

Modelling the Effect of Maintenance-Induced Failures from Periodic Testing of Safety-Critical Equipment as Part of RCM Analysis in the Oil and Gas Industry

Jon Tømmerås Selvik^{1,2}(⊠) and Hans Petter Lohne²

¹ University of Stavanger, Stavanger, Norway jon.t.selvik@uis.no
² NORCE Norwegian Research Centre, Stavanger, Norway

Abstract. Determining appropriate maintenance programmes for technical inventory is recognized as important for quality reliability and safety management in the oil and gas industry. The programme could be achieved through reliabilitycentred maintenance (RCM) analysis, where safety-critical equipment with potential for hidden failures is given particular attention. Output of the analysis is seen in combination with relevant requirements to perform functional testing of the equipment. The testing involves collecting and analysing data for verification of acceptable reliability and safety levels during the operational phase. This testing is often required in periodic intervals, where shorter intervals might be required initially or after failures for more control. Despite the intention of such activity, it could however influence equipment conditions in a negative way and over time contribute to a reduced reliability performance, i.e., lead to maintenance-induced failures. In this paper, focus is on periodic testing of the component 'downhole safety valve' (DHSV), and mechanisms leading to its failure. We consider the use of an age-adjusting imperfect repair model for analysing the effect of maintenanceinduced DHSV failures and discuss the influence of recommended industry guidance. We particularly discuss the benefits of a test strategy having initially one to three months intervals, compared with an alternative strategy with constant sixmonth or one-year intervals. Based on the analysis, the 12-month interval gives the highest overall probability of failure on demand despite reducing the probability for maintenance-induced failures. There is a marginal difference between the other two alternatives, where then the selected distributions and uncertainties play a larger role. Barrier data collected by the Petroleum Safety Authority Norway (RNNP project data) is used for the analysis.

1 Introduction

Reliability centred-maintenance (RCM) is a widely used technique within reliability engineering practise, with over 40 years of history of successful application in various industries [1], including the oil and gas industry. It provides a way of selecting the appropriate maintenance policy and assessment of periodicity (scheduling), where high

attention is given to safety critical equipment, especially passive equipment with 'hidden failure' potential such as blow-out preventers (BOPs) and downhole safety valves (DHSVs). This paper will focus mainly on the latter.

RCM is a way to identify critical items and to develop appropriate maintenance programs maintaining inherent reliability [2]. Assessment of maintenance intervals (maintenance optimization) being a main step. A description of the steps is given in Rausand et al. [3, p. 392]. The assessment covers attributes such as requirements, consequences, cost, and the probability of failure on demand (PFD).

Functional testing is normally the default for passive safety critical equipment unless there is failure-alerting condition-monitoring. When considering DHSV reliability as part of the RCM analysis, for operations on the Norwegian Continental Shelf, such valves are normally subject to a six-month functional test interval as recommended in NORSOK D-010 [4]. The first year being a bit special, consisting of three one-month and then three three-month intervals, afterwards resuming six-month intervals unless some functional failure occurs. The interval could be adjusted depending on the reliability demonstrated, but also based on planning of other maintenance activity, i.e., grouping of maintenance tasks. Testing the functionality (proof-test) makes it possible to detect DHSV failures before a demand and a potentially dangerous situation occurs. A shorter time between the tests influences the expected time between failure and detection. The idea being that shorter intervals is in favour of reliability and safety. Cost aspects might pull in the opposite direction. One could also have maintenance-induced failures, challenging the actual reliability and safety benefits of frequent testing.

When optimising maintenance, the potential for maintenance-inducing failures could be included as part of an imperfect repair model, although the influence is often ignored in practice [5]. According to Dekker [6] a main reason for why, is the lack of adequate tools and methods for identification of such failures, besides the lack of good data. The lack of sufficient degradation experience leads to extensive use of the exponential failure distribution when estimating the DHSV mean time to failure and PFD. However, for DHSVs such a distribution might not be realistic as the industry recommended test schedule indicate a higher failure rate during the first year of operation. A main objective of this paper is to consider the use of an age-based model, comparing the exponential versus the Weibull distribution, as well as different test strategies for the DHSV maintenance optimisation part of RCM. A Weibull distribution might be more realistic and is often pointed to in qualification, but also way more complex to integrate analytical-wise. This as input for development of an imperfect repair model for scheduling of DHSV proof-tests, where the effect of maintenance-induced failures may be studied.

2 Imperfect Repair Modelling

A typical objective or objective function when optimizing maintenance, is cost; the optimization criteria typically expressed as expected cost over a time period. Reliability might also serve as a criterion, e.g., the number of failures or the PFD. When the prime focus is safety, a key is to ensure acceptable safety integrity performance. In general, the criteria can be expressed through time-based models, which are widely applied in the context of RCM; such as P-F interval models, block-replacement, and age-replacement

models (see model descriptions in e.g. Lindqvist [7]; Moubray [1] and, Rausand et al. [3, p. 549]). The influence of maintenance-induced failures for passive safety equipment is discussed in e.g., Hafver et al. [8], where two BOP test schemes are compared, i.e. a scheme with constant testing intervals and a scheme with adaptive scheduling. Adaptive scheduling means that time to next test is adjusted to compensate for changes in failure rate or failure frequency during the operational cycle to maintain the reliability. Such flexibility allows for more efficient test scheduling.

In modelling of repairable systems, such as DHSVs, the impact of imperfect repair is typically reflected by adjustment of the failure frequency w(t) as the starting point following (immediately after) a repair action. The basis is then that the failure frequency immediately after the repair should reflect the quality of the maintenance. The w(t) is sometimes referred to a *rate of occurrence of failure* (ROCOF).

One of the simplest ways is to assign a repair efficiency function (or index) for the maintenance quality, i.e., the probability of maximum condition improvement, and link this to the w(t) or the PFD. This could be modelled binary by assigning $\rho(t)$ as the probability of perfect repair (as-good-as new), with $1-\rho(t)$ as the probability of the item condition being as immediately before the failure occurs (as-bad-as old). Such a concept could be applied for any maintenance policy. The concept could also be extended to multistate as suggested in e.g., Doyen and Gaudoin [9]. It represents a type of imperfect repair modelling based on w(t) adjustment, where each maintenance event could lead to either perfect repair (as the best) or minimal repair (as the worst), or something in between. However, the steepness of the failure rate curve at time t, the w'(t), is not influenced by the adjustments. A different way to adjust, is to make time the basis for w(t) reduction. This could change the steepness of w(t), where the reduction factor or function moves the failure rate (virtually) back in time (age reduction). However, such an adjustment will not have much effect under assumption of an exponential failure distribution, where the w(t) is constant. The reduction factor then gives the percentage of time reduction based on the time elapsed since the previous repair or total elapsed time. The adjustment is then limited to the development recorded for the earlier period.

Assuming that failures only are revealed from proof-testing at intervals with length τ and as-good-as-new after repair, the average PFD can be calculated from:

$$PFD_{AVG} = \frac{1}{\tau} \int_0^\tau F(t)dt \tag{1}$$

Where F(t) is the PFD at time t; P(T < t). For each cycle, and immediately after the repair, the item is virtually moved back to t = 0. However, if the repair is imperfect, unless the failure rate is decreasing, the PFD_{AVG} will be increasing for each test cycle of length τ . F(t) can be calculated with reference to the failure rate z(t):

$$F(t) = PFD = 1 - \exp\left[-\int_0^t z(t)dt\right]$$
(2)

In (2) if z(t) is a constant $z(t) = \lambda$ the integral in (2) equals $\lambda \cdot t$, giving the exponential distribution. F(t) for the Weibull distribution is presented in Sect. 4. Following each cycle of tests, the underlying failure density distribution and F(t) may then change depending on the repair quality. Let $F_i(t)$ denote the PFD in the interval from immediately after

test number i-1 to test *i*. For constant test intervals with length $\tau: F_1(\tau) \to F_2(2\tau) \to F_3(3\tau) \to \dots \to F_{n-1}[(n-1)\tau] \to F_n(n\tau).$

A change in F_i would then reflects the change in z(t) induced by the maintenance. In addition, the maintenance effect could be modelled by adjusting the virtual start time for cycle *i*. Instead of start at t = 0, it could start at a point in time reflecting the current equipment condition at start-up to match the PFD. Such a modelling may allow for adjustment of the initial cycle level and for the development up to next test. The PFD_{AVG} for cycle *n* specifically could be expressed as:

$$PFD_{AVG}(\mathbf{n}) = \frac{1}{\tau} \int_{S(n)}^{S(n)+\tau} F_n(t)dt; \text{ for } n > 0$$
(3)

where S(n) is a function assigning virtual start-up time of cycle *n*. This is used as basis for modelling the effects of maintenance-induced DHSV failures in Sect. 4. For more details and presentation of different imperfect repair models, we refer to reviews of imperfect repair models given in e.g., Wang and Pham [10, p. 13], Pham and Wang [11]. See also Rausand et al. [3, p. 455] and Nakagawa [12, p. 171; 13].

3 The Downhole Safety Valve Situation at NCS

DHSV is a main barrier element in offshore wells and plays a key role for safety management for oil and gas facilities. Each year the Petroleum Safety Authorities Norway (PSA) publish a risk level report including a DHSV reliability status. Figure 1 gives an overview of the development in fraction of failed tests from 2002 to 2019 based on data reported from the oil and gas companies to the PSA (see [14]).

In total over the period 89,514 tests are registered, with 2,582 failures; see also Table 1. As indicated by the figure, there is a significant increase in fraction of failures per test, and data show that 35 out of 80 facilities have a fraction above the critical level of 0.02 in 2019. For the full period almost half of the facilities are above this critical level. Assuming a six-month test interval this gives a constant failure rate $\lambda = 6.68 \cdot 10^{-6}$ /h, meaning an expected time to failure of 17.3 years. Due to confidentially issues, field-specific data are not presented.



Fig. 1. Fraction of DHSV failures on number of tests

According to the PSA reporting, previous analysis shows that facilities with more than 20 years of operation, is more prone to DHSV failures compared to younger facilities (for the period 2008 to 2017). The analysis also shows that facilities with 6 to 20 years of operation have a significantly lower fraction of failures; supporting a lower failure frequency in the middle of the lifetime.

Year	Tests	Failures	Fraction of failed
2002	3 851	31	0.00805
2003	3 098	46	0.01485
2004	3 566	67	0.01879
2005	3 322	80	0.02408
2006	4 787	95	0.01985
2007	5 290	153	0.02892
2008	5 863	130	0.02217
2009	4 993	156	0.03124
2010	4 993	135	0.02704
2011	5 227	149	0.02851
2012	5 624	135	0.02400
2013	5 772	149	0.02581
2014	4 592	169	0.03680
2015	5 016	168	0.03349
2016	5 786	200	0.03457
2017	6 051	252	0.04165
2018	6 032	243	0.04029
2019	5 651	224	0.03964

Table 1. Number of DHSV tests with failures at the Norwegian Continental Shelf

A main challenge for the failed valves is the failure mode 'leakage in closed position'. A presentation by Molnes [15] shows that this specific failure mode accounts for around 35% of historical failures, being the dominating one. It represents a type of failure that could be traced to the maintenance activity and number of tests performed. The challenge being primarily the robustness of the seal. Number of tests may influence the performance, but also the time between tests if there are long dormant periods (as for having one-year test intervals) with fluids or sand particles eroding the seal, inducing 'sticking' and insufficient closing ability. Several publications, e.g., Vick et al. [16] and Vinzant et al. [17], focus on the sealing technology as a performance limiting factor. We refer to Selvik and Abrahamsen [18] for DHSV reliability review.

4 Modelling the Effect of Maintenance-Induced DHSV Failures

The modelling presented in this section is based on (3), where The PFD_{AVG} for each cycle is interpreted as the mean value for the specific period. The overall PFD_{AVG} can then be calculated from the arithmetic mean over the *n* periods with length τ :

$$Overall \ PFD_{AVG}(n) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\tau} \int_{S(i)}^{S(i)+\tau} F_i(t) dt$$
(4)

For a test cycle *i*, the function S(i) is here seen as a function assigning a value in $[0, \infty)$ based on the stresses accumulated by the *i* tests performed, assuming no use of the valves except the test demands, with $S(i-1) \leq S(i)$; $0 < i \leq n$. S(i) comprise as such the tests' probability of reducing DHSV condition or functionality; with 0 as 'as-good-as-new', τ as 'as-bad-as old', and $t > \tau$ as a decrease in condition and at worst a complete loss of function. It depends on the number of tests, quality of the tests, and the condition at the test. One way is to specify periods of equal length from τ . We select here 45 periods as the number of cycles *m* assumed before the valve is in a 'as-bad-as-old' condition after testing, giving (in hours) when τ is six months:

$$S(i) = \frac{\tau}{m}(i-1) = 96.0(i-1)$$
(5)

We also assume the F(t) to be independent on the cycle number, i.e., $F_0(t)$ equals $F_i(t)$. As input, we refer to the failure data presented in Sect. 4 and will use these as the field-specific results. Regarding the F(t), we consider two common distributions: the exponential distribution with parameter $\lambda = 6.68 \cdot 10^{-6} [h^{-1}]$; (giving a mean time to failure MTTF = 17.3 years), and, the Weibull distribution with rate parameter λ (and scale parameter $1/\lambda$), and shape parameter k = 1.3; and with F(t) = $1 - \exp[-(t \cdot \lambda)^k]$ (MTTF = 15.8 years). The Weibull k parameter is derived from available qualification testing results. The comparison is then achieved by considering three distinct testing strategies: the one with one-year shorter initial tests, a constant six-month strategy, and a maximum 12-month strategy.

For the overall PFD_{AVG} formula (4) under assumption of a constant failure rate λ , and fixed test intervals $\tau_i = \tau$, the formula can be expressed as:

Overall
$$PFD_{AVG}(n) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\tau_i} \int_{S(i)}^{S(i)+\tau_i} \left(1 - e^{-\lambda \cdot t}\right) dt$$
 (6)

$$= \frac{1}{n\tau} \sum_{i=1}^{n} \left[t + \frac{1}{\lambda} e^{-\lambda t} \right]_{S(i)}^{S(i)+\tau} = 1 + \frac{1}{n\tau} \sum_{i=1}^{n} \frac{1}{\lambda} \left(e^{-\lambda \cdot (S(i)+\tau)} - e^{-\lambda \cdot S(i)} \right)$$
(7)

The overall $PFD_{AVG}(n)$ when applying the Weibull distribution can then be expressed as:

Overall
$$PFD_{AVG}(n) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\tau_i} \int_{S(i)}^{S(i) + \tau_i} \left(1 - e^{-(\lambda_i t)^{k_i}} \right) dt$$
 (8)

In practical applications there will be initially higher failure probability. However, this is not included here. The effect would be a higher probability in the first year, although it is uncertain how large this effect would be.

When assuming an exponential distribution the probability of failure at each test is shown in Fig. 2. Longer test intervals have a higher probability of failure from the beginning, but the difference diminishes over the cycles due to maintenance induced wear. Due to the maintenance induced failures, the PFD at the start of a cycle will be larger for shorter test intervals. The PFD_{AVG} for a short test interval policy will therefore become greater than the PFD_{AVG} for a longer interval policy earlier than Fig. 2 might indicate. The 6-month intervals policy will have a higher PFD_{AVG} than the 12-month interval policy after 21.8 years, and with more frequent initial testing already after 17.8 years.



Fig. 2. Probability of the DHSV being in a failed state for a test with exponential distribution, conditioned on that it was functioning at the start of the interval, when the test is performed either at 6 or 12-month intervals or 6-month intervals with more frequent tests initially.

When considering the Weibull distribution, there can be seen an effect of wear in Fig. 3. The more frequent initial test intervals demonstrate a lower probability of being in a failed state for a test during the early years, at the cost of slightly higher probability due to test-induced wear in the later years. There is also a slight curvature to the lines in the plot caused by the increase in failure rate due to the time-adjusted aging. For the Weibull situation, the 6-month intervals policy will have a higher PFD_{AVG} than the 12-month interval policy after 23.0 years, and with more frequent initial testing after 19.2 years; 1.4 years later than for the exponential.



Fig. 3. Probability of the DHSV being in a failed state for a test with Weibull distribution, when the test is performed either at 6 or 12-month intervals or 6-month intervals with more frequent tests initially.

Alternative modelling-approaches are shown below. Figure 4 shows a Weibull distribution where the scale parameter is reduced by a fixed amount per interval. Even though the four additional tests performed in the first year only changes the probability by a small amount initially, the effect grows larger later in life. If the scale is reduced by a percentwise reduction per test a similar difference increase is seen, as shown in Fig. 5. Finally, a scenario where the maintenance activity does not modify the distribution, but rather an independent increasing probability of failure occurring due to the maintenance itself that is then added to the failures that can occur within the interval is shown in Fig. 6. In this case each maintenance activity causes a fixed increase that remains constant over time.



Fig. 4. Probability of DHSV being in a failed state for a test using Weibull distribution, where the scale parameter is reduced linearly by each test, when the test is performed either at 6 or 12-month intervals or 6-month intervals with more frequent tests initially.



Fig. 5. Probability of DHSV being in a failed state for a test using Weibull distribution, where the scale parameter is reduced by a percent for each test, when the test is performed either at 6 or 12-month intervals or 6-month intervals with more frequent tests initially.



Fig. 6. Probability of DHSV being in a failed state for a test using Weibull distribution, where the maintenance activity can cause failures independently of wear state, when the test is performed either at 6 or 12-month intervals or 6-month intervals with more frequent tests initially.

Overall PFD_{AVG} for the different cases are shown in Table 2 below.

Table 2. PFD_{AVG} for different maintenance schemes and different maintenance-induced failure models.

Test intervals	Weibull			Exponential		
	6 months	Initially more frequent	12 months	6 months	Initially more frequent	12 months
Time adjustment	1.54%	1.59%	1.64%	3.85%	3.94%	4.04%
Linear scale reduction	0.72%	0.75%	1.36%	2.01%	2.07%	3.31%
Percentage scale reduction	0.79%	0.83%	1.46%	2.16%	2.22%	3.50%
Independent failure increase	0.85%	0.88%	1.31%	4.24%	4.51%	3.06%

Table 2 presents then the average probability of failure over the 39 years calculated for the different models. The maintenance scheme with 6-month intervals results gives the lowest values in most cases, although for failure independent of running time longer intervals are better with the exponential distribution.

5 Discussion

5.1 The Effect of Maintenance-Induced DHSV Failures

A key is the difference in effect between the test policies. As seen in the above calculations, frequent testing in the initial phase will reduce the maximum probability reached of being in a failed state at a test, at the cost of slightly higher probabilities in later years. From the overall average probability of failure on demand, it is seen that 12-month intervals are worse off in most cases. However, this depends on the severity of damage caused by the testing. It is nevertheless obvious that maintenance is a trade-off between reducing the highest probability of being in a failed state and the overall probability of being in a failed state (such as during the initial phase). This depends on the failure distribution as well.

How maintenance-induced failures occur matters. In most of the examples where the increased testing shifts towards a state of increased wear, either through time-adjustment or through parameter adjustment, the changes seen are not drastic but will accumulate to a greater overall probability of failure over the lifetime. In the examples where failures are actually induced by maintenance, rather than just causing an increase in failure rate during the next cycle, the failure probability will here shift the curves upwards, which illustrates how frequent initial testing can deteriorate the valves quickly. Similar behaviour is difficult to achieve with models that simply "age" the valves when wear-out failures occur relatively late in the example distribution used. Distributions with a finite support (i.e., components with limited lifespan) would be able to develop much stronger differences.

5.2 RCM Value

The modelling of maintenance-induced failures is important to understand if frequent early initial testing is worth the wear to reduce the probability of being in a failed state within a test interval. If there is a fixed cost (in terms of failures) to wear, such as if there is a constant probability of failure due to the maintenance, then the frequency of maintenance can be done by focusing on keeping the maximum probability of failure in an interval as low as possible without considering additional long-term effects. In other cases, the long-term effect of maintenance should be considered, as the total probability of failure during the time the valve is installed can be larger.

When discussing maintenance, other aspects can also be considered, such as the associated costs as well as the needs. If during the early stages there is greater uncertainty in subsurface conditions there is also a greater need for a functioning DHSV, while if the uncertainty is low and therefore low probability of a demand for the DHSV the condition of the valve is less precarious. In such situations one could discuss having more frequent tests when a demand is more likely, and rather increase the test intervals when a demand is less likely to avoid excessive wear.

While several deterministic models are shown here to illustrate some different alternatives to modelling maintenance induced failures, when making decision related to maintenance based on reliability modelling the uncertainty of these models and the underlying failure causes they attempt to represent must be considered.

6 Conclusions

In this paper, DHSV failures and imperfect repair modelling has been presented, including a dataset from the NCS. A Weibull distribution and exponential distribution based on this dataset was then used to illustrate the effect of different models for maintenanceinduced DHSV failures. The results illustrate the cost frequent testing can have on later reliability. To make correct decisions regarding testing, it is important that also the uncertainty around maintenance-induced failures, wear during operation and their interactions are included.

References

- 1. Moubray J (2012) Reliability centered maintenance, 3rd edn. Industrial Press, Incorporated, New York
- Selvik JT, Aven T (2011) A framework for reliability and risk centered maintenance. Reliab Eng Syst Saf 96(2):324–331
- 3. Rausand M, Barros A, Høyland A (2021) System reliability theory models, statistical methods, and applications, 3rd edn. Wiley, Hoboken
- 4. NORSOK D-010 (2021) Well integrity in drilling and well operations, 5th edn. Standards Norway, Oslo
- 5. Jackson C, Mailler B (2013) Post-servicing failure rates: optimizing preventive maintenance interval and quantifying maintenance induced failure in repairable systems. In: Proceedings annual reliability and maintainability symposium (RAMS), Orlando, FL, USA
- 6. Dekker R (1996) Applications of maintenance optimization: a review and analysis. Reliab Eng Syst Saf 51:229–240
- Lindqvist B (2006) On the statistical modelling and analysis of repairable systems. Stat Sci 21(4):532–551
- 8. Hafver A, Oliveira LF, Pedersen FB (2019) Optimal scheduling of tests of safety systems, considering test-induced degradation. In: Proceedings of the 29th European safety and reliability conference, Hannover, Germany
- 9. Doyen L, Gaudoin O (2011) Modeling and assessment of aging and efficiency of corrective and planned preventive maintenance. IEEE Trans Reliab 60(4):759–769
- Wang H, Pham H (2006) Reliability and optimal maintenance. Springer, London. https://doi. org/10.1007/b138077
- 11. Pham H, Wang H (1996) Imperfect maintenance. Eur J Oper Res 94(3):425-438
- Nakagawa T (2005) Maintenance theory of reliability. Springer, London. https://doi.org/10. 1007/1-84628-221-7
- Nakagawa T (2008) Replacement and preventive maintenance models. In: Misra KB (ed.) Handbook of performability engineering. Springer, London. https://doi.org/10.1007/978-1-84800-131-2_49
- 14. PSA (2020) Summary report Norwegian Continental Shelf 2019 Trends in risk level in the Norwegian Petroleum industry. Petroleum Safety Authority Norway (PSA). https://www.ptil.no/en/technical-competence/rnnp/rnnp-2019/summary-report/
- Molnes E (2012) Leveraging data for high value decisions on well OpEx optimization. Presentation at the US-Norway technology partnership conference, Houston, Texas, USA, 29 March
- 16. Vick J, et al (2005) Unique safety valve design using a floating magnetic coupler and metalto-metal sealing capability to provide step-change improvements in reliability, safety, and cost. OTC paper 17455. http://www.onepetro.org
- 17. Vinzant M, et al (2004) Enhanced reliability with unique tubing-retrievable safety valve design. SPE paper 88627. http://www.onepetro.org
- Selvik JT, Abrahamsen EB (2013) How to describe well barrier reliability: a SCSSV review study. In: COMADEM 2013: proceedings of the 26th international congress on condition monitoring and diagnostics engineering management, Helsinki, Finland