-7. Nov. 2018 (pp. 2575–2582).

Arnoux, E. (2021). *Scenarios from real life: PFA proposals for AD safety validation, French Scenarios Library, EUCAD Conference 2021*. Retrieved from https://www.connectedautomateddriving.eu/wp-content/uploads/2021/05/EUCAD2021_D3_BO11_Emmanuel-Arnoux.pdf

Bast (Bundesanstalt für Straßenwesen). (2014). *Ergebnisse Fahrleistungserhebung 2014*. Retrieved from https://www.bast.de/BASt_2017/DE/Verkehrssicherheit/Fachthemen/U2-fahrleistung-2014/u2-Fahrleistung-2014-ergebnisse.html

Banks, V. A., Plant, K. L., & Stanton, N. A. (2019). Driving aviation forward; contrasting driving automation and aviation automation. *Theoretical Issues in Ergonomics Science, 20*(3), 250–264.

Benmimoun, M., Fahrenkrog, F., Zlocki, A., Eckstein, L. (2011). *Incident detection based on vehicle CAN-data within the large scale field operational test "euroFOT" 22nd Enhanced Safety of Vehicles Conference*. Retrieved from https://www-esv.nhtsa.dot.gov/Proceedings/22/files/22ESV-000042.pdf

Bjorvatn, A., Page, Y., Fahrenkrog, F., Weber, H., Aittoniemi, E., Bärgman, J., Borrack, M., Hermitte, T., Heum, P., Hiller, J., Innamaa, S., Itkonen, T., Lehtonen, E., Malin, F., Pedersen, K., Silla, A., Sintonen, H., Streubel, T., & Torrao, G. (2021). *Impact Evaluation Results, L3Pilot Deliverable D7.4*.

Blumenthal, M., Fraade-Blanar, L., Best, R., & Irwin, J. (2020). *Safe Enough: Approaches to assessing acceptable safety for automated vehicles. Approaches to Assessing Acceptable Safety for Automated Vehicles*. RAND Corporation. https://doi.org/10.7249/rra569-1

Bundesministerium für Verkehr und digitale Infrastruktur. (2015, October). *Halbzeitbilanz des Verkehrssicherheitsprogramms 2011–2020*. Retrieved from https://www.bmvi.de/SharedDocs/

DE/Publikationen/StV/halbzeitbilanz-verkehrssicherheitsprogramm.pdf?__blob=publicationFile.

Sayer, J. R., Bogard, S. E., Funkhouser, D., LeBlanc, D.J., Bao, S., Blankespoor, A.D., Buonarosa, M. L., & Winkler, C. B. (2010). *Integrated vehicle-based safety systems heavy-truck field operational test key findings report, The University of Michigan Transportation Research Institute, Report No. DOT HS 811 362*. Retrieved from https://deepblue.lib.umich.edu/bitstream/handle/2027.42/77988/102672.pdf?sequence=1&isAllowed=y

Bonnefon, J.-F., Cerny, D., Danaher, J., Deviller, N., Johansson, V., Kovacikova, T., Martens, M., Mladenovic, M. N., Palade, P., Reed, N., Santoni de Sio, F., Tsinorema, S., Wachter, S., & Zawieska, K. (2020). *Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility. Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659)*. Publication Office of the European Union. Retrieved from https://ec.europa.eu/info/sites/info/files/research_and_innovation/ethics_of_connected_and_automated_vehicles_report.pdf

Breuer, J., von Hugo, C., Mücke, S., & Tattersall, S. (2015). Nutzerorientierte Bewertungsverfahren von Fahrerassistenzsystemen. *Handbuch Fahrerassistenzsysteme* (pp. 183–196). Springer Fachmedien Wiesbaden.

CDER. (U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research). (2005). *Guidance for industry: Characterization and qualification of cell substrates and other biological starting materials used in the production of viral vaccines for the prevention and treatment of infectious diseases*. Retrieved from https://www.fda.gov/media/72309/download

Destatis (Statistisches Bundesamt). (2021). Verkehr – Verkehrsunfälle 2019 (Fachserie 8 Reihe 7). *Statistisches Bundesamt.* Retrieved from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/Publikationen/Downloads-Verkehrsunfaelle/verkehrsunfaelle-jahr-2080700197004.pdf?__blob=publicationFile

Destatis (Statistisches Bundesamt). (2019). *Verkehr: Verkehrsunfälle 2018 (Fachserie 8 Reihe 7)*. Statistisches Bundesamt. Retrieved from https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00052769/5462403187004_akt20191017.pdf

Destatis (Statistisches Bundesamt). (2018). *Verkehr – Verkehrsunfälle 2017 (Fachserie 8 Reihe 7)*. Statistisches Bundesamt. Retrieved from https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00040387/2080700177004_Erg16082018.pdf

Destatis (Statistisches Bundesamt). (2017). *Verkehr – Verkehrsunfälle 2016 (Fachserie 8 Reihe 7)*. Statistisches Bundesamt. Retrieved from https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00035071/2080700167004_Korr16042018.pdf

Destatis (Statistisches Bundesamt). (2015). *Verkehr – Verkehrsunfälle 2015 (Fachserie 8 Reihe 7)*. Statistisches Bundesamt. Retrieved from https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00016632/2080700151034.pdf;jsessionid=640EE976AC4916A939407BD8060D9053

Dietmayer, K. (2016). Predicting of machine perception for automated driving. *Autonomous Driving* (pp. 407–424). Springer.

Milakis, D., van Arem, B., & van Wee, B. (2017). Policy and society related implications of automated driving: A review of literature and directions for future research. *Journal of Intelligent Transportation Systems, 21*(4), 324–348. https://doi.org/10.1080/15472450.2017.1291351

Food and Drug Administration. (2018). *Benefit-risk assessment in drug regulatory decision-making: Draft PDUFA VI implementation plan (FY 2018–2022)*. Retrieved from https://www.fda.gov/files/about%20fda/published/Benefit-Risk-Assessment-in-Drug-Regulatory-Decision-Making.pdf

DIN EN ISO 14971:2020-07 Medical devices: Application of risk management to medical devices (ISO 14971:2019); German version EN ISO 14971:2019

DIN EN 50129:2019-06 Railway applications: Communication, signalling and processing systems— Safety related electronic systems for signalling; German version EN 50129:2018 + AC:2019

DIN EN 50126-1:2018-10 Railway Applications: The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS) - Part 1: Generic RAMS Process; German version EN 50126-1:2017

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., & Knipling, R. R. (2006). *The 100-car naturalistic driving study, Phase II: Results of the 100-car field experiment, Virginia tech transportation institute, Report No. DOT HS 810 593*. Retrieved from https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/100carmain.pdf

Draskóczy, M., Carsten, O. M. J., & Kulmala, R. (1998), *Road safety guidelines. CODE Project, telematics application programme, Deliverable B5.2.*

Dryselius, B. (1990). *Pro-Gen Safety group summary report: Estimation of the potential safety effects of different possible PROMETHEUS-Functions, Prometheus (Program)*. PRO-GEN Safety Group.

EMA European Medicines Agency. *Benefit-risk methodology project work package 3 report: Field tests*. European Medicines Agency; 31 Aug 2012. Retrieved from http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/09/WC500112088.pdf. Report no: EMA/718294/2011.

EMA European Medicines Agency. (n.d.). *PROACT-URL*. Retrieved from https://www.protectbenefitrisk.eu/PrOACT-URL.html

Ellway, J., Hallbauer, K., & Kerz, T. (2019). The development of a Euro NCAP far side occupant test and assessment procedure. *26th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. Retrieved from https://www-esv.nhtsa.dot.gov/Proceedings/26/26ESV-000278.pdf

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: THe Journal of the Human Factors and Ergonomics Society, 37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Di Fabio, U., Broy, M., Brüngger, R., Eichhorn, U., Grunwald, A., Heckmann, D., Hilgendorf, E., Kagermann, H., Losinger, A., Lutz-Bachmann, M., Lütge, C., Markl, A., Müller, K., & Nehm, K. (2017). *Ethic commission: automated and connected driving, Report of ethics commission appointed by the federal minister of transport and digital infrastructure*. Retrieved from https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

Fahrenkrog, F. (2016). Wirksamkeitsanalyse von Fahrerassistenzsystemen in Bezug auf die Verkehrssicherheit. Doctoral Dissertation, RWTH Aachen University.Schriftenreihe Automobiltechnik Issue Number: 196, fka Forschungsgesellschaft Kraftfahrzeuge mbH, ISBN 978-3-946019-01-5.

Fahrenkrog, F., Wang, L., Platzer, T., Fries, A., Raisch, F., & Kompaß, K. (2019). Prospective safety effectiveness assessment of automated driving function: From the method to the results. Enhanced Safety of Vehicles. Retrieved from https://www-esv.nhtsa.dot.gov/Proceedings/26/26ESV-000166.pdf

Färber, B. (2015). Kommunikationsprobleme zwischen autonomen Fahrzeugen und menschlichen Fahrern. *Autonomes Fahren* (pp. 127–146). Springer.

Farmer, C. M. (2004). Effect of electronic stability control on automobile crash risk. *Traffic Injury Prevention, 5*(4), 317–325. https://doi.org/10.1080/15389580490896951

Hakuli, S., & Krug, M. (2015). Virtuelle integration. *Handbuch fahrerassistenzsysteme* (pp. 125–138). Springer Fachmedien Wiesbaden.

Hancock, P. A., Simmons, L., Hashemi, L., Howarth, H., & Ranney, T. (1999). The effects of in-vehicle distraction on driver response during a crucial driving maneuver. *Transportation Human Factors, 1*(4), 295–309. https://doi.org/10.1207/sthf0104_1

Helmer T. (2014). *Development of a Methodology for the Evaluation of Active Safety using the Example of Preventive Pedestrian Protection*. Doctoral Dissertation. TU Berlin. Retrieved from https://depositonce.tu-berlin.de.https://depositonce.tu-berlin.de/bitstream/11303/4270/2/helmer_thomas.pdf

HLDI. (2021). Volvo city safety loss experience: An update Highway Loss Data Institute. *Bulletin 29*(23). Retrieved from https://www.iihs.org/media/48c6e9ae-d60b-4cc7-9bf6-a330ef1d177e/-808307776/HLDI%20Research/Bulletins/hldi_bulletin_29.23.pdf

Hunink, M., Weinstein, M., Wittenberg, E., Drummond, M., Pilskin, J., Wong, J., & Glasziou, P. (2001). *Decision making in health and medicine: Integrating evidence and values*. Cambridge University Press.

Innamaa, S., Smith, S., Barnard, Y., Rainville, L., Rakoff, H., Horiguchi, R., & Gellerman, H. (2018). *Trilateral impact assessment framework for automation in road transportation, Report of the EU-US-Japan ITS Cooperation*. Retrieved from https://connectedautomateddriving.eu/wp-content/uploads/2018/03/Trilateral_IA_Framework_April2018.pdf

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, B., Di Lillo, L., Merat, N., Metz, B., Page, Y., Shi, E., & Sintonen, H. (2020). *Evaluation plan, L3Pilot Project, Deliverable D3.4*. Retrieved from https://l3pilot.eu/fileadmin/user_upload/Downloads/Deliverables/L3Pilot-SP3-D3.4-Evaluation_plan-v1.0_DRAFT_for_website.pdf

International Civil Aviation Organization. (2008). Doc 9859 safety management manual. In *International Civil Aviation Organization* (Fourth edition). Retrieved from http://www.icao.int/fsix/_Library/SMM-9859_1ed_en.pdf, C:/Users/Danilo/Downloads/Safety_management_and_risk_modelling_in_aviation.pdf, http://www.easa.eu.int/essi/documents/Methodology.pdf

Isaksson-Hellman, I., & Lindman, M. (2012). "The effect of a low-speed automatic brake system estimated from real life data." Annals of advances in automotive medicine. Association for the Advancement of Automotive Medicine. Annual Scientific Conference (Vol. 56, pp. 231–40). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3503425/

ISO 4804. (2020). *Road vehicles: Safety and cybersecurity for automated driving systems: Design, verification and validation. ISO Technical report 4804:2020*.

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

ISO 26262. (2018). *Road vehicles: Functional safety*. ISO standard series ISO 26262.

ISO 21448. (2019). *Road vehicles — Safety of the intended functionality*. ISO standard ISO/PAS 21448:2019.

Kessler, C., Etemad, A., Alessandretti, G., Heinig, K., Selpi, Brouwer, R., Cserpinszky, A., Hagleitner, W., & Benmimoun, M., (2012), *Final Report, euroFOT project deliverable D11.3*. Retrieved from https://www.eurofot-ip.eu/download/library/deliverables/eurofotsp120121212v11dld113_final_report.pdf

Knoll, P., & Langenwieder, K. (2006). *Der Sicherheitseffekt von ESP in Realunfällen – Überlegungen zum volks-wirtschaftlichen Nutzen von prädiktiven Fahrerassistenzsystemen 2*. Tagung Aktive Sicherheit durch Fahrerassistenz.Retrieved from https://mediatum.ub.tum.de/doc/1145167/1145167.pdf

Kocherscheidt, H. (2004). *Wege zur Effizienzbetrachtung von Fahrerassistenzsystemen, 1. Tagung Aktive Sicherheit durch Fahrerassistenz*. Retrieved from https://mediatum.ub.tum.de/doc/1145552/1145552.pdf

Kron, H. H. (2004). *On the evaluation of risk acceptance principles*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.455.4506&rep=rep1&type=pdf

Krumbach, P., & Schnieder, L. (2019). Positive Risikobilanz als ein Zulassungsmaßstab für das Hochautomatisierte Fahren. Safetronic.

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

Larsson, P., Esberg, I., van Noort, M., Willemsen, D., Garcia, E., Fahrenkrog, F., Zlocki, A., Scholliers, J., Koskinen, S., Várhelyi, A., & Schönebeck, S. (2012). Test and evaluation plans, interactIVe project. *Deliverable, D7*, 4.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception, 5*(4), 437–459. https://doi.org/10.1068/p050437

Mai, M. (2017). *Fahrerverhaltensmodellierung für die prospektive, stochastische Wirksamkeitsbewertung von Fahrerassistenzsystemen der Aktiven Fahrzeugsicherheit*. Doctoral Dissertation. TU Dresden, Schriftenreihe des Lehrstuhls Kraftfahrzeugtechnik, ISBN 9783736995482

Malone, K., Wilmink, I., Noecker, G., Roßrucker, K., Galbas R., & Alkim, T. (2008) *Socio-economic impact assessment of stand-alone and co-operative intelligent vehicle safety systems (IVSS) in Europe, eIMPACT project, Deliverable D10*. Retrieved from https://trimis.ec.europa.eu/sites/default/files/project/documents/20130401_140253_77423_eIMPACT_D9_D10_v2.0.pdf

Merat, N., & De Waard, D. (2014). Human factors implications of vehicle automation: Current understanding and future directions. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 193–195. https://doi.org/10.1016/j.trf.2014.11.002

Metz, B., Rösener, C., Louw, T., Aittoniemi, E., Bjorvatn, A., Wörle, J., Weber, H., Torrao, G. A., Silla, A., Innamaa, S., Malin, F., Lehtonen, E., Fahrenkrog, F., Heum, P., Pedersen, K., Merat, N., Nordhoff, S., Beuster, A., Dotzauer, M., & Streubel, T. (2019). *Evaluation method, L3Pilot Project, Deliverable D3.3*. Retrieved from https://l3pilot.eu/fileadmin/user_upload/L3Pilot-SP3-D3.3_Evaluation_Methods-v1.0_DRAFT_for_website.pdf

Najm, W. G., Da Silva, M. P., & Wiacek, C. J. (2000). *Estimation of crash injury severity reduction for intelligent vehicle safety systems, SAE technical paper series 2000-01-1354, SAE 2000 World Congress*.

Najm, W. G., Stearns, M. D., Howarth, H., Koopmann, J., & Hitz, J. (2006). *Evaluation of an automotive rear-end collision avoidance system, U.S. Department of Transportation, Report No. DOT HS 810 569*. Retrieved from https://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/2006/HS910569.pdf

National Academies of Sciences, Engineering, and Medicine. (2014). *Naturalistic driving study: Field data collection*. The National Academies Press.

NHTSA. (National Highway Traffic Safety and Transportation). (2013). *Preliminary statement of policy concerning automated vehicles*. National Highway Traffic Safety Administration.

NHTSA. (National Highway Traffic Safety and Transportation). (2016). *Federal automated vehicles policy. Accerlerating the next revolution in roadway safety*. Retrieved from https://www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20guidance%20PDF.pdf

op den Camp, O., van Montfort, S., Uittenbogaard, J., & Welten, J. (2016). *CATS Final project summary report, CATS project Deliverable 6.1*. Retrieved from https://wiki.unece.org/download/attachments/60360943/AEBS-04-07%20%28NL%29%20TNO%20CATS%20Final%20project%20summary%20report.pdf?api=v2

OpenPass. (2021).OpenPASS Working Group. Retrieved from https://openpass.eclipse.org/

Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 252–263. https://doi.org/10.1016/j.trf.2014.04.009

Page, Y., Fahrenkrog, F., Fiorentino, A., Gwehenberger, J., Helmer, T., Lindman, M., Op den Camp, O., van Rooij, L., Puch, S., Fränzle, M., Sander, U., & Wimmer, P. (2015). A comprehensive and harmonized method for assessing the effectiveness of advanced driver assistance systems by virtual simulation: The p.e.a.r.s. initiative. *The 24th International Technical Conference on the Enhanced Safety of Vehicles (ESV)* (pp. 1–12). Retrieved from http://www-esv.nhtsa.dot.gov/Proceedings/24/isv7/main.htm

PEGASUS. (2019). *Scenario Database, PEGASUS Project, Final Event*. Retrieved from https://www.pegasusprojekt.de/files/tmpl/Pegasus-Abschlussveranstaltung/15_Scenario-Database.pdf

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

P.E.A.R.S. Consortium (2021). Prospective effectiveness assessment for road safety: Overview, UNECE VMAD Sub Group 2 Virtual Testing, Meeting 24th Feb. 2021. Retrieved from https://wiki.unece.org/download/attachments/123666581/VMAD-SG2-12-03%20PEARS%20Presentation.pdf?api=v2

Rödel, C., Stadler, S., Meschtscherjakov, A., & Tscheligi, M. (2014). Towards autonomous cars: The effect of autonomy levels on acceptance and user experience. *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1–8). https://doi.org/10.1145/2667317.2667330

Rösener, C. (2020). *A traffic-based method for safety impact assessment of road vehicle automation*. RWTH Aachen University. Doctoral Dissertation. Retrieved from https://publications.rwth-aachen.de/record/801589/files/801589.pdf

SAE. (2016). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles*. SAE.

Sander, U., & Lubbe, N. (2018). The potential of clustering methods to define intersection test scenarios: Assessing real-life performance of AEB. *Accident Analysis & Prevention, 113*, 1–11. https://doi.org/10.1016/j.aap.2018.01.010

Smith, S., Koopmann, J., Rakoff, H., Peirce, S., Noel, G., Eilbert, A., Yanagisawa, M. (2017). *Benefits estimation model for automated vehicle operations: Phase 2 final report*. U.S. Department of Transportation FHWA-JPO-18-636. Retrieved from https://rosap.ntl.bts.gov/view/dot/34458/dot_34458_DS1.pdf

Spicer, R., Vahabaghaie, A., Bahouth, G., Drees, L., Martinez von Bülow, R., & Baur, P. (2018). Field effectiveness evaluation of advanced driver assistance systems. *Traffic Injury Prevention, 19*(sup2), S91–S95. https://doi.org/10.1080/15389588.2018.1527030

Spitzhüttel, F., Petzold, M., & Liers, H. (2015). *PCM from iGLAD database, FAT-Schriftreihe 280*. Verband der Automobilindustrie. Retrieved from https://www.vda.de/dam/vda/publications/2015/FAT-Reihe/fat-schriftenreihe-280.pdf

Sternlund, S. (2017). The safety potential of lane departure warning systems: A descriptive real-world study of fatal lane departure passenger car crashes in Sweden. *Traffic Injury Prevention, 18*(sup1), S18–S23. https://doi.org/10.1080/15389588.2017.1313413

Sinha, A., Chand, S., Wijayaratna, K. P., Virdi, N., & Dixit, V. (2020). Comprehensive safety assessment in mixed fleets with connected and automated vehicles: A crash severity and rate evaluation of conventional vehicles. *Accident Analysis and Prevention, 142*, 105567. https://doi.org/10.1016/j.aap.2020.105567

TRStrab Brandschutz – Technische Regeln von Straßenbahnen – Brandschutz in unterirdischen Betriebsanlagen, Stand 24.06.2014

Treiber, M., Hennecke, A., & Helbing, D. (2000). Congested traffic states in empirical observations and microscopic simulations. *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, 62*(2), 1805–1824. https://doi.org/10.1103/PhysRevE.62.1805

UN ECE ALKS. (2020). *Proposal for a new UN Regulation on uniform provisions concerning the approval of vehicles with regards to software update and software updates management system*. ECE/TRANS/WP.29/2020/80.

UN ECE R-155. (2020). *Proposal for a new UN Regulation on uniform provisions concerning the approval of vehicles with regards to cyber security and cyber security management system*. UN ECE Regulation. Retrieved from https://unece.org/fileadmin/DAM/trans/doc/2020/wp29/WP29-182-05e.pdf

van Noort, M., Bakri, T., Fahrenkrog, F., & Dobberstein, J. (2015). 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

Wiedemann, R. (1974). *Simulation des Straßenverkehrsflusses*. Schriftenreihe des Instituts für Verkehrswesen der Universität Karlsruhe.

Winner, H. (2015). Quo vadis, FAS? In H. Winner, S. Hakuli, F. Lotz, & C. Singer (Eds.), *Handbuch Fahrerassistenzsysteme* (pp. 1173–1182). Springer Vieweg.

Witt, M., Kompaß, K., Wang, L., Kates, R., Mai, M., & Prokop, G. (2019). Driver profiling: Data-based identification of driver behavior dimensions and affecting driver characteristics for multi-agent traffic simulation. *Transportation Research Part F: Traffic Psychology and BehaviourISSN, 64*, 361–376. https://doi.org/10.1016/j.trf.2019.05.007

Witt, M., Wang, L., Fahrenkrog, F., Kompaß, K., & Prokop, G. (2019). Cognitive driver behavior modeling: influence of personality and driver characteristics on driver behavior. In N. Stanton (Ed.), *Advances in human aspects of transportation. AHFE 2018. Advances in intelligent systems and computing.* (Vol. 786). Springer.

Wood, M., Knobel, C., Garbacik, N., Robbel, P., Boymanns, D., Smerza, D., Maass, M., Löhning, M., Dalong, L., Radboud, D. T., Dehlink, B., Timmons, A., Meijs, M., Kaule, D., Bellotti, Harb, M., Krüger, R., O'Brien, M. et al. (2019). *Safety first for automated driving. White paper of different car manufacutres and suppliers*. Retrieved from https://www.press.bmwgroup.com/global/article/attachment/T0298103EN/434404

Zhao, D., & Peng, H. (2017). *From the lab to the street: Solving the challenge of accelerating automated vehicle testing*. MCity White Paper. Retrieved from https://mcity.umich.edu/wp-content/uploads/2017/05/Mcity-White-Paper_Accelerated-AV-Testing.pdf

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.