Planning and evaluation of reliability demonstration testing with uncertainties

Tobias Leopold
Esslingen University of Applied Sciences, Germany, Tobias.Leopold@hs-esslingen.de

Abstract: The testing of products with the aim of a reliability proof can be realized by reliability demonstration tests (RDT). However, the supposedly good plannability as well as the simple evaluability based on the binomial distribution is in most cases linked to the assumption of distribution parameters of the failure behaviour. These assumptions are subject to uncertainties about the required distribution parameters, the effects of which are highly significant. Case studies illustrate the possible consequences for the planning of the needed sample size and the influence on the confidence level. Potential improvement approaches to achieve the most robust planning and evaluation based on reliability demonstration tests are explained in the context of endurance testing without any failures.

1. INTRODUCTION

The reliability of a product is a key customer requirement and represents one of the most important purchasing criteria [1]. To ensure this important product characteristic, product reliability must be brought into focus during product development.
First, product reliability must be defined for the overall product and, if necessary, broken down to lower system levels. A quantitative target definition consists of a minimum product reliability or maximum failure probability at a certain point in time [2]. For the proof of product reliability, the confidence level has to be defined, too. With these three quantitative parameters, a target definition is complete and suitable as a basis for demonstrating product reliability.

In practice, there are often numerous challenges in planning tests to proof product reliability. First of all, the type of verification procedure must be defined, which can basically be divided into failure-oriented or failure-free testing. Further planning parameters are, for example, the number of test items and the test duration. Taking into account project boundary conditions and availability of specimen, test planning is often an interactive process resulting in different planning scenarios.

In addition to cost and time requirements, some information on the failure behaviour of the test specimens is already necessary in the planning phase, which is usually only available imprecisely or comes from estimates of similar products [6]. These inaccuracies cause risks in the plannability and trustworthiness of the proof of reliability. The consideration of uncertainties is thus of enormous importance in the practical reliability work of a product development.

2. THEORY OF RELIABILITY DEMONSTRATION TEST

In the following, a failure-free testing is assumed, whereby the principle consideration of uncertainties is transferable to a failure-oriented testing. The basic principle of RDT is based on successful testing when all test items survive the test duration without failure. With this procedure, the proof of a defined minimum reliability can be provided at a target time. In addition, the trustworthiness is quantified by the confidence level [6], [10].

The statistical basis is the binomial distribution, the so-called procedure based on number of failures is therefore also named binomial testing [7]. The binomial equation leads to [9]:

$$1 - CL = \sum_{i=0}^{r} \binom{n}{i} \left(1 - R_{min}(t_{target})\right)^{i} R_{min}(t_{target})^{n-i} \tag{1}$$

Where n is the number of test units, r is the number of failures, R_{min} is the minimum reliability to be demonstrated at the service life requirement t_{target} and CL is the given confidence level. An assumption of a specific failure time distribution is not required in equation (1).

In practice, RDTs aim at testing without failures (r = 0), which finally leads to the equation of success run [9]:

$$1 - CL = R_{min}(t_{target})^{n} \qquad \text{respectively} \qquad R_{min}\left(t_{target}\right) = (1 - CL)^{1/n} \tag{2}$$

In this case, the highest confidence level is achieved or the smallest number of test items is required. In practice, planning for this case is done, often without estimating the probability of occurrence [4], [5] for this scenario.

Products with a long service life usually pose a challenge in testing. For example, passenger cars or commercial vehicles are expected to have a service life of 10 to 15 years. However, this long service life requirement cannot be tested 1:1 within the scope of product development. In addition to the use of time-graded load spectra, the test duration in such use cases must be shorter than the service life requirement t_{target}. For the numerical analysis of RDTs, a lifetime ratio LR is now introduced from the ratio of the test duration t_{test} to the required service life t_{target} [2]:

$$LR = \frac{t_{test}}{t_{target}} \tag{3}$$

Taking equation (3) into account in equation (2), the verifiable minimum reliability R_{min} under the assumption of a Weibull-distributed failure characteristic is thus as follows [2]:

$$R_{min}\left(t_{target}\right) = (1 - CL)^{\frac{1}{LR^{\beta} n}} \tag{4}$$

Therefore, as soon as the test duration t_{test} is not equal to the required service life t_{target} and therefore a lifetime ratio $LR \neq 1$ is applied, a statement about the failure behaviour is necessary. For the assumption of a Weibull distributed failure characteristic the estimation of shape parameter β is required, which, however, can only be determined from failure data.

Often, the shape parameter β is estimated from failure data of similar products in operating conditions that are as comparable as possible. Another option within a product development project is the possibility to derive the shape parameter β from previous test phases of the product development project. The possible difference of the test specimens from different sample phases from design changes or manufacturing influences must be evaluated with regard to a changed failure characteristic.

The estimation of the shape parameter β for an RDT is therefore always associated with uncertainties, the resulting effects are examined below.

3. CONSIDERATION OF UNCERTAINTIES IN RDT

3.1. Confidence level and sample size

The possible impact of uncertainties in the estimation of a shape parameter β on the confidence level and required number of test specimens is to be analyzed already in the planning phase of an RDT. The possible consequences are investigated on the basis of case studies and their effects are discussed. For the exemplary investigation, a vehicle component is used whose reliability target is defined with a B_{10}-value of 100.000 load cycles, whereby comparable requirements also exist in other industrial sectors. The planning parameters for an RDT are therefore as follows:

$$R_{min}(t_{target}) = 90\% \; @ \; CL = 90\%, t_{target} = 100.000LC \quad (5)$$

The necessary test duration can often not be fully represented within the project schedule in the context of vehicle development. The test duration t_{test} is therefore smaller than the required service life t_{target}, that leads with equation (3) to a lifetime ratio $LR < 1$. For this assumed case, the life time ratio LR has to be considered as described in equation (4).

The maximum test duration of scenario I is limited to 75.000 LC, in another scenario II the maximum test duration is 50.000 LC.

First, the implications of these two scenarios for planning the necessary sample size of an RDT are shown. For this purpose, the required number of test specimens is calculated from equation (6):

$$n = \frac{\ln(1-CL)}{LR^{\beta} \cdot \ln R_{min}(t_{target})} \quad (6)$$

Figure 1 shows the evaluation of the required sample size for scenario I as a function of different shape parameters β, in order to derive the effects of uncertainties of the shape parameter β on the number of test specimens n. As already noted, a value for the shape parameter β of the Weibull distribution is already necessary in the planning phase, but this can only be determined from failure data. If the estimation of the shape parameter is faulty, it follows directly that the number of test items is faulty. If, for example, a shape parameter of 2.0 is assumed, but in reality the failure cause is described by a shape parameter of 3.5, test planning would provide a sample size of 39 instead of 60 and therefore too few test specimens. 21 specimen are missing.

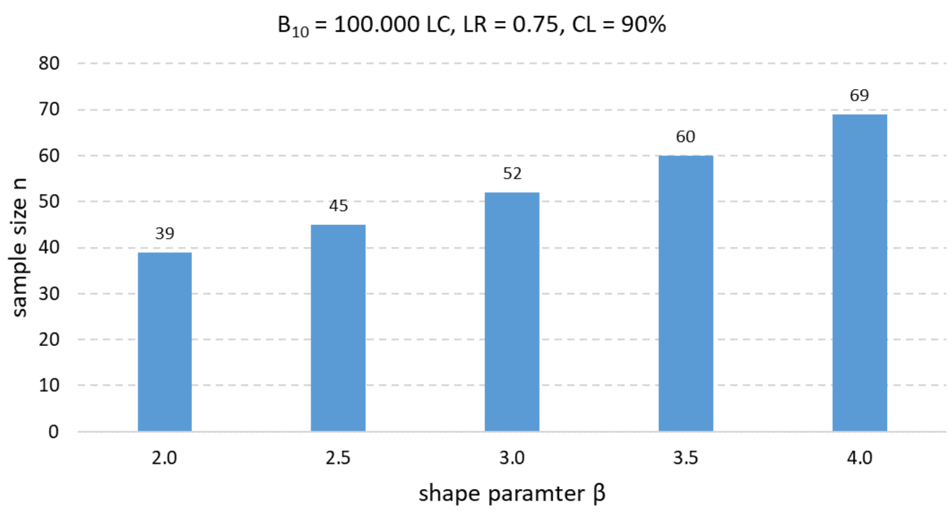

Figure 1: Required sample size n for scenario I of a RDT

The impact of an insufficient sample size, based on an uncertain or erroneous estimate of the shape parameter β, on the validity of the RDT and thus on the achieved confidence level is shown in figure 2. Here, the achieved confidence level is determined for scenario I, which results from the planning in figure 1 with a sample size of n = 39. The possible effects of an assumed shape parameter of β =2.0 on the confidence level is thus shown, when in reality the shape parameter is β > 2.0.

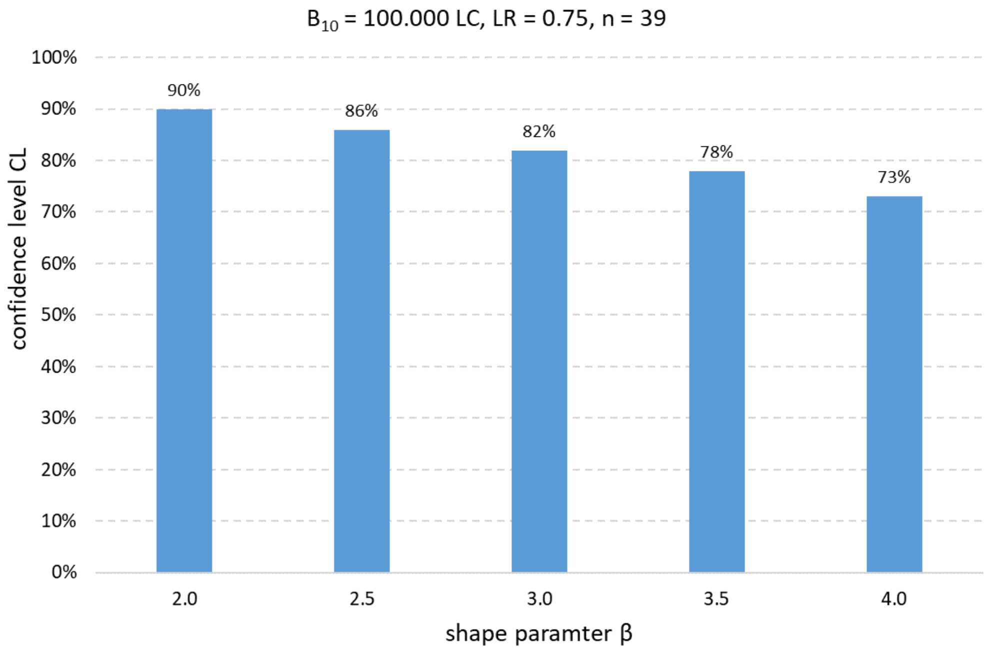

Figure 2: Achieved confidence level CL for scenario I of a RDT with uncertain shape parameter β

The too low number of test specimens n thus leads to a reduction of the achieved confidence level when an uncertain shape parameter β is used. However, since the estimation of the shape parameter of the Weibull distribution cannot be checked with an RDT, the reduced confidence level due to an uncertain shape parameter β initially remains undetected in practice.

In scenario II, the effects of an uncertain shape parameter β at a lifetime ratio of LR =0.5 are illustrated. The required sample sizes are shown in figure 3.

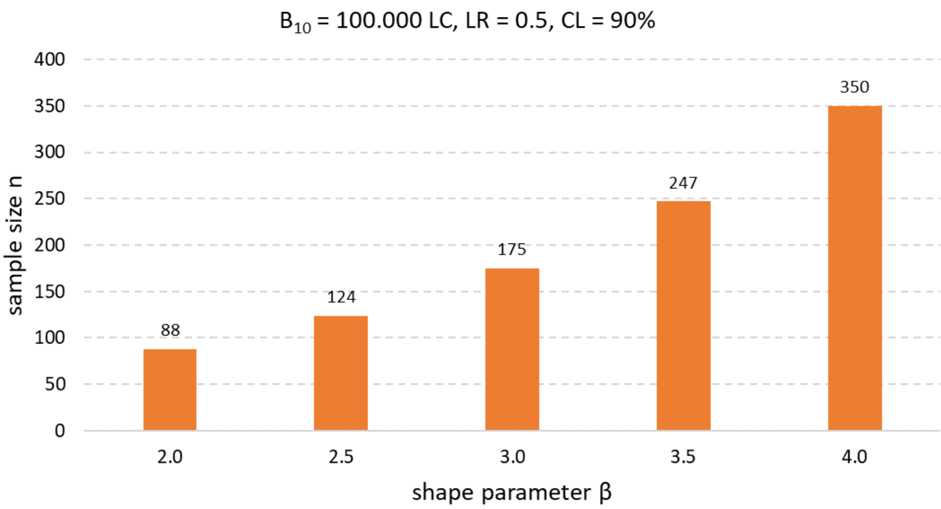

Figure 3: Required sample size n for scenario II of a RDT

In scenario II, an uncertain shape parameter β has a stronger impact on the confidence level than in scenario I, see figure 4. If the shape parameter is erroneously assessed as 2.0 instead of 4.0, the confidence level is halved from the required 90% to an alarming value of 44%

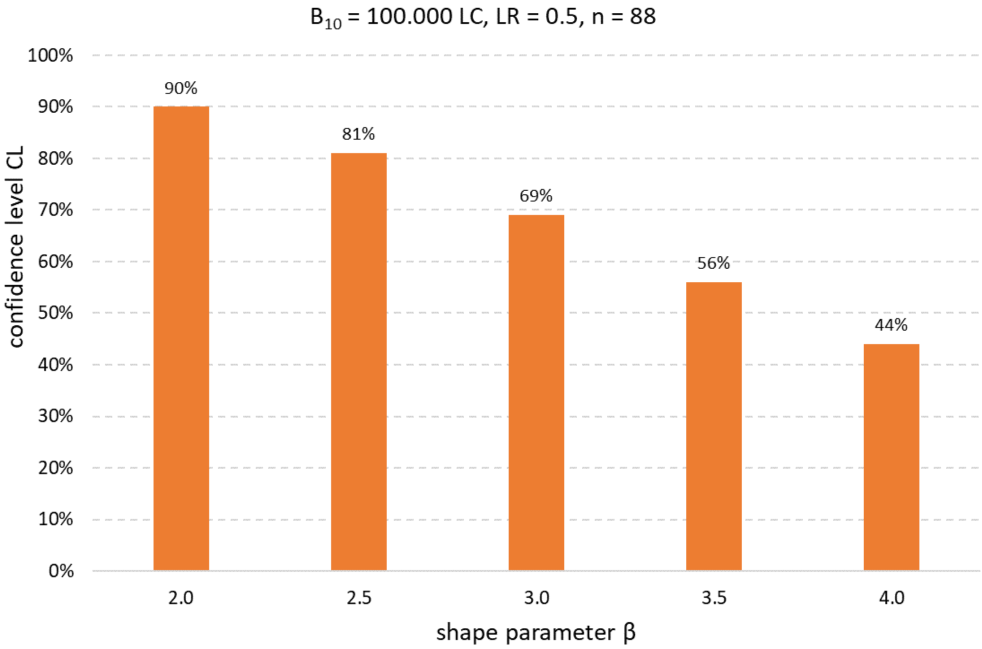

Figure 4: Achieved confidence level CL for scenario II of a RDT with uncertain shape parameter β

Scenarios I and II show the consequences of an uncertain or erroneous estimation of the shape parameter β of a failure-free test planning. In both scenarios, the shape parameter is assessed to be too small. Of course, it is also possible to estimate a shape parameter that is too large. From a statistical point of view, there is no increased reliability risk, quite the contrary: the sample size is too large. However, this means an economic risk, as unnecessary costs are incurred. The effects of an overestimated shape parameter can also be seen in figure 1 and figure 3. If a shape parameter of 4.0 is estimated for Scenario II, but in reality the shape parameter is only 2.0, a sample size of 88 would be sufficient instead of a sample size of 350. That means that almost four times as many specimens are tested than would be necessary.

The scenario study is intended to illustrate that it is of enormous importance to ensure an accurate estimate of a shape parameter at the beginning of a failure-free test planning. In addition, this shape parameter should be questioned, if possible, by performing failure-oriented test. For example, a RDT can be used to release a sample phase of a product development project. After this release, the failure-free test can be continued until a sufficient number of failures have occurred. This censored sample then allows the confirmation of the estimated shape parameter of the test setup for the RDT by a Weibull analysis of the censored sample.

3.2. Probability for a successful life test

In addition to estimating the required sample size and the resulting confidence levels, the probability of all test items surviving the required test duration without failure must also be determined. According to [4], [5] this probability for a successful life test can be determined as follows:

$$P_{zf} = \frac{number\ of\ test\ samples\ with\ zero\ failures}{total\ number\ of\ samples} \tag{6}$$

To estimate the probability P_{zf} for a successful success run or a failure-free trial of a sample, the failure behaviour of the specimen has to be characterized. For the case of a Weibull distributed failure behaviour, in addition to the shape parameter β, the second parameter of a 2-parameter Weibull distribution is now also necessary, the scale parameter η. Based on the binomial distribution, the probability can now be calculated that, for a defined number of specimens with a known failure probability respectively reliability, no specimen will fail until a certain point in time. This probability based on the binomial distribution thus corresponds exactly to the sought probability for a successful life test in which no failures occur [6].

$$P_{zf} = R(t_{test})^{n} \tag{7}$$

The reliability at time t_{test} of a sample with Weibull distributed failure behaviour can be described with:

$$R(t_{test}) = e^{-\left(\frac{t_{test}}{\eta}\right)^{\beta}} \tag{8}$$

The supposedly simple plannability of a reliability demonstration test is thus dependent on the estimation or prior knowledge of the failure behavior of the product. The impact of uncertain estimates on the probability of success P_{zf} of a success run test is analyzed below.

Figure 5 shows the success probabilities for a RDT of scenario I for a planning based on correct parameters of the Weibull distribution.

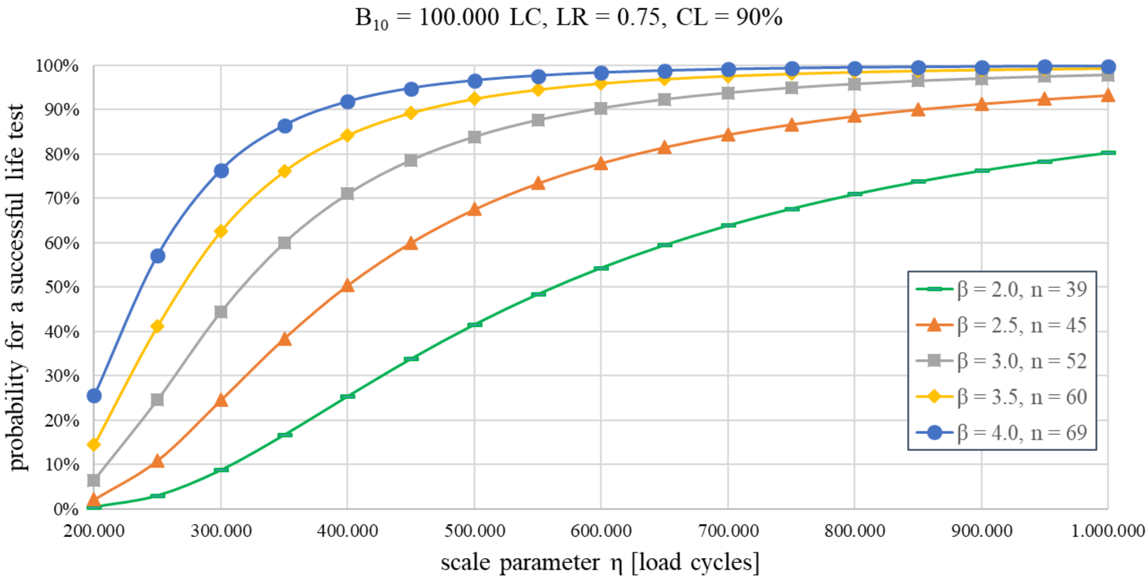

Figure 5: Planning of an RDT of scenario I on the basis of known, secure planning parameters

However, since Weibull parameters cannot be determined from the results of an RDT, which is necessary for the planning and evaluation of a RDT, both planning and evaluation must be performed with uncertain Weibull parameters.

In addition to determining the probability of success when the Weibull parameters are known, figure 5 also allows the probability of success to be estimated when the distribution parameters are imprecise. For example, if a shape parameter is estimated as 3.0 and a scale parameter is estimated as 800.000 LC, this results in a probability of success of 95.8%. However, if in reality the shape parameter is only 2.5 and the scale parameter is 400.000 LC, a sample size of 45 specimens (see figure 1) would result in a success probability of 50.4%.
In addition, the influence of the sample size must now be taken into account, which is shown for the described example in figure 6.

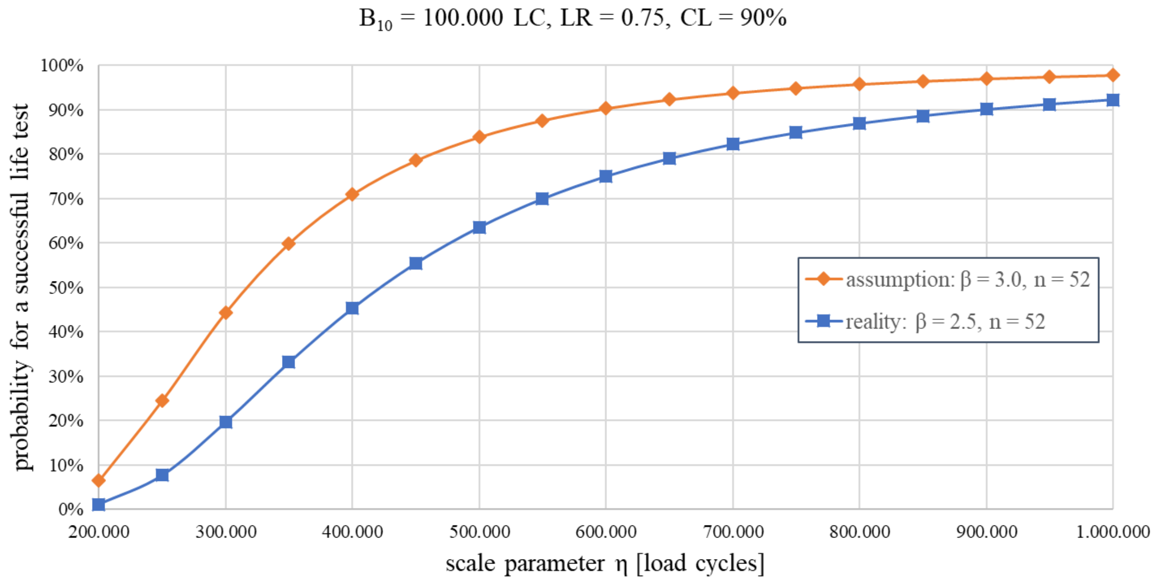

Figure 6: Planning of a RDT on the basis of uncertain planning parameters for scenario I

Planning based on the shape parameter with $\beta = 3.0$ would result in a sample size of $n = 52$. With these 52 test specimens, the probability of success is now to be checked, whereby the real distribution parameters with shape parameter $\beta = 2.5$ and scale parameter $\eta = 400.000$ LC must also be taken into account. The probability of success based on the assumption would be 95.8%, but in reality, a reduced probability of success of 45.3% for passing the test would have to be expected. Consequently, a supposedly safe success run test would in reality only be successful in every second case, see table 1.

table 1: Exemplary comparison of the effects of uncertain distribution parameters on success probability for scenario I

	assumption	reality
shape parameter β	3.0	2.5
scale parameter η	800.000 LC	400.000 LC
sample size n	52	52
probability of success P_{zf}	95.8%	45.3%

The exemplary effects of the uncertain Weibull parameters on the probability of success of a RDT for scenario II are shown in figure 7.

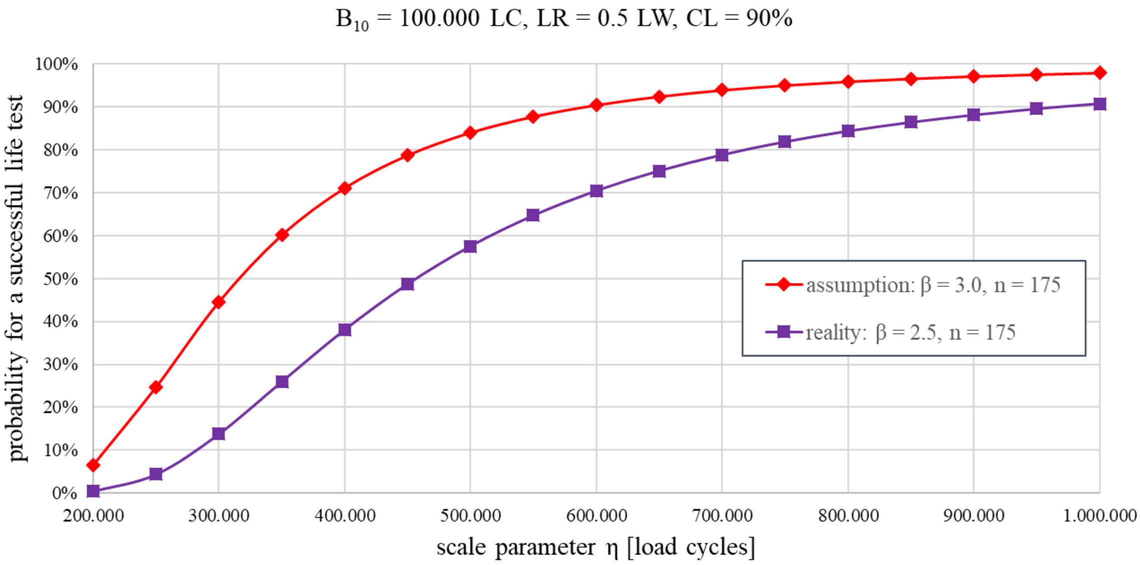

Figure 6: Planning of a RDT on the basis of uncertain planning parameters for scenario II

The deviations between assumption and reality are even more pronounced in scenario II compared to scenario I, see table 2.

table 2: Exemplary comparison of the effects of uncertain distribution parameters on success probability for scenario II

	assumption	reality
shape parameter β	3.0	2.5
scale parameter η	800.000 LC	400.000 LC
sample size n	175	175
probability of success P_{zf}	95.8%	38.0%

3.3. Approaches for improvement

The supposedly simple plannability of a RDT is directly dependent on the quality of the Weibull parameters used, but the assumed distribution parameters can neither be confirmed nor refuted by a RDT. From this finding, an improvement approach in the application of a RDT is that it should ideally be preceded by a failure-oriented testing. This is particularly recommended if no failure data from predecessor products are available or if a significantly changed failure behaviour of the new product is to be expected.

Frequently, failure-free testing is planned, carried out and evaluated, for example to release the next sample loop as part of a product development process. After the actual end of the RDT, however, it should be continued for as long as possible with the goal of continuing to test as many prototypes as possible until failure. This failure data can then be used to perform a Weibull analysis and thus determine the Weibull parameters for the investigated product design. These distribution parameters now allow a critical review of the planning and evaluation activities of the tested sample phase and enable a reduction of the uncertainties of the subsequent sample phase. Downstream failure-oriented testing should take place no later than the final release for series production of a new product in order to be able to confirm the assumptions for the planning and, above all, the validity of the results of the product testing and thus minimize a possible field risk.

Another way of creating the most robust possible assumptions about failure behaviour and thus a planning and evaluation basis with few uncertainties can be to use stochastic fatigue simulations [3]. From these analyses, initial product-specific estimates for the distribution parameters can then be determined.

4. CONCLUSION

The previous considerations show that for RDTs assumptions about the failure behaviour of the specimens are necessary, although failure-free testing is performed. The effects of uncertain Weibull parameters are sometimes very clear and must be taken into account; in particular, rough estimates of failure behaviour must be viewed critically. As shown in the case studies, an apparently solidly designed RDT can be planned with a supposedly high confidence level. However, based on imprecise, assumed distribution parameters, the conclusions drawn from the results of the RDT may not be very robust, which may lead to a risky product release with risks of field performance of the serial products. This risk must be identified and minimized with suitable measures, such as an adjusted sample size. Some possible improvement approaches for a planning and evaluation basis for RDTs that is as robust as possible always consists of an accurate estimation of the failure behaviour, which, however, originates from a failure-oriented product testing. Thus, RDTs should always be applied in the context of upstream or downstream failure-oriented product testing.

References

[1] B. Anderhofstadt, S. Spinler, "*Factors affecting the purchasing decision and operation of alternative fuel-powered heavy-duty trucks in Germany – A Delphi study*", Transportation Research Part D: Transport and Environment, Volume 73, pp. 87-107, (2019).

[2] B. Bertsche, "*Reliability in Automotive and Mechanical Engineering - Determination of Component and System Reliability*" Springer, 2008, Berlin, Heidelberg.

[3] M. Dazer, "*Zuverlässigkeitstestplanung mit Berücksichtigung von Vorwissen aus stochastischen Lebensdauerberechnungen*", University of Stuttgart, 2019.

[4] M. Dazer, D. Brautigam, T. Leopold, B. Bertsche, "*Optimal Planning of Reliability Life Tests Considering Prior Knowledge*", Proceedings - Annual Reliability and Maintainability Symposium, pp. 1-7, 2018.

[5] M. Dazer, T. Herzig, A. Grundler, B. Bertsche, „*R-Optima: optimal planning of reliability tests*", Proceedings - IRF2020: 7th International Conference Integrity-Reliability-Failure, pp.695-702, 2020.

[6] A. Grundler, M. Dazer, B. Bertsche, "*Effect of Uncertainty in Prior Knowledge on Test Planning for a Brake Caliper using the Probability of Test Success*", ", Proceedings - Annual Reliability and Maintainability Symposium, pp. 1-9, 2021.

[7] H. Guo, H. Liao, A. Gerokostopoulos, A. Mettas, "*Design of reliability demonstration testing for repairable systems*", Proceedings - Annual Reliability and Maintainability Symposium, pp. 1-6, 2011.

[8] H. Guo, H. Liao, "*Methods of Reliability Demonstration Testing and Their Relationships*", IEEE Transactions on Reliability, vol. 61, no. 1, pp. 231-237, March 2012.

[9] H. Guo, E. Pohl, A. Gerokostopoulos, "*Determining the Right Sample Size for Your Test: Theory and Application*", Proceedings - Annual Reliability and Maintainability Symposium, 2013

[10] D. B. Kececioglu, "*Reliability & Life Testing Handbook*", 2nd ed, PTR Prentice Hall, 1994, New Jersey.