Chapter 17
The New Role of Road Testing for the Safety Validation of Automated Vehicles

Walther Wachenfeld and Hermann Winner

17.1 Introduction

There is a difference between developing a vehicle that is driving itself as safely as today’s cars are driven and assessing this vehicle in terms of safety. But, in order to introduce highly automated vehicles to public traffic, both must be done and remaining open questions for both must be answered. When trying to introduce these vehicles not just for a few users with limited use cases but rather as a mass product for a whole society, the safety has to be shown for many kilometers over many years resulting in a vast number and combination of different situations.

For the introduction of highly automated vehicles that cover the SAE levels 4 and 5 (“High Automation” and “Full Automation”) \[1\] like an “Interstate Pilot Using Driver for Extended Availability,” a “Full Automation Using Driver for Extended Availability,” or a far advanced mobility concept like the “Vehicle on Demand” \[2\], special challenges arise for assessing the vehicle in terms of safety as the controller, being the human driver, isn’t available anymore to correct the automation’s behavior \[3\]. In each and every situation, the automation needs to choose an adequate behavior out of manifold possibilities.

In order to discuss the question: “How can road testing contribute to the safety assessment of highly automated vehicles?” in the first part of this chapter, we will specify the goals for assessment before explaining the challenges for road testing. Other test tools besides road testing are briefly introduced, and their demand for road tests is derived. The second part of this chapter will lay down a new train of thought on how to proceed with the “approval-trap” for highly automated vehicles. The
A statistical approach will be developed further to get a first idea how an introduction strategy for a highway pilot could look like.

### 17.2 The Goal of the Validation in Terms of Safety

Wachenfeld and Winner [3] describe the necessary relation $V_{\text{acc}}$ between additional risk $R_{\text{add}}$ and avoided risk $R_{\text{avo}}$ to human beings that emerges from the introduction of automated vehicles.

$$V_{\text{acc}} \rho = \frac{R_{\text{add}}}{R_{\text{avo}}} \quad (17.1)$$

The goal of the validation is to prove this relation. The exact level of this relation that needs to be met is unclear and depends on the benefit that comes with the introduction. As long as the validation can prove that the advantages offset the disadvantages for every stakeholder $\rho$, the vehicle can be released for production and can be introduced into traffic. The benefits caused by automated vehicles separate the people into two groups of stakeholders. On the one hand, the users or passengers directly benefit from the technology and can choose whether they want to use the technology or not. On the other hand, there are those affected by the technology outside the vehicle without being able to choose being affected or not. When assuming that the ones affected are exposed uniformly to the technology, we define these as the society. As both groups benefit and are affected in different ways, the validation has to take this into account. Dividing up the goal on users and society considers the freedom of individuals to also take a higher risk for themselves with the restriction not to endanger others above the currently accepted threshold. Examples are the usage of motorcycles and cars compared to trains for this freedom of choice.

At this point, we propose two different goals for the validation of automated vehicles in terms of safety:

1. A user or passenger who benefits from the use of the automated vehicle can be exposed to a higher risk than the one taken when driving by himself (reference risk) as long as he is informed about this additional risk. The user should get the option to weigh conservatively defined advantages and disadvantages.
2. As long as no major additional benefit for society is agreed on, we demand that the society is not exposed to a higher risk than to being part of today’s traffic.
17.3 Challenges for Safety Validation of Automated Vehicles Based on Road Testing

Road testing for safety validation of highly automated vehicles means bringing a nearly final version of the object under test (OUT) into real traffic. In focus are systems that could endanger humans if not functioning properly. It is a safety critical unsupervised/non-correctable system [4] that will be released after the validation step. As it is unknown if the required safety is reached, a trained test driver is necessary to supervise the OUT although it is not originally made for being supervised (like all demonstrators of self-driving vehicles presented up to now).

The idea of road testing is that the OUT is confronted with enough relevant situations in real traffic. If the OUT passes these situations without the intervention of test drivers, the assumption is made that it will be able to do this in everyday use as well. The question is: How many situations in real traffic are enough? To answer this question, the properties of road testing, meaning the way situations are formed in public traffic, need to be studied. The most important property of road testing is that the reality is not manipulated, as the validity of the test would suffer. For example, there should not be some kind of actor (traffic participant) that imitates behavior to test the OUT. As he or she is conscious about what happens, by definition the behavior and especially his or her reactions will be artificial. If no manipulation on real traffic occurs, parameters of different kind form situations. For example, the authors in [5] distinguish dynamic objects, scenery, and actions that are described by parameters like position in time, width of a lane, or a deceleration value.

These parameters can be classified by the possibility to consciously select or cover situations that are formed by these parameters:

- By choosing the route (spatial), static objects and correlated effects can be gathered consciously. For example, if a certain form of tunnel is relevant for testing, planning a route along this tunnel easily leads to the OUT confrontation with this situation.
- By choosing a certain time (temporal), the parameters like lighting, weather conditions, season of nature, etc. can be selected consciously.

To a certain extent, both categories are predictable and as a result easy to aim for. On the other hand, constellations and situations that can be influenced by a living being, such as the two that follow, are more difficult to aim for:

- The exposure to high traffic density can, for example, be increased by choosing the route and time in combination. This is possible as the large number of living beings evens out different behavior. Some of many human beings, e.g., need to commute back home at a certain time, thus making it predictable.
- What is difficult to experience is behavior that is not correlated to time and space, such as a human overlooking an obstacle or being distracted from another passenger or his mobile phone.
Different driving behavior and thus different trajectories are caused by different combinations of these parameters. As not all parameters can be chosen actively without manipulating the road test, trying to cover dangerous situations with low exposure is challenging. It is more like a random selection that a certain combination of parameters is experienced during a test drive. The higher the number of kilometers or time driven, the more probable it gets to come upon a certain combination of parameters. The question of how many kilometers need to be driven on the road to assure sufficient coverage of combinations now arises.

It is relevant to highlight that the number of parameter combinations between the different levels of automation (SAE 0–2 vs. SAE 3–5) differs significantly. Today’s vehicles (SAE 0–2) need to be assessed [3] for being a controllable tool supervised by a driving human. The possibility for the driver to brake, steer, and accelerate don’t depend on the surrounding vehicles, whereas for more highly automated vehicles (SAE 3–5), the whole vehicle behavior also depends on the other vehicles’ behavior as well as the environmental conditions. Accordingly, the challenge explained in the following is also valid for today’s vehicles but the human capabilities to perceive and interpret the surrounding vehicles simplify the assessment and reduce the relevant parameter combinations strongly. This human driver is removed for highly automated vehicles, and therefore this simplification is missing.

The challenge for assessment consequently follows from the high level of safety that is defined by the benchmark and by the random occurrence of situations for road testing. Transferring the high level of safety into the theory of parameters forming situations in real traffic means that parameter combinations that end in challenging situations for the OUT appear rarely.

One benchmark is accident statistics that provide numbers (see Table 17.1) on average distances between two accidents of different severity levels [7].

Accidents can consequently be seen as parameter combinations that are relevant for safety assessment. To answer the question of how many kilometers need to be driven, the probability distribution function of a random number is necessary in addition to the average distance between two events. When the number of road accidents for a given amount of kilometers driven can be seen as a Poisson distributed number [8, 9], a statistical argumentation can be laid out how many kilometers are necessary to scientifically prove safety. This leads to the so-called “approval-trap” [3]. The approval-trap explains that although a safe vehicle is developed, there is no economical way to prove its safety. For the example of the

| Severity level | Average distance between two accidents of this level (km) |
|----------------|----------------------------------------------------------|
| S3             | $660.0 \cdot 10^6$                                       |
| S2             | $53.2 \cdot 10^6$                                        |
| S1             | $12.5 \cdot 10^6$                                        |
| S0             | $7.5 \cdot 10^6$                                         |
Highway-Pilot, this means that although a vehicle is developed that is twice as good as today’s traffic, approximately 10 times the distance between two events is necessary to have a chance of 50% to prove (level of significance 5%) that the developed vehicle is safer than the comparison group. For severity S3 (fatality), this leads to 10 times $660 \cdot 10^6$ km $= 6.6 \cdot 10^9$ km, which is economically not feasible. The safety can’t be shown by just driving on German highways, and thus the lack of feasible safety assessment hinders the vehicle release. Up to this point, we haven’t even discussed changes that would require retesting affected kilometers of the road test.

A first conclusion can be drawn: Road testing before start of production (SOP) will not be suitable to statistically prove the same safety level of higher automated vehicles compared to today’s traffic. Thus, other ways out of the approval-trap need to be found. The next two sections will focus on alternative ways. Firstly, alternative approaches replacing the road test will be discussed before a new interpretation of the statistical approach is motivated.

### 17.4 Challenges for New Approaches on Safety Validation

Due to rising costs that are a consequence of increasing complexity of additional functionality that is implemented in new vehicles, the automotive industry is looking for a way to optimize verification and validation (V&V) processes. The question is whether these approaches can solve the approval-trap for highly automated vehicles, explained above.

#### 17.4.1 New Approaches on Safety Validation

From the road test perspective, three approaches to optimize V&V processes are described in literature:

First of all, one could improve road testing itself. Above, different parameters of road testing were explained. Some of these can be actively aimed for. The goal of this approach is to reduce irrelevant kilometers and to aim for potentially relevant kilometers/events in real traffic. In [10], this idea is laid out in more detail for advanced driver assistant systems (ADAS). Based on a road test, events are monitored and interpreted as relevant or irrelevant, and future test routes are planned based on this monitoring.

Besides improving road testing, approaches are described to substitute road testing by artificial or virtual tests [11]. For higher automated driving, Fig. 17.1 structures on a generic level the different tools used. A more detailed classification can be found in [12]. The distinction between real, artificial, and virtual is illustrated by the example of a human as part of the environment: The human can either behave
like he would in reality, or he could show artificial behavior like a person being aware that he is monitored. In both cases the human is a living being. On the other hand, the human can also be replaced by a technical system like an artificial dummy. The third approach would be a virtual representation of the human in software. In this case, the behavior of the human is modeled by software in a virtual world.

The different tools like test tracks, Vehicle-in-the-Loop, Hardware-in-the-Loop, and Software-in-the-Loop (SiL) replace some of the environmental or vehicle parts with artificial or virtual parts. Virtual and artificial parts can be manipulated and observed more easily with the motivation to set up relevant test cases and neglect irrelevant ones. Additionally, these tools try to reduce additional risk that is introduced when testing technical systems. A special motivation for SiL is the independence of time and hardware. Simulations can be parallelized and accelerated, limited only by computational power.

The third approach tries to combine road testing with other tools like SiL. The approach “Virtual Assessment of Automation in Field Operation” (VAAFO [13]) as well as other methods described in literature [14–16] follow a similar method: Human- or ADAS-controlled vehicles are equipped with hardware to perceive and interpret the situation as the OUT would. The virtual behavior of the OUT is assessed in all situations covered by the real human driver. As the OUT cannot act on the actuators, this method can be executed without introducing additional risks to public traffic. Not only test drivers, but in principle every driver with a driver license can execute this combined road and SiL testing.

All three approaches are based on simplifications and assumptions to either replace real parts of a test with artificial/virtual ones or neglect irrelevant situations/parameter combinations. These simplifications and assumptions can be invalid when applied on OUT assessment.

17.4.2 Validation of Alternative Approaches by Road Testing

To avoid using simplifications and assumptions that are not proper for OUT assessment, real driving such as road testing is necessary. This time road driving
is used for safety validation of test tools and for safety validation of assumptions. The validation of tools for a defined number of test cases seems possible. But again, who can tell whether the selection met the necessary situations? Therefore, we come back to the challenge raised by the statistical train of thoughts. How can we show that the tools and assumptions are valid for OUT safety assessment?

An advantage for tool and assumption validation is that the number of kilometers doesn’t need to be driven with the OUT. This simplifies the collection of kilometers. Another factor that would reduce the huge amount of possible situations that need to be covered for tool or assumption validation could be the independence of different parameters of a situation. For example, properties of traffic models are independent of properties of radar sensor models and therefore don’t need to be modeled and validated in combination. This independence doesn’t exist for the validation of the OUT as errors in real sensors lead to different behavior depending on the surrounding traffic. A disadvantage is that even more things need to be validated. For example, the behavior of other road participants needs to be reflected by the tool, at least to a certain extent.

Until now we have not seen any proof that the advantages outweigh the disadvantages resulting in less kilometers necessary to be driven (no matter who collects the stated amount). On the one hand, the more components are replaced and the more cases are neglected, the more validation effort for tools and assumptions has to be made. On the other hand, the more cases are left for road testing, the more validation of the OUT has to be done on the road. This seems to be a trade-off between OUT validation on the one hand and tool validation for OUT validation on the other hand. Additionally, the long-term perspective has to be considered as well. It can be that the first validation of tools needs a higher effort as the road testing itself, but when validating another version, vehicle type, or new generation, the overall effort could be reduced by orders of magnitude. An example for that effort reduction is described in [11] for ESC testing.

At this point, a second conclusion can be drawn: When pursuing approaches to replace or reduce road testing, road tests will still be of interest as these approaches need to be validated. At least until now, it’s unclear whether or not other approaches reduce the validation effort for the first vehicle.

Of course, if a tool or an assumption is validated, its advantages and potential to increase efficiency can be utilized. But up to that point, validation activities based on real driving are and will be necessary.

### 17.5 A Confession About the First Introduction of Automated Vehicles

The proof of safety of the OUT by simply road testing before SOP is economically infeasible with statistical significance. For alternative approaches, it is at least uncertain if the required validation effort is reduced. Tool and assumption validation could equal out the reduction of OUT validation.
This leads to the conclusion that, from a statistical perspective, the first vehicles that will be introduced will not satisfy a scientific proof of comparable safety.

This seems to be an obstacle on the way to everyday automated driving. But, if safety cannot be proved, the scientific question to ask is if the hypothesis “the automated vehicle achieves the requirements in terms of safety” can be falsified. If using a conservative approach to test for being worse without a result, one could take the chance and the risk to introduce the vehicle.

For this introduction, the goals defined in the beginning of this chapter must still be achieved.

- The risk for society shall not be increased.
- The individual as the passenger should be able to weigh imminent risks and, depending on that weighting, be able to choose whether or not to use the vehicle.

In the following, we will explain an argumentation that has the potential to enable introducing automated vehicles based on the same statistics that before have led to the approval-trap.

### 17.6 Argumentation for Introduction of Automated Systems Motivated by Statistics

The following argumentation is based on the assumption that accidents are Poisson distributed. This Poisson distribution is explained before the generic theory is laid out and, based on that, the introduction of automated driving on German autobahn is illustrated with examples.

To represent the distribution of accident events, we use the Poisson distribution [8, 9]:

\[
P_{\lambda}(k) = \frac{\lambda^k}{k!} e^{-\lambda}
\]  

(17.2)

This distribution assumes that the occurrence of an accident is an independent and non-exhaustive random process \(P_{\lambda}(k)\). In the equation, \(k\) corresponds to the number of accident events and \(\lambda\) to the expected value with which this event occurs. The expected value \(\lambda\) is defined by the quotient

\[
\lambda = \frac{d_{\text{test}}}{SP}
\]

(17.3)

Whereby \(d_{\text{test}}\) stands for the observed test kilometers and \(SP\) for the performance of the system. The performance represents the expected number of kilometers between the accidents. Two distributions of a probability distribution function with different expected values are depicted in Fig. 17.2. For this example, it is assumed that a certain number of kilometers \(d_{\text{test}}\) were driven and one event occurred. Can we now
define a worse and a better performance level $SP$ from this test? Equation (17.3) connects this with the search of an expected value $\lambda$. Based on a probability of error value\(^1\) $e = 1\%$, these questions can be mathematically formulated with two equations:

\[
P_{\lambda_-} (k > 0) \leq 1\% \tag{17.4}
\]

For which expected number $\lambda_-$ is the probability to have at least\(^2\) one accident less than or equal to 1%? A numerical search provides the value $\lambda_- = 0.01$. This tells us that when one accident occurs after $d_{\text{est}}$ kilometers, we statistically prove with $e = 1\%$ error probability that the vehicle is worse in terms of safety compared to a performance level of $SP = \frac{d_{\text{est}}}{\lambda_-} = \frac{d_{\text{est}}}{0.01}$. In other words, Eq. (17.4) tells that with a probability of 99%, no event occurs assuming $\lambda_- = 0.01$.

\[
P_{\lambda_+} (k \leq 1) \leq 1\% \tag{17.5}
\]

The second equation (17.5) asks for which $\lambda_+$ the probability that one or no accidents happened is at most 1%. In this case, the numerical search provides the

\(^1\)The value (5\%, 1\%, 0.1\% etc.) that is taken needs further considerations but is just one variable in that theory. For this chapter, we intentionally use 1\% and thus a different value as used in [3]. We don’t want to be misunderstood that one or the other is the “right” value for that.

\(^2\)Please be aware that $P (k > a) = \sum_{a+1}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$ is the cumulative distribution function. The same counts for $P (k < a) = \sum_{0}^{a-1} \frac{\lambda^k}{k!} e^{-\lambda}$. 

---

Fig. 17.2 Poisson probability distribution function (left) and cumulative distribution function (right) for two different expected values: one for the proof of being worse and one for the proof of being better when one event occurred.
value $\lambda_+ = 6.64$. This tells us that when one accident occurs after $d_{\text{test}}$ kilometers, we statistically prove with $e = 1\%$ that the vehicle is better in terms of safety compared to a performance level of $SP = \frac{d_{\text{test}}}{\lambda_+} = \frac{d_{\text{test}}}{6.64}$. Both probability functions (PDF and CDF) are plotted in Fig. 17.2. With a chance of $99\%$, the expected value of the OUT is

$$\lambda_- \leq \lambda_{\text{OUT}} \leq \lambda_+. \quad (17.6)$$

As the smaller the expected value, the better the OUT, the value for $\lambda_-$ defines a kind of best case and the value for $\lambda_+$ defines similar a worst case in that sense.

Based on that theory of accidents being Poisson distributed, we explain in [3] why using road tests for the proof being better than today’s human drivers are economically not feasible.

In the following, we will use the test to the “other hand,” meaning how and when one can be sure and state that the vehicle that is introduced into traffic contravenes the goals stated above.

### 17.6.1 Universal Theory on a “Brave Introduction” of Automated Systems

Based on the explanation of the Poisson distribution of accidents, we will now lay out a universal theory on the introduction of automated systems where certain events (accidents) exist which have to be considered that happen after a certain time or distance of travel. Additionally, it is assumed that a comparison level (benchmark) is known.

Let’s assume that the performance level for the benchmark is $SP = SP_{\text{bench}}$. So, in statistical average, after $SP_{\text{bench}}$ kilometers, one of the relevant events should happen. Now we want to introduce the OUT after it was tested for

$$d_{\text{test}} = \frac{SP_{\text{bench}}}{\tau} \quad (17.7)$$

kilometers and one event occurred. $\tau$ describes the ratio between the benchmark and the test kilometers. We know from Eq. (17.4) that the performance level of the OUT is equal or worse:

$$SP_{\text{OUT}} \leq \frac{d_{\text{test}}}{\lambda_-}. \quad (17.8)$$

Combining both Eqs. (17.7) and (17.8), the performance level of the OUT is

$$SP_{\text{OUT}} \leq \frac{SP_{\text{bench}}}{\tau \cdot \lambda_-}. \quad (17.9)$$
Equation (17.9) tells us that the OUT is worse than \( \frac{1}{\tau \lambda_-} \) times the benchmark. On the other hand, we can only tell that the vehicle is better than

\[
SP_{\text{OUT}} \geq \frac{SP_{\text{bench}}}{\tau \cdot \lambda_+},
\]

(17.10)

With this test, we can’t prove that we are less safe than \( \frac{1}{\tau \lambda_-} \) times the benchmark, and on the other hand we only prove that we are safer than a \( \tau \cdot \lambda_+ \) times worse system.

Why should this system not be introduced into traffic when we can’t prove it is less safe than \( \frac{1}{\tau \lambda_-} \) the benchmark? Why should we take the risk to introduce the system when we just know that it is safer than a \( \tau \cdot \lambda_+ \) times worse system? Let’s compare the result of the test with the goals stated above separately for the user and the society.

### 17.6.1.1 User’s Perspective

From the user’s perspective, the system has a certain benefit. This benefit comes not from safety but from mobility, freedom, money, etc. The user knows that he or she probably needs to take a higher risk to access the benefits, but as long as the counted events don’t lead to a \( \lambda_- \) below the line of Fig. 17.3, less safety can’t be proven. This actually leads to an introduction that matches the goal for the user.

For the operating company of the system, there should be an option to stop and block the automated mode when from a monitoring of relevant events a non-acceptable safety can be proven. This limit is not necessary the whole area below the line as the users’ acceptance defines this limit. How this monitoring should look and how the option to stop should be implemented is intentionally left open.

**Fig. 17.3** Area of proven less safety than benchmark system for \( e = 1\% \)
17.6.1.2 Society’s Perspective

On the other hand, for the society that in the first place has no direct benefit from the individual using the automated system, it is justified to ask if the society is exposed to a higher risk. The question is how to prove that society is exposed to that higher risk. One way to study the risk this society faces when using this technology would be to study the number of events that happened before the technology was introduced. This number of events, recorded, for example, each year, should not be affected negatively by this introduction. Figure 17.4 displays such an example. Each year, a discrete number of events were recorded. Over the year, the number decreased following a certain monotonic trend. This trend is not given but can be fitted by, for example, a least square approximation of a suitable mathematical function like a quadratic function\(^3\) \(f_{\text{fit}}\). This is done for the example by the (red/solid) line (see Fig. 17.4). All points differ from the trend line. These deviations are still independent from the technology we want to introduce. This fact leads to the question: How does the next number of recorded events have to differ from the values in the past, fitted by the trend line, to be sure that it was affected negatively?

\(^3\) \(k = 0.18 \cdot i^2 - 41.16 \cdot i + 1202.\)
To answer this, we derive the standard deviation of these events compared to the trend line.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (k_i - f_{\text{fit}}(i))^2}$$

(17.11)

In this equation, $N$ is the number of years and $k_i$ is the number of events recorded in year $i$. We now define that

$$k_{N+1} \leq f_{\text{fit}}(N + 1) + \frac{\sigma}{\beta}$$

(17.12)

is indistinguishable for society, where $\frac{\sigma}{\beta}$ expresses the number of events caused by the introduction of the new technology. We assume this being indistinguishable for society as it is $\beta$ times smaller than the standard deviation and therefore disappears in the noise of numbers each year. There is actually no way to prove or detect that the trend is affected negatively, as the number is too small and lies below the limit of detection.\(^4\)

To match the goal for society, we now want to prove that the risk will not increase above the stated limit of $\frac{\sigma}{\beta}$. This can be done by limiting the numbers of kilometers driven on public roads. The slower the technology is introduced, the fewer events will happen in a single year. The number of systems or correspondingly the time of usage or number of kilometers will be limited by this approach. But how many systems $I_{\text{OUT}}$ can be introduced?

If we assume that one system covers an average of $\bar{d}$ each year by introducing $I_{\text{OUT}}$ systems, we get $d_{\Sigma} = I_{\text{OUT}} \cdot \bar{d}$ driven during 1 year. With the requirement

$$P_{\lambda_{\Sigma}} \left( k \leq \frac{\sigma}{\beta} \right) \geq 99 \% = 1 - e$$

(17.13)

where $\lambda_{\Sigma} = \frac{d_{\Sigma}}{SP_{\text{OUT}}}$. Equation (17.13) requires that the number of events caused by the OUT should be smaller than $\frac{1}{\beta}$ of the standard deviation with a probability of error of 1 %. The value for $\lambda_{\Sigma}$ can again be derived numerically. The performance level of the OUT can be estimated from the test drive with Eq. (17.10) which, together with the result from Eq. (17.13), results in

$$d_{\Sigma} \leq \lambda_{\Sigma} \cdot SP_{\text{OUT}} = \lambda_{\Sigma} \cdot \frac{SP_{\text{bench}}}{\tau \cdot \lambda_{\Sigma}}.$$  

(17.14)

Equation (17.14) shows that if less than $d_{\Sigma}$ kilometers with the new technology are driven during 1 year, the danger to contravene the requirements from society is smaller than 1 %.

\(^4\)There exists more theory to define this limit but at this point we use a variable to keep it simple.
17.6.2 EXAMPLE: Introducing Highly Automated Driving on German Autobahn

Let’s try to apply this introduction approach on automated driving. Events in this case can be accidents with different levels of severity. The most challenging and best recorded number of accidents are the ones that involve fatalities. Focusing on Germany, we can discuss this example for the highway pilot, as numbers are given for this use case (see Table 17.1).

\[ SP_{\text{bench}} = 660 \cdot 10^6 \text{ km} \] (17.15)

Equation (17.15) expresses that on average on German autobahn, a fatality occurred every 660 million km. This defines the benchmark for the introduction. When assuming that the OUT is tested for \( \tau \) times less kilometers and maximum one accident with fatality was generated, the test delivers two results. First, the vehicle in best case has a certain maximum level of performance in terms of safety [conclusion drawn from Eq. (17.9)]. Second, Eq. (17.10) shows the performance level is reliably higher than

\[ SP_{\text{OUT}} \geq \frac{SP_{\text{bench}}}{\tau \cdot \lambda_+} = \frac{660 \cdot 10^6 \text{ km}}{\tau \cdot \lambda_+}. \] (17.16)

At this point, a conservative approach is chosen as the worst case estimation is selected. Now we know the users’ perspective for the introduction of the highway pilot.

For the society’s perspective, we refer to [17] where the accident statistic (Destatis) for the years 1992–2014 is reported. These numbers and the fitted curve are depicted in Fig. 17.4. From Eq. (17.11)

\[ \sigma \approx 48 \] (17.17)

can be calculated. Combining the performance level from Eq. (17.16) with the standard deviation from Eq. (17.17) and defining a value for \( \beta \), the numerical search delivers a certain \( \lambda_\Sigma \) (see Table 17.2). Together with Eq. (17.14), one gets

\[ d_\Sigma \leq \lambda_\Sigma \cdot \frac{SP_{\text{bench}}}{\tau \cdot \lambda_+}. \] (17.18)

All vehicles together shouldn’t activate the highway pilot for more than this amount of kilometers \( d_\Sigma \). Based on the numbers of 2013 where 52.4 million vehicles were registered and 224.2 billion km where driven on autobahn

\[ \bar{d} = \frac{224.2 \cdot 10^9 \text{ km}}{52.4 \cdot 10^6 \text{ km}} \approx 4278 \text{ km}. \] (17.19)
This number can be converted into a number of vehicles:

\[ I_{OUT} = \frac{d_S}{d} \]

(17.20)

assuming the usage doesn’t change.

With these numbers for the German autobahn, we get different cases for introduction depending on the variables we have defined in the theoretical part (see Table 17.2).

### 17.7 Conclusion

Besides the efforts spend on test tracks, today, road testing is the most important tool to validate the safety defined for a new vehicle. A huge effort is made to manage new functionalities resulting in higher complexity of new vehicles. Since the statistical proof of safety for highly automated systems (SAE-level 4 and 5) by classical road testing is economically infeasible, other approaches are requested. Three ways can be distinguished: improving road test efficiency; applying simplifications as well as virtualizations to reduce the effort with tools like test tracks, test benches, or XiL; and combining both with tools like the VAAFO concept. Unfortunately, up to now the evidence that these approaches can reduce the effort for the validation of highly automated systems is missing due to the reason that the approaches also need to be validated. Today, this seems to be a trap for the approval of the vehicles.

But, when looking back to the safety requirements, these can be separated for a user that benefits and a society that is ideally not burdened with additional risks. With these two goals, the statistical approach can be used for another second way of argumentation: (1) Whether we only introduce a new technology when we have unambiguously proven that the new technology does not introduce additional risks independently of every additional advantages or (2) we focus on the additional advantages and introduce the technology at a conservative estimate to profit from the advantages until one can reliable prove we are really introducing inappropriate
risks. Both ways of argumentation are valid interpretations from today’s perspective, but their consequences differ significantly. Argumentation two could pave the way for highly automated vehicles should it withstand the open debate of relevant stakeholders.5

We have seen that road testing was and will be relevant for the introduction of new vehicles. However, road testing will be used in a different way. One way will be the validation of alternative approaches, their tools, and simplifications. This will also lead to a big effort before one can profit the new approach. A second way will be the described partial validation for a cautious introduction strategy. Whether or not these kilometers need to be run on a real road or with other tools remains open.

Acknowledgement The research work has been funded by the Daimler and Benz Foundation.

References

1. SAE International Standard J3016: Taxonomy and Definitions for Terms related to On-Road Motor Vehicle Automated Driving Systems, SAE International (2014)
2. W. Wachenfeld et al., Use Cases for Autonomous Driving, in Autonomous Driving: Technical, Legal and Social Aspects, ed. by M. Maurer, J.C. Gerdes, B. Lenz, H. Winner (Springer, Berlin, 2016)
3. W. Wachenfeld, H. Winner, The Release of Autonomous Vehicles, in Autonomous Driving: Technical, Legal and Social Aspects, ed. by M. Maurer, J.C. Gerdes, B. Lenz, H. Winner (Springer, Berlin, 2016), pp. 425–449
4. W. Wachenfeld, H. Winner, Do Autonomous Vehicles Learn? in Autonomous Driving: Technical, Legal and Social Aspects, ed. by M. Maurer, J.C. Gerdes, B. Lenz, H. Winner (Springer, Berlin, 2016), pp. 451–471
5. S. Geyer, M. Baltzer, B. Franz, S. Hakuli, M. Kauer, M. Kienle, S. Meier, T. Weißgerber, K. Bengler, R. Bruder, Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance. IET Intell. Transp. Syst. 8(3), 183–189 (2013)
6. German Federal Statistics Agency (DESTATIS): Verkehrsunfälle 2013
7. H.-P. Schöner, Challenges and Approaches for Testing of Highly Automated Vehicles, in CESA 3.0—Congress on Automotive Electronic Systems, Paris, 4 Dec 2014
8. A. Nicholson, Y.-D. Wong, Are accidents poisson distributed? A statistical test. Accid. Anal. Prev. 25(1), 91–97 (1993)
9. M. Gründl, Fehler und Fehlverhalten als Ursache von Verkehrsunfällen und Konsequenzen für das Unfallvermeidungspotenzial und die Gestaltung von Fahrerassistenzsystemen (2005)
10. P. Glauner, A. Blumenstock, M. Haueis, Effiziente Felderprobung von Fahrerassistenzsystemen, in 8. Workshop Fahrerassistenzsysteme, ed. by UNI DAS e.V., pp. 5–14, Walting (2012)
11. U. Baake, K. Wüst, M. Maurer, A. Lutz, Testing and simulation-based validation of ESP systems for vans. ATZ Worldw. 116(2), 30–35 (2014). doi:10.1007/s38311-014-0021-6

5This is what we would like to encourage at that point. Please give positive and negative feedback on this introduction strategy to overcome the hindrances for introduction of higher automated systems.
12. F. Schuldt, T. Menzel, M. Maurer, Eine Methode für die Zuordnung von Testfällen für automatisierte Fahrzeugfunktionen auf X-in-the-Loop Simulationen im modularen virtuellen Testbaukasten, in 10. Workshop Fahrerassistenzsysteme, ed. by UNI DAS e.V., Walting (2015)

13. W. Wachenfeld, H. Winner, Virtual Assessment of Automation in Field Operation. A New Runtime Validation Method, in 10. Workshop Fahrerassistenzsysteme, ed. by UNI DAS e.V., Walting (2015)

14. H. Winner, Einrichtung zum Bereitstellen von Signalen in einem Kraftfahrzeug, Patent DE 101(02), 771 (2001)

15. B. Hoye, D. Lambert, G. Sutton, Autonomous Driving Comparison and Evaluation, US Patent 20,150,175,168 (2015)

16. http://www.heise.de/newsticker/meldung/Autonome-Autos-Danke-dass-Sie-das-Auto-von-morgen-testen-2760591.html. Accessed 23 Jul 2015

17. http://www.dvr.de/betriebe_bg/daten/unfallstatistik/de_autobahn.htm. Accessed 21 Aug 2015