

New method for updating failure rates and proof test intervals of equipment groups within safety instrumented systems

Solfrid Håbrekke

Software Engineering, Safety and Security, SINTEF Digital, Norway. E-mail: solfrid.habrekke@sintef.no

Mary Ann Lundteigen

Department of Engineering Cybernetics, NTNU, Norway. E-mail: mary.a.lundteigen@ntnu.no

Stein Hauge

Software Engineering, Safety and Security, SINTEF Digital, Norway. E-mail: stein.hauge@sintef.no

Safety instrumented systems (SIS) are to be followed up during operation at a facility. An important part of SIS follow-up is to ensure that all relevant SIS requirements from the design phase are fulfilled during the entire lifetime of the facility. The safety integrity level (SIL) and the corresponding probability of failure on demand (PFD) requirements should be verified regularly. The PFD of a component is a function of the dangerous undetected failure rate and the proof test interval of the component. For follow-up purposes it is relevant to group comparable components within the same equipment group, such as shutdown valves or level transmitters. For analysis purpose, the method focuses on equipment groups where all components within the same group are assumed to have a comparable (common) failure rate and PFD budget. This paper presents an approach for estimating updated failure rates and assessing the proof test intervals based on operational experience. The method for updating failure rates is based on a Bayesian failure rate estimation with additional periodisation. For proof test interval assessment two methods are given: One method based on comparison of the “experienced” failure rate and the failure rate assumed in design, and one method based on comparison of the “experienced” PFD and the PFD budget requirement.

Keywords: Reliability, safety instrumented systems, safety instrumented functions, probability of failure on demand, low demand mode, failure rate, proof test interval, follow-up.

1. Introduction

Follow-up of safety instrumented system (SIS) is necessary to ensure that the safety integrity and the corresponding requirements and assumptions from design of every SIS are maintained throughout the operational lifetime of a given facility. Updating failure rates and assessing proof test interval of SIS components are key parts of the SIS follow-up. The probability of failure on demand (PFD) is a safety integrity requirement from design that needs to be followed up during the entire operation phase of a facility. The PFD is a function of the failure rate and the proof test interval. Thus, updated failure rate estimates based on recorded operational data determine if the proof test interval can be increased or should be reduced compared to the PFD requirement. Requirements to be followed up during SIS operation are given in Petroleum Safety Authority (PSA) regulations and international standards such as IEC 61508 and IEC 61511 (IEC 2010, 2016), in company

governing documents as well as facility specific requirements.

This paper provides methods and formulas for estimating updated failure rates and proof test intervals based on operational data. The approach has been developed during the update of the report “Guidelines for follow-up of Safety Instrumented Systems (SIS) in the operating phase” (Hauge and Lundteigen 2008). This guideline has been widely adapted by the Norwegian petroleum industry and has been used as a basis for vendors and consultancy companies to develop SIS follow-up applications for several oil and gas operators. The new approach will be described in the updated guideline (Håbrekke et. al. 2020).

Alternative methods and variants of methods for calculating the proof test interval based on operational experience have been identified from modifications of the original method given in Hauge and Lundteigen (2008), extensive experience from use in operational reviews, identified improvement areas, and a survey among operators on their methods in use.

The main application of the SIS follow-up guideline is for the oil and gas industry. However, methods and formulas are also relevant for other equipment (e.g. process safety valves) and other industry sectors, particularly for the process sector where SIS functions operate in the so-called low-demand mode of operation. This means that the SIS functions are demanded seldom, which according to IEC 61511 and IEC 61508, the standards on functional safety of SIS, is less than once per year (IEC 2010, 2016).

The remaining structure of the paper is as follows: Firstly, some SIS follow-up practices are given. Secondly, the main steps and the assumptions of the new method are listed. Thirdly, the new method and the corresponding algorithm and formulas are provided. Finally, the paper discusses some aspects related to the new method together with concluding remarks.

2. SIS follow-up Practices

SIS follow-up is used to describe the activities needed in the operational phase of a facility to monitor and maintain the functional safety and safety integrity of the safety instrumented functions (SIFs) that are performed by the SIS. The functional safety is about the functional requirements describing how the SIFs *shall perform* upon demands and is documented in system control diagrams (SCDs), piping and instrument diagrams (P&IDs), cause and effects (C&Es), and the safety requirements specification (SRS). The safety integrity relates to how the SIFs (actually) respond, considering the potential presence of degradations or faults. The safety integrity of a SIF can be measured by the “average probability of failure on demand”, i.e. the PFD, and is influenced by the failure rates, proof test intervals, redundancy, and vulnerability to common cause failures.

IEC 61511 (IEC 2016), is widely adopted by the process industry sector and defines general requirements to the SIS follow-up activities. The aforementioned SIS follow-up guideline is a practical supplement to the standard and the content of the guideline has been adapted by the Norwegian guideline on the application of IEC 61508 and IEC 61511 (Norwegian Oil and Gas 2020).

There have also been other international initiatives on practical guidelines for SIS follow-up, for example, the ISA guideline TR84.00.02-

2015 (ISA 2015), which provide formulas and practical guidelines on the failure classification. Specific issues that are related to SIS follow-up such as partial proof testing is discussed ISA-TR96.05.01 (ISA 2017). These guidelines may be considered in conjunction with another guideline ISA-TR.84.00.03 (ISA 2019), which provides more support on the practical procedures, roles, and responsibilities.

The method presented in this paper can be regarded as a supplement to existing standards and practices, with more guidance on failure rate estimation based on operational data, and on the assessment of proof test intervals.

3. Main Steps and Assumptions

3.1 Main steps

The method presented in this paper requires a set of input data to update the failure rates and proof test intervals for an equipment group. The main steps of the method are illustrated in Fig. 1. The Bayesian approach has been selected for the failure rate calculations as it weights the prior knowledge with new experience. SIS components are built to be highly reliable, which means that there may be zero critical failures experienced. This does not mean that the failure rate is zero, and the estimate can benefit from also having knowledge about the number of failures in the past.

As shown in Fig. 1, the input data constitutes:

- Operational experience, i.e. failure data for an equipment group during an *observation period*. The failures of interest are the *dangerous undetected (DU) failures*.
- Prior knowledge about the DU failure rate, proof test interval and PFD budget (or target) for the equipment group.

An observation period is the interval of time (calendar time) between the start date and end date of a failure data collection interval. DU failures are failures that prevent components to perform their safety functions and where the failures are not revealed immediately after occurrence, but rather during proof tests, on demand or by random observations (corrective maintenance, inspection, etc.). An example of a DU failure is a shutdown valve that fails to close within the response time requirement upon a proof test.

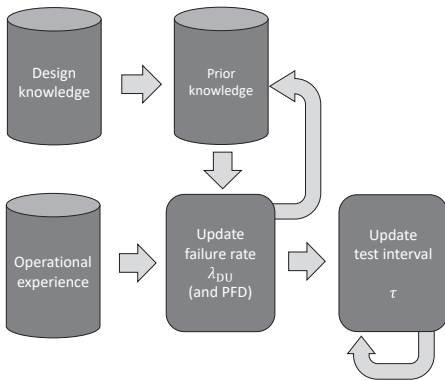


Fig. 1. The method steps.

The DU failure may have been latent for a while but is not revealed until the valve is activated by a proof test or demand. For simplicity, 'failure rate' refers to 'DU failure rate' through the rest of this paper.

The prior knowledge about the failure rate is the failure rate from the preceding observation period (which also could be the design phase when the facility is newly put into operation) together with the uncertainty of this (input) failure rate. The PFD budget (or target) denotes maximum PFD allocated to a subsystem of a SIF, so that, when summed over all subsystems, will meet the safety integrity level (SIL) requirement of the SIF. With this input, the method to assess the proof test interval is split into the following sub-steps:

- (i) calculate the failure rate of the component type using a Bayesian approach,
- (ii) calculate the associated 70% upper credibility limit of the failure rate as conservative estimate (see below),
- (iii) update the PFD,
- (iv) calculate a new proof test interval.

The output of the method is a recommendation about a new (or maintained as before) proof test interval. Before making the final decision upon changing the proof test interval, a qualitative checklist should be consulted as provided by Håbrekke et. al. (2020). This is to assess the quality of the data collection, confirm underlying assumptions, and to follow manufacturer recommendations, operational and maintenance constraints.

3.2 Assumptions

The main assumptions regarding the method are presented below.

3.2.1 Grouping of components

For practical reasons and to achieve a better statistical confidence, it is common to group components that share similar function and properties, so that the failure rate is estimated on the basis of failures reported to components that belong to the same equipment group. Examples of typical SIS equipment groups are smoke detectors, level transmitters and shutdown valves. The components within a defined equipment group are assumed to be homogenous, i.e. attached with the same failure distribution. Every component within the same equipment group is then associated with the same failure rate, proof test interval and PFD target. Equipment groups including various technologies for the same function (e.g. various measuring principles for level transmitters) may be considered as inhomogeneous. Definitions of SIS equipment groups is suggested in Hauge et. al. (2019). It must be ensured that sufficient aggregated operational experience has been gathered to assess if the proof test interval should be updated. The more components within the group and/or the longer the observation period is, the more extensive becomes the operational experience. Grouping will therefore be a trade-off between having enough aggregated operating time and capturing the important attributes influencing the failure rate of the equipment group for a given period (see below).

3.2.2 Modelling assumptions

The assumptions regarding the method are as follows:

- All components within the equipment group have been proof tested or activated at least once in the observation period.
- The input failure rate and its conservative estimate (prior knowledge) includes both random hardware and systematic failures, i.e. the expected operational failure rate.
- Time to perform proof tests, repair or to replace a failed component or other downtime is negligible compared to the time between proof tests (or activations/demands).

- The proof test coverage is 100%, i.e. all DU failures are possible to reveal upon proof tests.

3.2.3 Aggregated operating time

The aggregated operating time (T_i) for observation period i is determined by the length (calendar time) of the observation period (t_i) and the number of components within the equipment group (n_i). The following should be considered when calculating the aggregated operating time for an observation period: Adjusting for removed or added components during the observation period or since last observation period, excluding components not tested or activated in the observation period, and excluding components that have been out of service or in passive standby during a longer observation period (e.g. > 5% of the aggregated operating time).

3.2.4 Selecting observation period

The length of the observation period for which operational experience is gathered for an equipment group should be selected so that it provides sufficiently aggregated operating time within the observation period to give the necessary confidence in the updated failure rate $\lambda_{DU,i}$ and the recommendations about proof test intervals. The operational experience of an equipment group within an observation period should preferably be weighted more than 50% compared to the prior data (from previous observation periods). For the case where the conservative estimate of the input failure rate is twice the input failure rate, this corresponds to $\lambda_{DU,i-1} \cdot T_i > 1$ (e.g. with an input failure rate of $1 \cdot 10^{-6}$ and operating time of 10^6 hours). Here $\lambda_{DU,i-1}$ is the failure rate from the previous observation period (or from design). Based on the above, the suggested minimum length of an observation period becomes:

$$t_i > \frac{1}{n_i \cdot \lambda_{DU,i-1}}. \quad (1)$$

4. Proposed Method

The method proposed in this paper has two main steps:

1. Update the failure rate (and the corresponding PFD).
2. Assess the proof test interval.

4.1 Step 1. Updating the failure rate

The failure rates are calculated using the Bayesian approach, where the operational experience is used in combination with prior knowledge about the failure rate. In addition, an iterative approach with several possible observation periods has been introduced.

A prerequisite for the Bayesian approach is an evaluation of the uncertainty of the input failure rate by providing a conservative estimate of this failure rate.

Fig. 2 illustrates how the input failure rate is selected for the Bayesian update for some consecutive observation periods. The updated failure rate from the previous observation period together with its conservative estimate becomes the prior knowledge for the next failure rate update. For the first observation period, the prior knowledge typically is the failure rate from design and its corresponding conservative estimate.

The estimate for the updated failure rate for observation period i based on operational data combined with prior knowledge is, see e.g. Vatn (2006):

$$\lambda_{DU,i} = \frac{\alpha_i + x_i}{\beta_i + n_i \cdot t_i} \quad (2)$$

where the uncertainty parameters β_i and α_i for the observation period are:

$$\beta_i = \frac{\lambda_{DU,i-1}}{(\lambda_{DU-CE,i-1} - \lambda_{DU,i-1})^2} \quad (3)$$

and

$$\alpha_i = \beta_i \cdot \lambda_{DU,i-1} = \frac{\lambda_{DU,i-1}^2}{(\lambda_{DU-CE,i-1} - \lambda_{DU,i-1})^2} \quad (4)$$

Here $\lambda_{DU-CE,i-1}$ is the conservative estimate of the input failure rate expressing the uncertainty (confidence) in the input failure rate $\lambda_{DU,i-1}$. This uncertainty will implicitly decide how the existing failure rate is weighted against the new operational experience. Suggested choices for $\lambda_{DU-CE,i-1}$, for observation period $i = 1$ and observation periods $i > 1$, respectively, are:

$$\lambda_{DU-CE,0} = \max\{2 \cdot \lambda_{DU,0}, 10^{-7}\} \quad (5)$$

and

$$\lambda_{DU-CE,i-1} = \max\left\{\frac{z_{0.15} \cdot 2(\alpha_{i-1} + x_{i-1})}{2(\beta_{i-1} + n_i \cdot t_{i-1})}, 10^{-7}\right\} \quad (6)$$

The first value in each bracket are suggested as defaults.

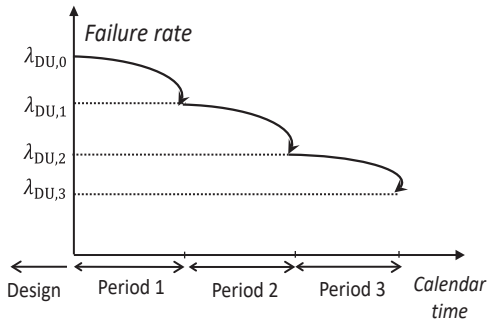


Fig. 2. Periodisation and input failure rates.

For observation period $i > 1$ this corresponds to the upper bound of the 70% credibility interval for the failure rate $\lambda_{DU,i-1}$ (Rausand and Høyland 2004). Here $z_{0.15,v}$ denotes the upper 15% percentiles of the χ^2 -distribution with v degrees of freedom, i.e. $P(\chi^2 > z_{0.15,v}) = 0.15$. In situations where v is not an integer, an interpolation in the χ^2 -distribution may be performed. For observation period 1, the user may not have any additional information about the uncertainty of the design failure rate. Then $\lambda_{DU-CE,1} = 2 \cdot \lambda_{DU,0}$, which reduces Eq. (3) and Eq. (4) to $\beta_i = 1/\lambda_{DU,0}$ and $\alpha_i = 1$, respectively.

Note that the lower failure rate limit of 10^{-7} per hour is specified to avoid that very low input failure rates totally outweighs the operational experience. This can also be interpreted as if the operational failure rate for any equipment group is never assumed to be lower than 10^{-7} per hour, based on generic data from PDS data handbook, Håbrekke et. al. (2013) and thorough reviews of operational data on Norwegian oil and gas facilities, both offshore and onshore.

4.2 Step 2. Updating the proof test intervals

If operational experience proves that the equipment has a reliability that differs significantly from what was assumed in the design phase or in previous observation period(s), there may be room for changing the proof test interval. It is implicitly assumed that the current proof test interval (based on data from the previous observation period), τ_{i-1} , in combination with the assumed failure rate fulfils, the relevant SIL requirements.

A limitation on maximum doubling or halving of the current proof test interval shall ensure that

the proof test interval is not altered too much at a time. Allowed proof test intervals are (on a discrete scale): 1 month, 2 months, 3 months, 4 months, 6 months, 9 months, 12 months, 18 months, 24 months, 36 months, and 48 months. The maximum allowed length of the proof test interval is 48 months assuming that the functional status of all safety critical equipment shall never be verified less frequently than this.

Below are suggested two approaches for calculating the updated proof test interval; one approach based on PFD requirement and one approach based on failure rate assumption from design. Some operators may use one method while others may use the other. However, it is recommended to keep to the same method throughout the lifetime of a facility.

4.2.1 Approach 1 – PFD target

In this PFD approach, the test interval is optimised based on how the recent operational PFD relates to the PFD target from design – regardless of the actual failure rate and test interval from design. The recent operational PFD is calculated from the updated failure rate and existing test interval. A prerequisite for this approach is a common PFD target established for the equipment group under consideration. This may e.g. be the maximum PFD target for a component (belonging to the group) within a SIF. The algorithm is as follows: Find the highest allowable proof test interval τ_i that fulfils the PFD target (PFD_t) for the equipment with the upper 70% credibility interval value of $\lambda_{DU,i}$. With the 70% credibility interval given by e.g. Rausand (2014) and the relation $PFD = \lambda_{DU} \cdot \tau/2$ we get the following inequality for the optimized proof test interval:

$$\tau_i \leq 2 \cdot \frac{PFD_t}{\lambda_{DU,i}^{70U}} = \frac{4 \cdot PFD_t \cdot (\beta_i + n_i \cdot t_i)}{z_{0.15,2(\alpha_i + x_i)}} \quad (7)$$

where β_i and α_i are given by Eq. (3) and Eq. (4), respectively. Note that the proof test interval must be rounded down to the nearest allowable proof test interval, and that the proof test interval cannot be more than doubled or halved compared to the existing proof test interval.

4.2.2 Approach 2 – Failure rate

In this failure rate approach, the estimated operational failure rate is compared to the failure

rate assumed from design, which has been used to verify that the SIL and PFD requirements are fulfilled. A doubling of the operational failure rate indicates a possible halving of the proof test interval and vice versa. When having several observation periods, the new operational failure rate is compared to the previous. The prerequisite for this approach is that the failure rate and proof test interval assumed from design is known and fulfils the PFD requirement. The algorithm is as follows: Find the highest allowable proof test interval τ_i with the upper 70% credibility interval value of $\lambda_{DU,i}$, corresponding to the design failure rate and proof test interval assumed in design, i.e.

$$\tau_i \leq \frac{\lambda_{DU,0} \cdot \tau_0}{\lambda_{DU}^{70U}} = \frac{\lambda_{DU,0} \cdot \tau_0 \cdot (\beta_i + n_i \cdot t_i)}{Z_{0.15,2(\alpha_i+x_i)}} \quad (8)$$

Note that for periods $i > 1$, the doubling/halving restrictions is to be compared to the current proof test interval, τ_i , rather than the proof test interval assumed in design, τ_0 .

5. Discussions and Considerations

5.1 Periodic vs. aggregated failure rate

The Bayesian failure rate estimate weights the past experiences (prior failure data) against the (new) operational data. Within this envelope, there are two main approaches to how the updated failure rate is calculated:

- *Aggregated failure rate:* All past operational experience is aggregated in the updated failure rate estimate, i.e. not only the most recent observation period.
- *Periodized failure rate:* There have been several adjacent observation periods in the past, and only the present observation period is included in the operational data. The previous observation periods become the fundament for the prior.

The aggregated Bayesian failure rate is estimated based on all aggregated experience up to current observation period, always with the design failure rate and its conservative estimate as the prior knowledge. The periodic Bayesian failure rate, on the other hand, is estimated based on the experience from the most current period only, with the result from the previous Bayesian failure rate estimation (from previous observation period) as the prior knowledge.

The periodic failure rate is more sensitive to new operating experience compared to the aggregated failure rate from observation period two and out, as the periodic failure rate more heavily weights the current observation period (which also may be the most relevant). The aggregated failure rate is not able to distinguish remote operating experience from more recent experience. Failures experienced several years ago may not be relevant (are “out-dated”) due to out-dated equipment, implementation of mitigating measures or other changes in operation and maintenance ensuring that the failures will not re-occur in future operation at the facility. A special consideration if using the aggregated failure rate is to remove “out-dated” failures from the operational data prior to the failure rate calculation. To remove “out-dated” failures from the historical statistics must always be justified and documented.

Due to the periodisation, the comparison of failure rates or PFD is always performed against the previous observation period. This implies also that the reassessment of the test proof interval is made against the proof test interval from the previous observation period. While the original method (Hauge and Lundteigen, 2008) had a limitation that the proof test interval could not be more than doubled or halved compared to what was assumed in the design phase, the periodisation allows more incremental changes and could lead to a proof test interval that is more than doubled or halved compared to initial design assumptions.

5.2 Credibility interval vs. confidence interval

Most of the methods identified for calculating updated proof test intervals consider the uncertainty level associated with the new updated failure rate. This uncertainty is often expressed as the confidence level, for example in IEC 61511 (IEC 2016). However, when using Bayesian approach to calculate the updated failure rate, it is more correct to use the credibility interval over the confidence interval. Unlike the confidence interval, the credibility interval captures our current uncertainty in the location of the failure rate within an interval, as opposed to the confidence interval that only captures the uncertainty about whether the interval contains the true failure rate or not. Also, the credibility interval captures the previous experience in

addition to the most recent experience and is particularly to prefer in the periodisation approach compared to the confidence interval.

An observation from analyses carried out, indicates that the credibility interval (in most cases) is narrower than the confidence interval, and not very sensitive to the new experience but rather influenced by its input failure rate. The confidence interval on the other side is independent of the initial failure rate but rather sensitive to changes in the operational data (from one observation period to another).

5.3 70% upper credibility limit

The result of the proof test interval assessment, i.e. the suggested proof test interval, should neither be too conservative nor to optimistic. In Eq. (7) and Eq. (8) the 70% upper credibility limit has been applied for the proof test interval assessment. Higher credibility limits, e.g. 90%, is found to give rather conservative approach (suggesting short intervals) unless rather comprehensive operational experience have been gathered.

IEC 61511 states that any failure rate data used should have a confidence level of at least 70%, and a similar requirement is found in IEC 61508 (IEC 2020, 2016). Hence, both IEC standards indicate that one should be conservative, and the recommended approach is to choose the upper 70% confidence value for the failure rate. This means that we are 70% confident that the "true" failure rate (and corresponding PFD) is below this conservative 70% failure rate (or PFD) estimate.

6. Concluding Remarks

This paper presents an approach for updating failure rates and a pair of simple methods for calculating proof test interval for equipment groups of safety instrumented systems.

Oil and gas operators have commented that they will follow up the PFD target and perform proof test interval assessment based on the estimated PFD rather than comparing the experienced failure rate with the failure rate from design, which the operator may not consider as the "correct" failure rate. In addition, an analogue method comparing failure rates may be relevant. Hence, it is beneficial to have a pair of analogue methods – one PFD method and one failure rate method – that give the same result, i.e. suggest the

same proof test interval based on the same operational experience.

The estimation of the failure rate (and the PFD) should also be analogue to the estimation of the corresponding interval, i.e. by applying the credibility interval for the Bayesian failure rate estimate.

The method should be applicable both for facilities with limited operational experience and for facilities that are approaching their end of lifetime. Modifications, changes in operation or maintenance, etc. may cause changes in the failure rate and such changes will be captured faster with a periodisation than without. Periodisation is also relevant when updating the proof test interval several times. The method proposed in Lundteigen and Hauge (2008) stated that the proof test interval could not be more than doubled or halved (aggregated method) compared to the proof test interval from design. This restriction is no longer relevant for the methods comparing with the proof test interval from the previous observation periods. There are examples of equipment groups, e.g. smoke detectors with an original proof test interval of 12 months having rather good failure record for several years indicating the possibility for increasing the proof test interval beyond 24 months, e.g. up to 36 months. Finally, periodisation ensures that the observation period considered is weighted the most compared to historic observation periods.

Acknowledgement

The paper presents results from APOS, a joint industry project on automatized process for SIS follow-up. The project is supported by the Norwegian Research Council (Project no. 295902) and 11 industry partners representing oil companies, engineering, consultants and vendors of control and safety systems.

References

- Hauge, S., Håbrekke, S., Lundteigen M. (2019). *Standardised failure reporting and classification of SIS failures in the petroleum industry*. SINTEF report.
- Hauge, S., Lundteigen, M. (2008). *Guidelines for follow-up of Safety Instrumented Systems (SIS) in the operating phase. Edition 1*. SINTEF Report A8788.
- Håbrekke, S., Hauge, S., Onshus, T. (2013). *Reliability Data for Safety Instrumented Systems*. SINTEF Report A24443. ISBN 978-82-536-1334-5.
- Håbrekke, S., Lundteigen, M., Hauge, S. (2020). *Guidelines for follow-up of Safety Instrumented*

- Systems (SIS) in the operating phase*. Edition 2 (draft). SINTEF report 2020:00014.
- IEC (2010). *IEC 61508. Functional safety of electrical/electronic/programmable electronic safety-related systems*. International Electrotechnical Committee.
- IEC (2016). *IEC 61511. Functional safety – Safety instrumented systems for the process industry sector*. International Electrotechnical Committee.
- ISA (2015). *ISA-TR84.00.02 Safety Integrity Level (SIL) Verification of Safety Instrumented Functions*. International Society of Automation.
- ISA (2017). *ISA-TR96.05.01 Partial stroke Testing of Automated Valves*. International Society of Automation.
- ISA (2019). *ISA-TR84.00.02 Automation Asset Integrity of Safety Instrumented Systems (SIS)*. International Society of Automation.
- Norwegian oil and gas (2020). *Guideline 070 on the application of IEC 61508 and IEC 61511 on the Norwegian Petroleum Industry*.
- Rausand, M. (2014). *Reliability of Safety-Critical Systems. Theory and Applications*. Wiley. ISBN 978-1-118-11272-490000.
- Rausand, M., Høyland A. (2004). *System Reliability Theory. Models, Statistical Methods, and Applications*. Wiley. ISBN 0-471-47133-X900000.
- Vatn, J. (2006). *Procedures for updating test intervals based on experience data*. ESREDA seminar on Reliability of Safety-Critical Systems.