Macroscopic Safety Requirements for Highly Automated Driving

Philipp Junietz
Institute of Automotive Engineering
TU Darmstadt
Otto-Berndt-Straße 2
64289 Darmstadt, Germany
junietz@fzd.tu-darmstadt.de

Udo Steininger
TÜV SÜD Rail GmbH
Barthstr. 16
80339 München, Germany
udo.steininger@tuev-sued.de

Hermann Winner
Institute of Automotive Engineering
TU Darmstadt
Otto-Berndt-Straße 2
64289 Darmstadt, Germany
winner@fzd.tu-darmstadt.de

Word count: 7,000 words text + 2 tables x 250 words (each) = 7500 words

Citation:
Junietz, Steininger, Winner: Transportation Research Record SAGE Publications Sage, Los Angeles, CA, Issue 2673(3), pp. 1-10, 2019, DOI 0361198119827910

Submission Date 12/03/2018

Abstract

The common expectation for highly automated vehicles (HAV) is that an introduction will lead to an increased road safety and a reduction in traffic fatalities – at least in relation to the mileage. However, quantizing the safety requirements is still in discussion. This paper analyzes the risk acceptance in other fields and applies the safety level on today's traffic to derive references for acceptable risks. The focus is on macroscopic safety requirements meaning accident rates per mileage and not the behavior in individual driving situations. It is concluded that the acceptable risk varies with the focus group involved and with the field share of automated vehicles. Increased safety of conventional driving in the future could lead to higher requirements as well. We also point out that it is not guaranteed that the given acceptable risk levels are also accepted by the customer because other factors besides the accident statistics are relevant. However, as none of these risk levels can be proven before introduction, a monitoring of vehicles in the field is suggested. Despite increased efforts in the research of safety validation, an uncertainty of the safety of HAV will remain at the time of introduction. Different introduction and risk management strategies are briefly introduced.

Keywords: Automated Driving, Safety Requirements, Risk Analysis

1 Contribution of this Paper

With the development of highly automated vehicles of SAE level 3 and higher (HAV), the issue of a valid safety approval is discussed recently. Following a requirement-based development process, the fundamental safety requirements should be defined a priori. The common expectation is that the introduction of HAV will reduce the number of accidents at least long-term. The *Vision for Safety* program of the National Highway Traffic Safety Administration (NHTSA) of the United States relies on the expected safety benefit of automated driving. (*1*) However, more skeptical voices also point out the risks of the new technology. The Ethics Commission on Automated and Connected Driving of the German Federal Ministry of Transport concluded that "*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*" (*2*). At the same time, we know that a statistical proof of superior safety cannot be performed without introducing HAV to the market (*3–6*). There is the EU policy of *Visio Zero* (*7*), which demands zero victims in traffic, but there are also concepts that promote an introduction of automated vehicles because with the knowledge and the recordings from the mileage recorded in the market, the systems could be improved faster and lives saved (*8*). We also know that the acceptance of risk depends on the individual benefit (*9*).

These statements raise some questions:

- How can this "*promise of increased safety"* be monitored and measured?
- What needs to be done before an introduction to the market?
- What are quantitative safety requirements from the different stakeholders?

To address those questions, quantitative safety requirements are deduced in the following sections. The HAV system in this paper is designed for controlled-access highways. We start with a review of existing studies that compare accepted risks from different technologies influenced by different types of exposition. The goal is to find theoretically acceptable quantitative values that can be transferred to HAV. Afterwards, consequences for the introduction of HAV are concluded.

However, we point out that there is currently no guarantee that those values are accepted in the end, as this depends on many factors, not least the response in the media. As the occurrence of accidents is not always a question of driving skill but often coincidence, the accident occurrence at the beginning of the introduction is crucial for acceptance. If the very unlucky case of a fatal accident in the first days of public driving occurs, communication in the media will be decisive. The resulting risk figures are acceptable from a scientific viewpoint but not necessarily accepted for a final product.

In the end, we discuss how to handle the presented requirements. As proof of safety before introduction of the system is unlikely, the two main factors of an introduction will be the observation of all incidents and the deployment of necessary updates, whenever a safety issue is detected.

2 Fundamental Safety Requirements and Current Road Safety

2.1 Motivation

The recent publication of the German Federal Statistical Office (*10*) again reports an increase in average life expectancy of newborns. Since the recording of mortality tables began in 1871, we observed a doubling of the average life expectancy. One should assume that people are very lucky with this development. But a look into daily media reporting shows that people are not only very skeptical about technical achievements but they are even afraid of effects that are obviously responsible for the aforementioned increase in life expectancy - for example medical and agricultural advances. People are concerned, for example, about man-made radiation and air pollution by industrial plants or transportation but also about the side-effects of medicine, the consumption of meat of uncontrolled origin, bacteria in green salad, dioxin in free-range eggs, etc.

Nevertheless, the facts speak for themselves: The most common natural causes of death in Germany - and this is representative for industrial countries - are cardiovascular diseases with 39%, followed by cancer with nearly 25% and, well behind, by diseases of respiratory and digestive system with 7% and 4%. It is interesting to note that non-natural causes of death, i.e. mainly suicides and accidents, contribute only 4% (*11*). From the medical point of view there is no doubt regarding the factors that really kill us - smoking, overweight, high blood pressure, diabetes, and physical inactivity. Everybody of us is able to control those factors and to prevent the consequences but why are we not doing this consistently? Moreover, why are we so concerned regarding other factors that are less risky but not readily controllable by ourselves? Why do our risk perception and risk acceptance seem contradictory?

There is obviously a discrepancy between objectively existing risks on the one hand and their perception and acceptance by individuals as well as by society on the other. It is important for that purpose, among other things, whether people enter into the risk voluntarily or not, whether they feel a personal benefit, and whether the risk is natural or synthetic. Moreover, risk perception depends on risk communication (*12; 9; 13*).

We have to ask ourselves whether it is even possible to deal with risk in an objective manner. What the consequences of the described difficulties with risk perception and acceptance are for the introduction of new, complex technologies like for example highly automated driving.

2.2 Quantitative Risk Assessment

The usual quantitative risk definition “risk equals frequency times severity” is illustrated in FIGURE 1. Considering occurrence of unintended events (frequency) and extent of damage (severity), we find two typical areas. In the green area, the system is in a safe state; the corresponding risk is accepted. In the red area, the system is in an unsafe state; the corresponding risk is not accepted. The borderline between these areas is probably not sharp; there can be a kind of transition area.

Although this simple definition is very useful for many questions in technology and insurance industry, it neglects aspects like aversion against high severity, lack of controllability, and personal benefit, which are relevant for risk perception and acceptance by individuals and society. Intensive research on risk perception and risk acceptance started in the second half of the last century. Different authors analyzed risk acceptance and risk-benefit constellations in various studies (*14–21*). Fritzsche discussed risk acceptance relating to voluntary nature of exposure based on the studies (*12*). Slovic concludes similar numbers in a more recent, updated publication (*13*). The results are summarized in FIGURE 2. It is interesting that both authors conclude similar risk numbers despite the major gap of several decades. The reason might be that risk perception studies reached their peak in the 70’s with the introduction of nuclear power. We suggest correcting the numbers with a factor derived from the change in mortality rate, as explained in the next sections.

For voluntary activities, Fritzsche found that the willingness to accept risks is nearly unlimited, depending on the experienced personal benefit. We can see this by the example of high-risk sport or other leisure activities, e.g. free climbing, motorcycling etc. Job-related activities are important for a deeper understanding of the subject. Acceptance is relatively well investigated in this field and there is a common understanding of accepted individual mortality risk in the order of 10^{-5} per person and year, for example by professional associations and insurance companies, on the one hand. On the other hand, job-related risks are useful to bridge the gap between voluntary and involuntary risks.

Fritzsche found that for involuntary risks, e.g. death of passengers due to a train or airplane crash, the acceptance level is an order of magnitude lower than for job-related risks. Moreover, acceptance decreases another order of magnitude if the risk is caused by major technology, e.g. chemical industry or nuclear power generation. Beside the fact that the experienced personal benefit of those technologies is low (at least from a subjective point of view), the low degree of self-determination or rather controllability by individuals plays an important role for the low acceptance level as well as the potentially high number of mortalities (severity). Nevertheless, FIGURE 2 shows that it is generally possible to deal with risk, risk perception, and acceptance in a quantitative manner.

To implement safety requirements based on risk acceptance, several concepts have been developed in different application areas. Because of the relationship between railway and road traffic, it is useful to refer to the CENELEC safety standard EN 50126. The development of this standard has been started in the 1990’s, where safety requirements based on quantitative risk analysis have been implemented and ALARP, MEM, and GAMAB have been introduced as principles for risk acceptance. Those principles shall be shortly explained in the following clauses.

2.2.1 As low as reasonably practicable (ALARP)

ALARP tries to assess what is technically feasible considering economic sense and social acceptance. Between the two regions of generally unaccepted and broadly accepted risk, there is a tolerance range where risk is undertaken only if a benefit is desired and where each risk must be made as low as reasonably practicable.

This is not applicable because EN 50126 failed to give certain values for generally unaccepted and broadly accepted risk. However, other authors, for example Risk & Reliability Associates, deliver both values (*22*): The two key levels seem to lie around road death statistics (about 10^{-4} per person and year) and the chances of being struck by lightning (about 10^{-7} per person and year). If something is more dangerous than driving a car, the risk is unacceptable. If something is less dangerous than being struck by lightning, then we do not expect anyone to do anything about it. In the range between these two figures, cost benefit studies are appropriate to reduce the risk to as low as reasonably practicable. Especially this lower ALARP limit corresponds very well with the acceptance criterion for major technology risks shown in FIGURE 2.

2.2.2 Minimum endogenous mortality (MEM)

MEM is based upon age- and gender-specific mortality rates (*10*). Although the absolute values of the mortality rates change with birth cohort, they show a typical development over age as well as a significant minimum at an age of about 10 years. The related mortality at an age of 10 years is defined as “minimum endogenous mortality”. The MEM principle demands that a new system does not significantly contribute to the existing minimum endogenous mortality.

EN 50126 specifies that the individual risk due to a certain technical system must not exceed 1/20th of the minimum endogenous mortality, taking into account that people are normally exposed to the risk of several technical systems. This means that the accepted individual risk of a certain technical system should be about $2.5 \cdot 10^{-6}$ per person and year, when using latest mortality rates as a basis (EN 50126 uses mortality rates from the 80’s). This value corresponds very well with the acceptance criterion for involuntary risks shown in FIGURE 2.

2.2.3 Globalement au moins aussi bon (GAMAB – English: generally at least as good as)

GAMAB, (or GAME globalement au moins équivalent), requires, unlike MEM, the existence of a reference system with – currently – accepted residual risks. According to GAMAB, residual risks caused by a new system must not exceed those of the reference system. In other words: a new system must offer a level of risk generally at least as good as the one offered by any equivalent existing system. This makes it necessary to identify the risk of an equivalent existing system.

Looking for the acceptable risk of highly automated driving in a certain application area according to GAMAB, we have to identify the current risk of the equivalent existing system in the same application area. To derive acceptance requirements for a controlled-access highway pilot, we analyze the current risk on German controlled-access highways during manual driving. TABLE 1 shows average distances between two accidents referring to severity levels according to ISO 26262 (*23*).

TABLE 1 Accidents on German controlled-access highways

FIGURE 1 shows the observed accident rates versus severity levels according to ISO 26262 (*23*). Comparing risks of the different severity categories requires weighting of the different levels. However, there is no standardized way (see also (*24; 25; 5*)). FIGURE 1 assumes that the difference between adjacent severity levels is one order of magnitude, or in other words that an accident with fatalities is ten times worse than an accident with severe injuries. This assumption allows to define a band of constant risk in current traffic, which can be used as reference. As already discussed in section 2.2.1, the risk will not be accepted above the upper envelope. Beyond the lower envelope, the risk might be accepted. Between both lines is a transition area. In accordance with the aforementioned ALARP principle, this is a tolerability region where risk is undertaken if a benefit is desired and where each risk must be made as low as reasonably practicable.

In FIGURE 2, the results of the application of the different risk acceptance principles are displayed related to the risk acceptance limits of the different expositions explained above.

FIGURE 1 Left: Illustration of risk; Right: Quantitative accident risk on German highways (*26; 23*)

In FIGURE 2, the different approaches for the mortality risk are summarized. On the one hand, it shows that the application of different risk acceptance principles delivers comparable and consistent results. On the other hand, it demonstrates that we have to deal with a relatively broad range of applicable acceptance criteria.

Taking into account the impact of voluntary exposure, different groups of users have to be distinguished, for example users of highly automated driving systems and other traffic participants. Finally, comparison with other technologies

– especially other traffic systems and technologies that deliver a high personal benefit – seems to be useful. Additionally, a decrease in total mortality risk is expected in the future, following the trend in the last decades and centuries. Therefore, risk acceptance might change over time.

FIGURE 2 Application of different risk acceptance principles to highway accidents, translated from (*27*), based on (*12*), GAMAB is based on the risk on German controlled-access highways.

2.3 Introduction of New Technologies in Aviation

Let us take a digression to aviation. Here, passengers are exposed to a technical system without having personal control. Although severe accidents happen, its safety is accepted by most of the population. Aviation has become increasingly automated in the past (although today's systems are still SAE level 2 because they are supervised by the crew). Due to the long travelling distance and the fact that accidents mostly happen during take-off and landing, accident rates are typically given per flight and not per travel distance. Accidents and critical situations are strictly reported and collected in a database, so we have even more profound data compared to road traffic. Depending on the number of flights per year, we can observe an annual risk that is similar to driving a car on a highway. One fatal accident happens about once per ten million flights (*28*). With a typical exposure of two flights per year, the risk of a fatal accident would be lower than the risk of involuntary exposure f_{inv} and about one order of magnitude lower than driving on a highway. However, with 20 flights per year, one would be exposed to a risk that is in the same order of magnitude. Therefore, the levels of risk are in fact comparable if only driving on controlled-access highways is considered. However, typically users drive on all types of roads. The risk of car traffic is at least on order of magnitude higher in total, so the superior reputation of air traffic is justified.

As mentioned before, aviation has become increasingly automated over the past decades. The detailed collection of data in aviation allows an analysis per generation of airplanes, which was summarized by Airbus Industries (*28*). As depicted in FIGURE 3, with every introduction of a new generation, the fatal accident rate for this new generation was higher than state of the art. Due to the low number of new airplanes at introduction, this trend cannot be observed in the total accident rate (comp. (*28*)). Nevertheless, the introduction was clearly beneficial to society in total because after an introduction phase of five to ten years, the new generation had the lowest accident rate of all.

Judging from this data, new generations of airplanes are not tested in a way to prove statistically that the system is superior to the former. In fact, this is impossible; because the knowledge about the new system's behavior is incomplete and only field experience can reduce the unknowns. Similar to HAV, statistical testing is neither economically feasible nor necessary because the strict supervision of air traffic allows efficient improvement in case of critical situations or accidents. However, the highest automation in commercial air traffic is still comparable to level 2, so human error is still a factor. Nevertheless, the leap in accident rate occurred with the introduction of technology, be it because of flaws in human-machine-interaction or in the technology itself. One could argue that it is unethical to release a system that is not tested in the best way possible. The authors would argue the opposite. First, it is impossible to completely test a system operating in an uncontrolled environment because there might be situations that the tester was not aware of. These "unknown unknowns" cannot be tested. Second, a stricter approval process would prevent technical progress because a profit-oriented development would become impossible. It seems possible if not likely that the accident rate of automated vehicles will behave in a similar way. We should be aware of that possibility and focus on the improvement of the system in case of a detected critical situation or accident. A similar thought is also expressed in (*8*). The delayed introduction of HAV could in fact risk the lives of many people because the system is believed to improve safety over time.

FIGURE 3 Fatal accidents with different generations of airplanes in commercial traffic. Dotted line means less than one million flights a year. First generation: Early commercial jets, Second generation: More integrated Auto Flight System, Third generation: Glass cockpit and Flight-Management-System, Fourth generation: Fly-By-Wire with flight envelope protection. (*28*)

With a combined testing strategy of simulation, proving ground tests, and real traffic tests, it is still unlikely to complete a logical proof of safety because every validation test has certain underlying assumptions. In order to deal with this uncertain safety performance, accidents, unexpected critical situations, and near misses must be monitored similar to air traffic, in order to find flaws in the system (including infrastructure and human interaction) with the chance to improve them. This is discussed in detail in section 4.

3 Safety Requirements of Different Focus Groups

In general, an automated vehicle is a risk for different focus groups. The first two groups are the users of the vehicle and the potentially involved accident partner, who can be any individual traffic participant. Grunwald (*9*) explains that the reason for the different views on accepted risk of the two groups results from the benefits the groups get. The third group is the society. Different from the first two groups, the fate of an individual is not relevant for society but the total accident number is. In this paper, we only discuss the occurrence of fatalities (index d), so instead of risk, we give quantitative requirements for the occurrence rate or frequency of fatal accidents. Quantitative requirements for different types of usage are given in FIGURE 2. It is concluded that the accepted frequency for a person's death per year f_{inv} is $10^{-6}\ \mathrm{k_d/a}$ for involuntary exposure, $10^{-5}\ \mathrm{k_d/a}$ for professional exposure (f_{prof}), and a theoretically unlimited risk for voluntary exposure with typical acceptance rates f_{vol} of up to $10^{-2}\ \mathrm{k_d/a}$. In general, the accepted risk varies with the benefit for the user or focus group. Most of these considerations are from the 70's and 80's regarding the discussions on safety of nuclear power plants. However, similar to MEM, the assumptions are still valid in general, but should be adapted to today's level of safety. The authors suggest a factor of one forth, similar to the development of MEM. (Reduction by a change in accident rate would be another approach with a similar outcome.) In the following, the lower risk today compared to the numbers above is indicated by its index with year and country of the underlying statistic.

3.1 User

The fatal risk for the user is assumed equivalent to the risk of a fatal accident of a HAV (neglecting a higher damage with more than one user at the same time). As depicted in FIGURE 2, the type of exposition is relevant for accepting risks. In most use cases, HAV functions are used voluntarily; they must be actively bought and activated. Professional use is also plausible but the use for the job is not expected during the first introduction phase. Involuntary use is excluded in typical use cases. This consideration suggests following the risk acceptance rate of $f_{\text{prof, 2016, GER}}$ equal to $1.4\cdot 10^{-9}\ \mathrm{k_d/km}$. Similar rates are also present in the US ($2.5\cdot 10^{-9}\ \mathrm{k}_d/\mathrm{km}$) (*29*). For other countries, data about the mileage on different road types is not always available. The accident rate on all roads' combined mileage is in a similar order of magnitude for most developed countries. (*30; 31*)

However, the substitution of conventional driving also suggests comparing the risk of today's driving with the suggested rate of f_{prof}. In the following, both considerations will be examined and compared. For today's driving risk, driving on Autobahn in Germany will be taken as a reference. This has several advantages. First, driving on controlled-access highways is one of the safest, if not the safest way of travelling in a car, especially when taking the accident rate per mileage as reference. Second, it is likely that the first HAV will drive on a controlled-access highway. Third, accident data on highways are well documented. Even minor accidents often result in the involvement of police because of the traffic disturbance and the measured traffic density estimates the travelled distance. Assuming an average travel distance $\bar{d}$, the time-based frequency (index t) can be transmitted to a distance-based frequency (index s) and vice versa. In this example, 4000 km/a are assumed as an average travel distance according to (*32*) and assuming an average velocity of 100 km/h.

In the following equation (1), the GAMAB principle and the MEM principle are combined. We consider this the upper limit for tolerable frequency because a new technology is introduced that comes with new risk. Additional risk is acceptable because the user experiences a benefit from that new technology. Note that we only consider the risk for the user in this section. This is not applicable to non-users or society at all as a whole, what will be discussed in the following sections.

$$
\begin{aligned}
&k_{d,\text{User}} \leq k_{\text{GAMAB}} + k_{\text{MEM}/20} \\
&\Longrightarrow f_{\text{t,d,User}} \leq f_{\text{s,d,gamab}} \cdot \bar{d} + f_{\text{t,MEM}/20} \\
&\Rightarrow f_{\text{s,d,User}} \leq f_{\text{s,d,gamab}} + f_{\text{t,MEM}/20}/\bar{d} \\
&f_{\text{s,d,User,2016,GER}} \leq 2.15 \cdot 10^{-9}\,\frac{\mathrm{k_d}}{\mathrm{km}};\ \ f_{\text{t,d,User,2016,GER}} = 8.6 \cdot 10^{-6}\,\frac{\mathrm{k_d}}{\mathrm{a}}
\end{aligned}
\tag{1}
$$

Interestingly, the order of magnitude according to equation (1) corresponds to the accepted frequency for professional exposure. This strengthens the hypothesis that both estimations result in acceptable values for users of automated vehicles. However, higher risk could be accepted by the user (similar to motorbikes or extreme sport) but the user should be aware of this potentially increased risk.

3.2 Passers-by

For all other traffic participants, the HAV has no direct benefit (besides the decreased total risk for all traffic participants assuming that the HAV is safer than the average driver). However, non-users could have a lower risk acceptance threshold because they are skeptical about the new technology or might even have (subjective) disadvantages e.g. due to slow vehicles on the road. Extraordinarily critical is the risk of new types of accidents (comp. (*33*)) because non-users would blame HAV for those accidents despite a potential reduction of the total number. New risks could be caused for example by systematic software failures or cyber-attacks. The total new risk of the technology for an individual non-user should be below $f_{\mathrm{inv,2016,GER}}$ equal to $2.5 \cdot 10^{-7}\,\mathrm{k_d}/\mathrm{a}$.

So how can the individual risk for a non-user be calculated? As long as there are not many HAVs on the market, the exposure is very low and the probability that the individual traffic participant is involved in a HAV's accident is low. So, the risk is multiplied with the field share μ. The risk for passers-by is diluted by the exposure to vehicles equipped with HAV.

$$\begin{aligned} &f_{\mathrm{t,d,new}} \cdot \mu \leq f_{\mathrm{inv,2016,GER}} = 0.25 \cdot 10^{-6}\,\frac{\mathrm{k_d}}{\mathrm{a}} \\ &\Leftrightarrow f_{\mathrm{s,d,new}} \leq 6.25 \cdot 10^{-11}\,\frac{\mathrm{k_d}}{\mathrm{km}} \cdot \frac{1}{\mu} \end{aligned} \tag{2}$$

According to equation (2), the accepted risk for a single HAV is lower with increasing number of HAV. This is intuitively obvious because the exposure multiplies with the number of potential single threats. Comparing the risk level with equation (1) results in a number of $1.625 \cdot 10^6$ HAV in Germany, until the risk acceptance of the other traffic participants becomes dominant. This deliberately neglects that the non-user also has benefits if the system is safer than the human driver it replaces. The authors believe that this will only be acknowledged by non-users if there is an undeniable difference in accident statistic. Otherwise, the (subjective) disadvantage of the new technology stays dominant.

3.3 Society

For society, the fate of individuals is of lesser importance. Benefits and costs of HAV are measured by the total number of accidents and whether they are reduced over time. In general, a decreasing trend of accident rate throughout the years can be observed in Germany (*34*) and the US (*35*). However, we can observe that this trend has been diminishing over the last 5 years for accidents with injuries and even had a slight (but insignificant) increase during these years. For fatal accidents, this trend of a more slowly decreasing rate is observable as well, but less significant. One could suspect that there is a natural limitation with current road network, traffic density, and state-of-the-art vehicles.

When introducing HAV, there will still be a non-zero risk of severe accidents and therefore it is likely that HAV will be involved in those severe or even fatal accidents. So, what are the requirements by society if individual accidents do not influence the total number significantly?

What is the upper total accident rate limit accepted by society?

The overall target is to reduce the amount of accidents over time with the introduction of new technology. If we follow the argumentation of Wachenfeld (*5*) and Kalra (*8*), we should allow a certain risk in order to bring HAV to the market and allow to gain further knowledge. At the same time, it is not acceptable for the whole society that the total risk is increased in a noticeable way.

However, there is no way to check how accident numbers would have evolved without the technology as soon as it has entered the market. Wachenfeld interpolates the accident numbers of the years 1992-2014 and suggest a standard deviation of 39 fatal accidents per year as a reference (*5*) for a maximum deviation caused by HAV. However, in the last decade, the decrease of fatal accidents and accidents with injuries diminished. At the same time, the annual travel distance increased. Hence, it seems justified to use recent numbers as reference. When using the accident rate for fatal accidents $f_{\mathrm{s,d}}$, an exponential regression is a better fit than a linear regression. Interestingly, this is also the case for accidents in aviation (comp. (*28*)). The standard deviation for the exponential regression for all years since 2010 results in:

$$\sigma_{\mathrm{7y,exp}} = \sqrt{\frac{1}{N_{\mathrm{year}}} \sum_{i=2010}^{2016} \left(f_{\mathrm{s,d,i}} - f_{\mathrm{7year,exp}}(i) \right)^2} = 9.4 \cdot 10^{-11}\,\frac{\mathrm{k_d}}{\mathrm{km}} \tag{3}$$

Multiplying the standard deviation with the average annual mileage in 2016 results in 22.9 fatal accidents per year, which is only slightly lower than what Wachenfeld calculated. However, it must be pointed out that the type of regression and the number of years influence the result. It is also possible to use the double or triple standard deviation as a measure. However, the results will be in a similar order of magnitude. In the following, the result from equation (3) will be used.

The requirements by society should be that the risk from HAV is significantly lower than the described exponential trend observed in the latest data, so HAV should be at least one standard deviation $\sigma_{7y,exp}$ better than the predicted performance of conventional driving. However, society should give HAV time to reach this high safety reference. Similar to air traffic, it is necessary to monitor the performance to allow improvement in functions, infrastructure, and user experience. In the following formula, it is suggested to allow additional risk of one standard deviation at the beginning of introduction and demand a risk three standard deviations lower than the extrapolation, when full market share is reached. Therefore, the acceptable risk not only depends on the development of the risk in conventional traffic over the years, but also on the market share of HAV μ.

$$\begin{aligned} &f_{\mathrm{s,d,soc}}(t)\cdot\mu + f_{\mathrm{7y,exp}}(t)\cdot\big(1-\mu(t)\big) \le \left(f_{\mathrm{7y,exp}}(t)+\sigma_{\mathrm{7y,exp}}\right)\cdot\big(1-\mu(t)\big) + \left(f_{\mathrm{7y,exp}}(t)+3\cdot\sigma_{\mathrm{7y,exp}}\right)\cdot\mu(t) \\ &\Leftrightarrow f_{\mathrm{s,d,soc}}(t) \le f_{\mathrm{7y,exp}}(t) + \sigma_{\mathrm{7y,exp}}\frac{1-\mu(t)}{\mu(t)} - 3\cdot\sigma_{\mathrm{7y,exp}} \end{aligned} \tag{4}$$

In the following, a field share μ is assumed that develops similar to the field share of other driving functions such as electronic stability control (comp. (*5*)). Full field share is assumed to be reached after 30 years and described by a sine function $(1+\sin\pi\cdot t/T)/2;\ 0\le t\le T,$ see FIGURE 4. But other parameters are also time dependent because the actual safety on the roads is expected to change over time, even without HAV.

3.4 Summary of Safety Requirements

In the previous sections, safety requirements for the three different stakeholders were deduced. For society, the acceptable risk depends on the market share of HAV. The authors suggest to allow an increase in total risk by one standard deviation of the predicted accident rate, so HAV can be introduced although the knowledge about its safety level is not yet complete. In addition to society's requirements, passers-by (as part of society) have increased requirements for new risks that come with automation. For users, we suggest constant risk requirements although they might increase with the current traffic safety over the years. However, the user's requirements are only dominant in the early introduction phase (comp. FIGURE 4) when the market share is relatively low. From a market share of about 10% on, the requirements of society (and non-users) are dominant. However, if the field share reaches 100%, there are no non-users remaining.

FIGURE 4 Safety requirements

In the following table, the requirements are summed up. Note that we currently only give values for Germany, because statistics about mileage on controlled-access highways are available. Since data for the accident rate on the whole road network is in the same order of magnitude for developed countries (see above), we do not expect significant changes in safety requirements.

TABLE 2 Summary of Safety Requirements

4 Introduction and Testing Strategy

The results from last section emphasize the difficulty of proving safety before introduction. In this section, an alternative introduction and testing strategy is presented briefly.

4.1 Test Strategy and Requirements for Technical Systems

Safety requirements for individual HAV systems or vehicles, respectively, cannot be directly derived from criteria for individually (MEM) or socially accepted risk (GAMAB). Otherwise, we are allowed to take credit from development of those systems according to ISO 26262:2011. Fulfillment of the standard ensures absence of unreasonable risk. I.e. risk, judged to be unacceptable in a certain context according to valid societal moral concepts. ISO 26262:2011 deals with hazardous events caused by malfunctioning behavior of E/E systems. Immediately after emission of the standard in November 2011, ADAS-related hazards caused by normal operation of the systems (i.e. without malfunctioning behavior) have been addressed in the discussions between safety experts. During the activities for the 2nd edition, which started in January 2015, the responsible working group ISO TC22/SC32/WG08 decided to develop a publicly available specification ISO/PAS 21448 as a separate specification for Safety of the intended function (SOTIF), which addresses the nominal performance in order to get a safe function. SOTIF specification deals with hazardous events without any malfunctioning behavior of E/E systems. However, product development according to those specifications does not replace validation and the development and validation of product tests. An evaluation of different test strategies can be found here (*36*).

4.2 Limited Introduction and Field Observation

Despite all consideration, normative specification, and product tests, there will be an uncertainty about the future performance of HAV at the time of the initial introduction to the market. This uncertainty either can be accepted if all stakeholders agree that the residual risk is sufficiently small, or controlled by reducing the number of sold vehicles in the introduction phase. (*36*) This last concept is also called risk-limited introduction. (*5*) The field share μ is controlled and the expose for the society and non-user reduced. Only the user has to accept the uncertainty if he wants to use HAV. However, the performance of HAV should be observed and statistics publicly discussed to build trust in customers and the society, but also to identify critical situations and improve the system with updates.

5 Conclusion

The introduction of HAV will probably cause a paradigm shift in traffic as we know it and it is likely that the distribution of accident types will change as well. While a reduction of fatal accidents is obviously beneficial, other changes in statistics might not be obviously right or wrong.

On the one hand, HAV have a high potential to reduce risk by avoiding accidents and reducing their consequences. As discussed before, related expectations have to be derived under consideration of

- socially accepted risk in general (GAMAB),
- several focus groups with different personal benefits, and
- time-dependent effects like field penetration and changes of social acceptance related thereto.

On the other hand, HAV will generate automation risks because of performance limitations and inadequate interaction with drivers and other traffic participants. Those risks have to be orders of magnitude lower in comparison to individual risk in a way that there can be no doubt about a positive risk balance – in analogy to the situation during introduction of safety belts or airbags.

The fulfilment of those safety expectations is supposed as a prerequisite for social acceptance in general and for cooperativeness of legislators, regulators, and standardization bodies for further development of legislation, regulation and standardization. A suitable market observation is necessary to prove, whether traffic safety develops itself as expected. Technical and organizational prerequisites need to be put into place, therefore.

Despite those measures, the future accident rate is unknown and cannot be predicted without high uncertainty. While the impact on safety by the accident rate of HAV was discussed assuming a certain market share of HAV, changes in the surrounding traffic are difficult to estimate. However, a statistic proof is only feasible after start of production with series cars, so we should not directly demand proof for the described requirements. A software update (which can also be a mandatory update) might be the only economically feasible way to apply updates onto a large fleet. In order to monitor whether the requirements for society are fulfilled, the total accident numbers should be monitored in addition to the accidents of HAV.

6 Author Contribution Statement and Acknowledgement

All authors were involved in the conception and the draft of this paper and approve the final version to be published. The work for this paper was part of the project PEGASUS funded by the German Ministry for Economic Affairs and Energy.

7 References

1. *A Vision for Safety 2.0,* U.S. Department of Transportation NHTSA, 2017. https://www.nhtsa.gov/document/automated-driving-systems-20-voluntary-guidance. Accessed July 17, 2018.
2. Federal Ministry of Transport and Digital Infrastructure. *Report of the Ethics Commission Automated and Connected Driving*, 2017. Accessed December 8, 2017.
3. Kalra, Nidhi, Paddock, and S. M. *Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?* http://www.rand.org/pubs/research_reports/RR1478.html. Accessed April 15, 2016.
4. Wachenfeld, W., and H. Winner. The Release of Autonomous Vehicles. In *Autonomous Driving. Technical, Legal and Social Aspects,* M. Maurer, J.C. Gerdes, B. Lenz and H. Winner, eds. Springer, 2016, pp. 425–449.
5. Wachenfeld, W. H. K., Dissertation, Technische Universität Darmstadt, 2017.
6. Winner, H., and A. Weitzel. Die Freigabefalle des autonomen Fahrens. In *Mensch und Fahrzeug*, Darmstadt, 2011, p. 10.
7. Tingvall, C., and N. Haworth. Vision Zero. An ethical approach to safety and mobility. In *6th ITE International Conference Road Safety & Traffic Enforcement: Beyond*, 2000, pp. 6–7.
8. Kalra, N., and D. G. Groves. The Enemy of Good. Estimating the Cost of Waiting for Nearly Perfect Automated Vehicles, 2017.
9. Grunwald, A. Societal Risk Constellations for Autonomous Driving. Analysis, Historical Context and Assessment. In *Autonomous Driving.* Springer, 2016, pp. 641–663.
10. *Kohortensterbertafeln für Deutschland. Methoden- und Ergebnisbericht zu den Modellrechnungen für Sterbetafeln der Geburtsjahrgänge 1871 – 2017*, 2017. https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Bevoelkerungsbewegung/Kohortensterbetafeln5126101179004.html.
11. Statistisches Bundesamt (Destatis). *Pressemitteilung Nr. 022*, 19.01.2017.
12. Fritzsche, A. F. *Wie sicher leben wir?* Verlag TUV Rheinland, 1986.
13. Slovic, P. *The perception of risk.* Earthscan, London [u.a.], 2011.
14. Crouch, E. A. C., and R. Wilson. Risk/benefit analysis, 1982.
15. Douglas, M., and A. Wildavsky. How can we know the risks we face? Why risk selection is a social process. *Risk analysis*, Vol. 2, No. 2, 1982, pp. 49–58.
16. Gibson, S. B. Risk criteria in hazard analysis. *Chemical engineering progress*, Vol. 72, No. 2, 1976, pp. 59–62.
17. Kinchin, G. H. Design Criteria, Concepts and Features Important to Safety and Licensing. ANS. In *ENS International Meeting on Fast Reactor Safety Technology, Seattle, Washington (19 -23 August, 1979)*, 1979.
18. Kletz, T. A. Hazard analysis, its application to risks to the public at large. *Reliability Engineering*, Vol. 3, No. 4, 1978, pp. 325–338.
19. Starr, C. Social benefit versus technological risk. *Science*, 1969, pp. 1232–1238.
20. Starr, C. Benefit-cost relationships in socio-technical systems. In *Environmental aspects of nuclear power stations*, 1971.
21. Webb, G. A., and A. S. McLean. *Insignificant levels of dose. A practical suggestion for decision making,* National Radiological Protection Board, 1977.
22. Risk & Reliability Associates Pty Ltd, Consulting Engineers. *Risk and Reliability - An Introductory Text.* Risk and Reliability Associates Pty Ltd, 2004.

23. Steininger, U., H.-P. Schöner, M. Schiementz, and J. Mazzega. *Validation of Assisted and Automated Driving,* Munich, April 19-20, 2016.
24. Baum, H., T. Kranz, U. Westerkamp, and B. für Strassenwesen. *Volkswirtschaftliche Kosten durch Straßenverkehrsunfälle in Deutschland.* Wirtschaftsverl. NW, Verlag für neue Wiss, 2010.
25. Hydén, C., Lund Institute of Technology. Department of Traffic Planning and Engineering, 1987.
26. Schöner, H.-P. *Challenges and Approaches for Testing of Highly Automated Vehicles,* Paris, 04.12.2014.
27. Steininger, U., and L. Wech. Wie sicher ist sicher genug? Sicherheit und Risiko zwischen Wunsch und Wirklichkeit. *VDI-Berichte, No.* 2204, 2013.
28. Airbus. *Commercial Aviation Accidents 1958-2016. A Statistical Analysis*, 2017. Accessed August 1, 2017.
29. *FATALITY RATE PER 100 MILLION ANNUAL VMT - 2013,* U.S. Department of Transportation Federal Highway Administration. https://www.fhwa.dot.gov/policyinformation/statistics/2013/pdf/fi30.pdf. Accessed July 17, 2018.
30. Oguchi, T. Achieving safe road traffic—the experience in Japan. *IATSS research*, Vol. 39, No. 2, 2016, pp. 110–116.
31. *Comparison of 2013 VMT Fatality Rates in U.S. States and in High-Income Countries,* U.S. Department of Transportation NHTSA, October 2016. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812340. Accessed July 17, 2018.
32. *VDA 702 Situationskatalog E-Parameter nach ISO 26262-3. VDA-Empfehlungen,* VERBAND DER AUTOMOBILINDUSTRIE E. V. (VDA), 2015. https://www.vda.de/de/services/Publikationen/situationskatalog-e-parameter-nach-iso-26262-3.html.
33. Gasser, T. M. Fundamental and Special Legal Questions for Autonomous Vehicles. In *Autonomous Driving: Technical, Legal and Social Aspects,* M. Maurer, J.C. Gerdes, B. Lenz and H. Winner, eds. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016, pp. 523–551.
34. Statistisches Bundesamt (Destatis). Verkehrsunfälle - Fachserie 8 Reihe 7 - 2015, 2015.
35. *Fatality Analysis Reporting System,* U.S. Department of Transportation NHTSA, 2017. https://www-fars.nhtsa.dot.gov/Main/index.aspx. Accessed July 18, 2018.
36. Junietz, P., W. Wachenfeld, K. Klonecki, and H. Winner. Evaluation of Different Approaches to Address Safety Validation of Automated Driving. Accepted Paper. In *2018 Intelligent Transportation Systems Conference (ITSC)*, 2018.

TABLE 1

Severity	**ISO 26262 Severity level**	**Average distance between two accidents of this level**	**Accident rate per driven distance**
Fatal	S3	$660 \cdot 10^6$ km	$1.52 \cdot 10^{-9}$/km
Severe Injuries	S2	$53.2 \cdot 10^6$ km	$1.88 \cdot 10^{-8}$/km
Injuries	S1	$12.5 \cdot 10^6$ km	$8.00 \cdot 10^{-8}$/km
w/o Injuries	S0	$7.5 \cdot 10^6$ km	$1.33 \cdot 10^{-7}$/km

TABLE 2

Description	**Symbol**	**Value based on German data from 2016**

User requirements		
Per Distance	$f_{\mathrm{s,d,User}}$	$2.2 \cdot 10^{-9}\,\mathrm{k_d/km}$
Per Time	$f_{\mathrm{t,d,User}}$	$8.6 \cdot 10^{-6}\,\mathrm{k_d/a}$
Passers-by requirements for new risks		
at μ=0.1	$f_{\mathrm{s,d,new}}$	$6.3 \cdot 10^{-10}\,\mathrm{k_d/km}$
at μ=0.1	$f_{\mathrm{t,d,new}}$	$2.5 \cdot 10^{-6}\,\mathrm{k_d/a}$
at μ=1	$f_{\mathrm{s,d,new}}$	$6.3 \cdot 10^{-11}\,\mathrm{k_d/km}$
at μ=1	$f_{\mathrm{t,d,new}}$	$2.5 \cdot 10^{-7}\,\mathrm{k_d/a}$
Society requirements		
In 5 years at μ=0.095	$f_{\mathrm{s,d,soc}}$	$1.8 \cdot 10^{-9}\,\mathrm{k_d/km}$
In 5 years at μ=0.095	$f_{\mathrm{t,d,soc}}$	$7.2 \cdot 10^{-6}\,\mathrm{k_d/a}$
In 30 years at μ=1	$f_{\mathrm{s,d,soc}}$	$2.9 \cdot 10^{-10}\,\mathrm{k_d/km}$
In 30 years at μ=1	$f_{\mathrm{t,d,soc}}$	$1.2 \cdot 10^{-6}\,\mathrm{k_d/a}$

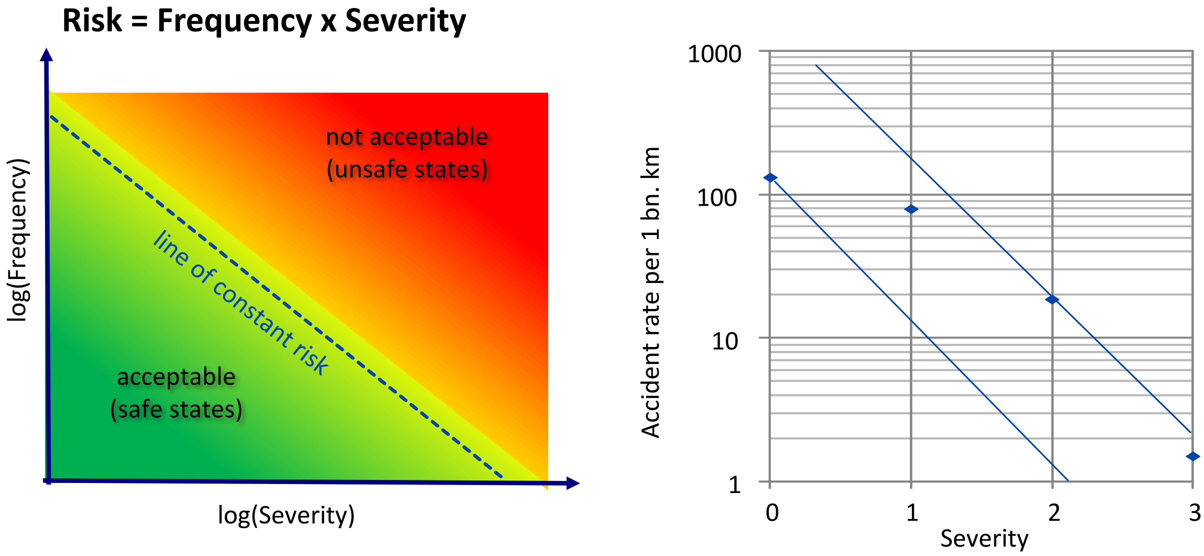


FIGURE 1

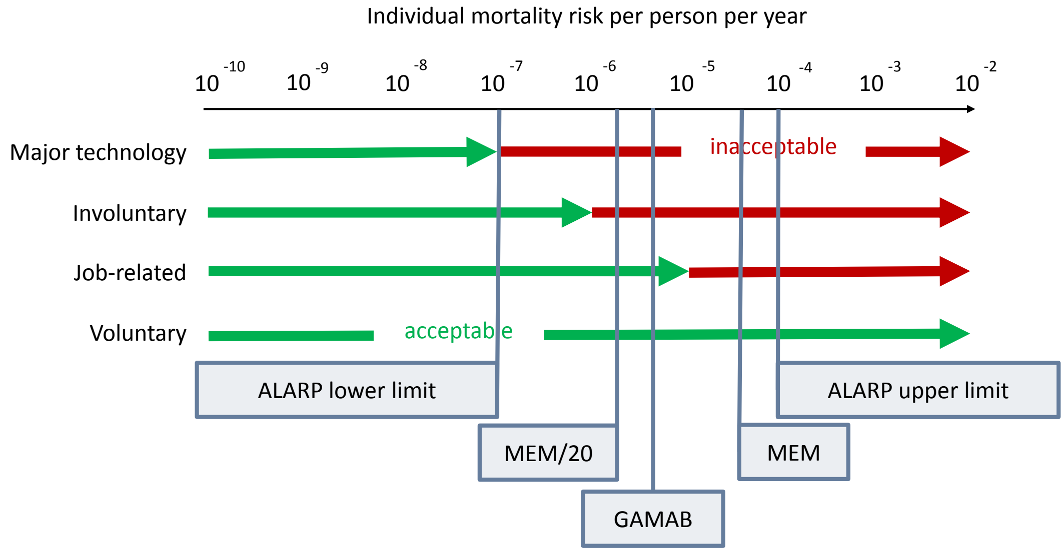
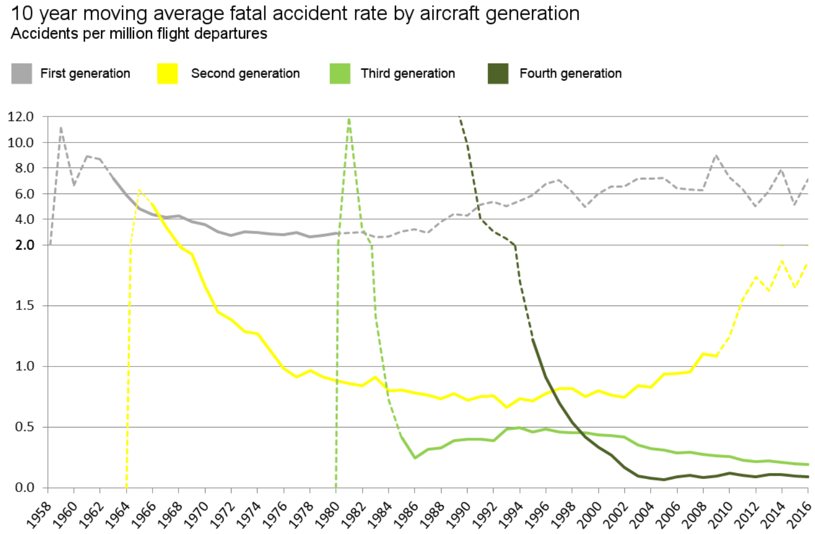
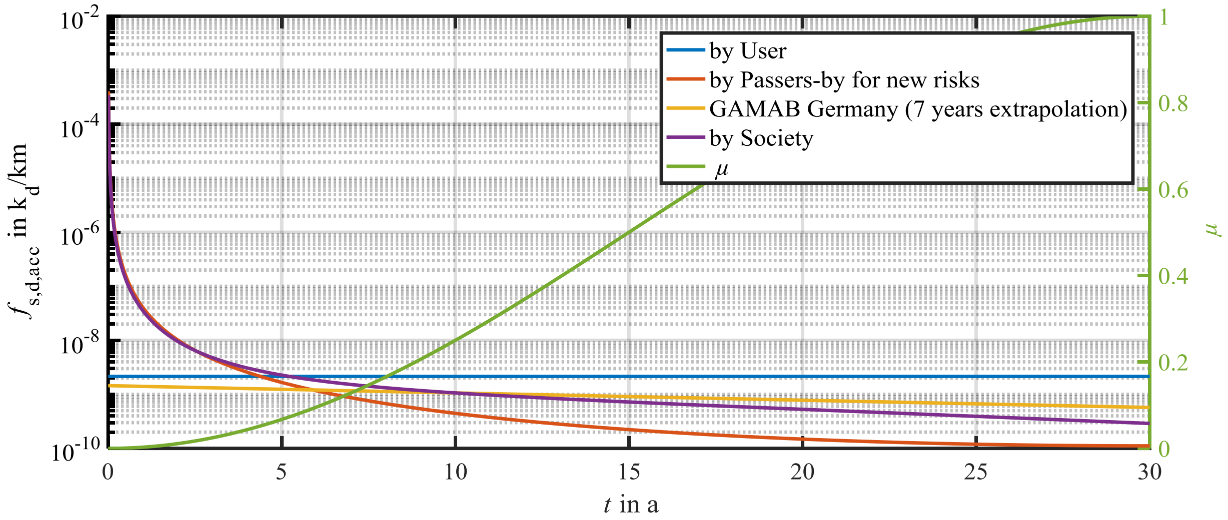


FIGURE 2



Source: Airbus "A statistical analysis of commercial aviation accidents 1958-2016"

FIGURE 3



10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
$f_{\mathrm{s,d,acc}}$ in $\mathrm{k_d}$/km
by User
by Passers-by for new risks
GAMAB Germany (7 years extrapolation)
by Society
μ
1
0.8
0.6
0.4
0.2
0
μ
0
5
10
15
20
25
30
t in a

FIGURE 4