Chapter 2

Safety demonstration process

This document is basically a summary of chapter 2 in the textbook *Demonstrating safety of software-dependant systems*. Some repetition from chapter 1 is also included.

2.1 What is safety demonstration?

With safety demonstration we mean that we use arguments to claim that our solution is safe. Argumentation theory, or argumentation, is the interdisciplinary study of how conclusions can be reached from premises through logical reasoning. It includes the arts and sciences of civil debate, dialogue, conversation, and persuasion. It studies rules of inference, logic, and procedural rules in both artificial and real-world settings (van der Meulen and Myhrvold, 2022).

2.2 The Safety 4.0 main steps to safety demonstration of novel technology

Figure 2.1 shows the main steps of demonstrating safety of novel technology.

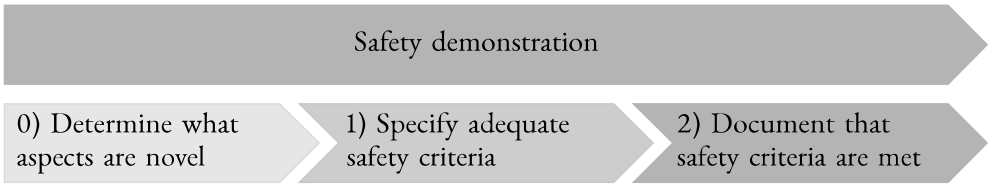

Figure 2.1: Safety 4.0—the main steps to safety demonstration of novel technology (van der Meulen and Myhrvold, 2022)

Safety 4.0 defines safety demonstration as: Documentation, based on evidence and struc-

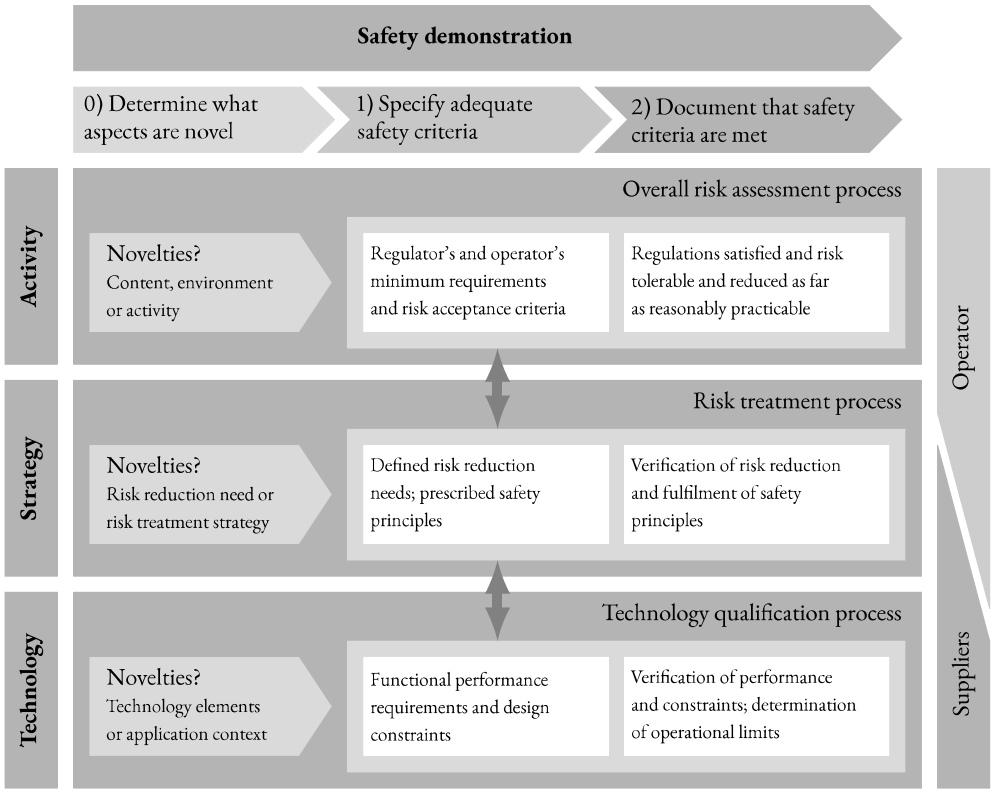

Figure 2.2: Overview of the safety demonstration process, consisting of three subprocesses with different perspectives and stakeholder involvement (van der Meulen and Myhrvold, 2022)

tured reasoning, that adequate safety criteria are specified and met. From this definitions it follows that the safety demonstration process essentially consists of two main steps:

1. Specify adequate safety criteria
2. Provide arguments and evidence that the specified safety criteria are met

However, when demonstrating the safety of novel solutions, it is primarily the implications of the novel aspects that are of concern. An additional preliminary step is therefore included to determine what aspects of the solution are novel:

0. Determine what aspects of the solution are novel

The three steps shown in Figure 2.1 are not performed as one linear process. In practice, the safety demonstration process consists of several subprocesses that may be conducted by different parties with different focus and perspectives.

To standardize the safety demonstration process, three main perspectives are defined van der Meulen and Myhrvold (2022):

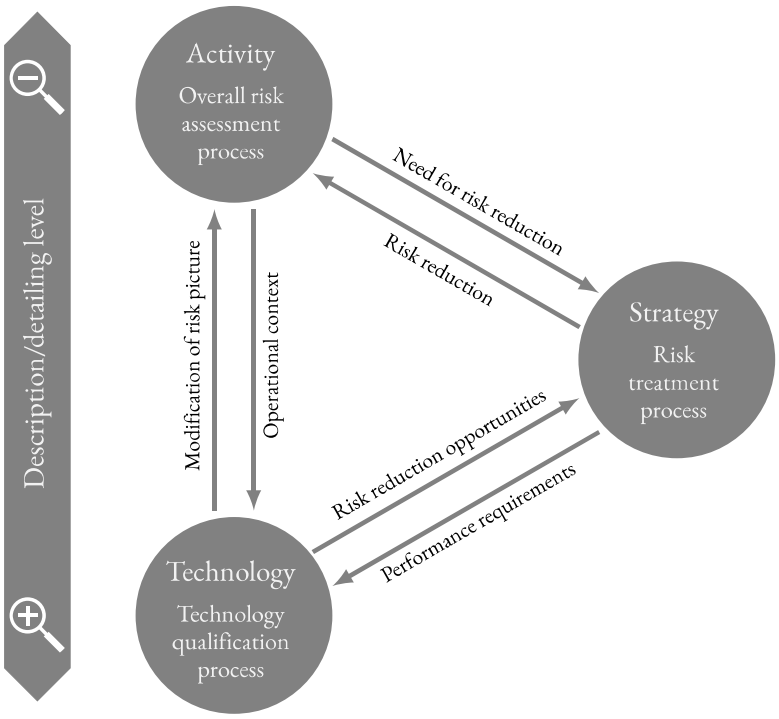

Figure 2.3: Illustration of three interlinked perspectives on safety demonstration (van der Meulen and Myhrvold, 2022)

- *Activity perspective (as part of risk assessment) - Is the activity safe?* Demonstrate that the overall risk of the activity (i.e., the consequences of the activity, with associated uncertainty) is tolerable and reduced as far as reasonably practicable. If the risk is not tolerable, determine the need for risk reduction.

- *Strategy perspective (as part of risk treatment) - Is the strategy safe?* Demonstrate that the chosen barrier strategy (including technical, operational, and organizational barrier elements) is able to provide sufficient risk reduction to ensure the activity is safe, and that the design and operation of systems are robust and prudent. The latter involves the strategy complying with specific requirements in the regulations, or that special conditions exist on which alternative strategies may be granted.

- *Technology perspective (as part of technology qualification) - Is the technology safe?* Demonstrate that the technology elements used fulfil the technical performance requirements and operational constraints needed for the activity and strategy to be safe.

2.3 The interlinked perspectives on safety demonstration

Figure 2.3 illustrates the three interlinked perspectives on safety demonstration.

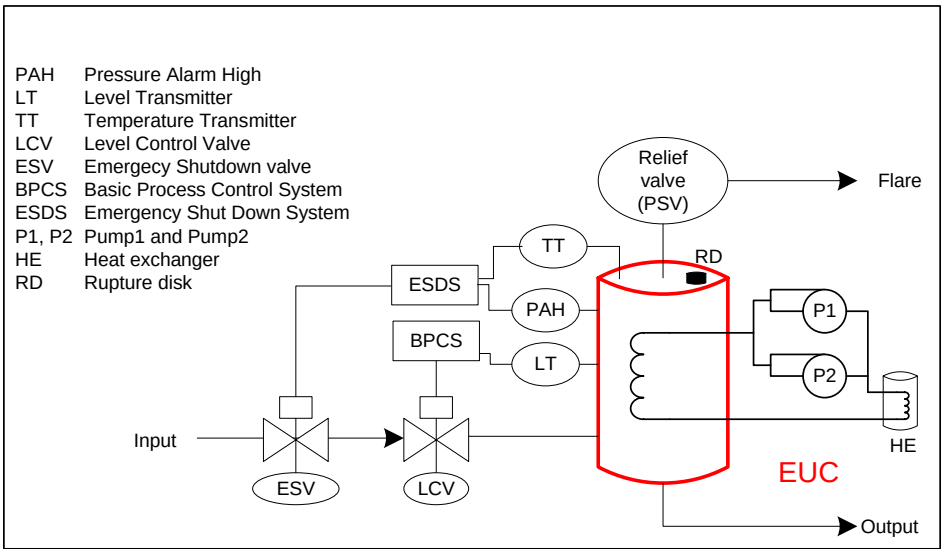

Figure 2.4: Equipment under control - Process system

2.3.1 Activity perspective - assessing the overall risk of the activity

As an example consider the process system in Figure 2.4. The equipment under control (EUC) should be taken care of. Without the emergency shutdown system, the relief valve and the rupture disk the risk is considered far to high.

The consequence dimension covers at least:

- Harm to people and potential loss of lives
- Potential for major accidents with multiple fatalities
- Material damages
- Environmental damages, e.g., release of hydrocarbons

A comprehensive discussion on risk acceptance criteria and other safety evaluation criteria is beyond the scope of this presentation. However, an example on the activity level is related to major accident risk. For example in the LNG-storage facility project at Risavika outside Stavanger the use of risk acceptance criteria was an issue. Figure 2.5 shows the risk acceptance criteria used. Note that these criteria also illustrates “historical” risks to help lay people related risk to something which might be more familiar. In the Risavika case it was a significant debate regarding the risk levels shown in Figure 2.5. It is also a debate regarding the value of using risk acceptance criteria which we will not pursue here.

Now, if the use of risk acceptance criteria is used, how will this translate to something tangible to be used at the strategic and technological level? First of all, having Figure 2.4 in mind,

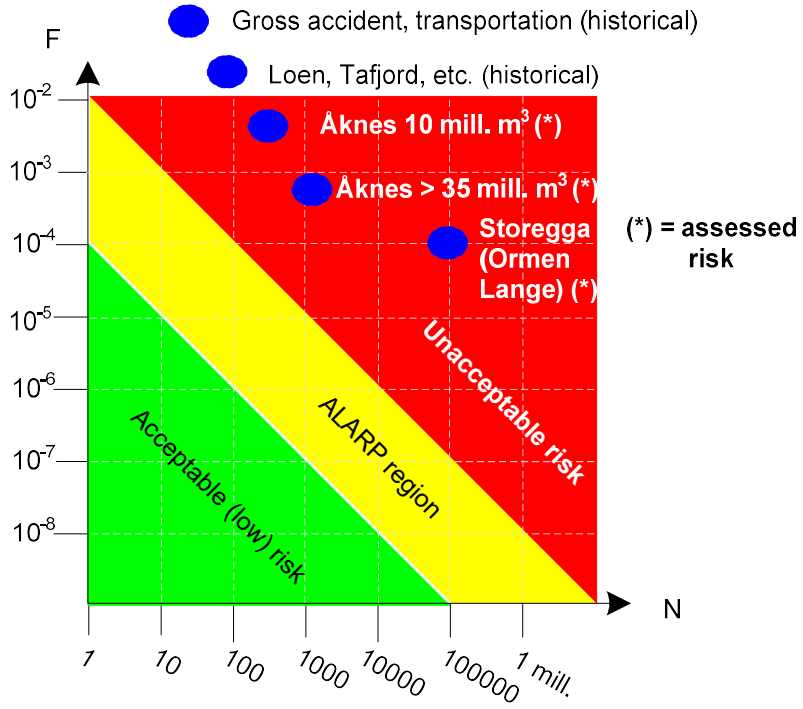

Figure 2.5: Risk acceptance and ALARP with major accident perspective

without the emergency shutdown system, the relief valve and the rupture disk the risk is considered far to high compared to the levels shown in Figure 2.5. Thus there is a need for risk reduction.

Since there are many EUCs' that contributes to the total risk, the risk associated with a particular risk should be much lower than the total risk covered in Figure 2.5. To simplify we consider a rupture of the vessel in Figure 2.4 to have a potential of 10 or more fatalities and there are in total some 10 EUCs representing major accident potential, the total frequency of rupture should not exceed one per 10 000 years.

Figure 2.6 depict the risk reduction perspective in IEC (2010). The idea is to identify risk reducing measures to bring the risk to a tolerable risk level which with our argument would be less than one (significant) vessel rupture per 10 000 years. For our vessel in Figure 2.4 the risk reduction is achieved by the following safety barriers:

- The safety instrumented system (E/E/PE), i.e., the ESD-system
- The pure mechanical PSV
- The static rupture disc

In relation to the overall safety demonstration process depict in Figure 2.3 this relates to the interaction between the *activity* and *strategy* subprocesses for safety demonstration. The

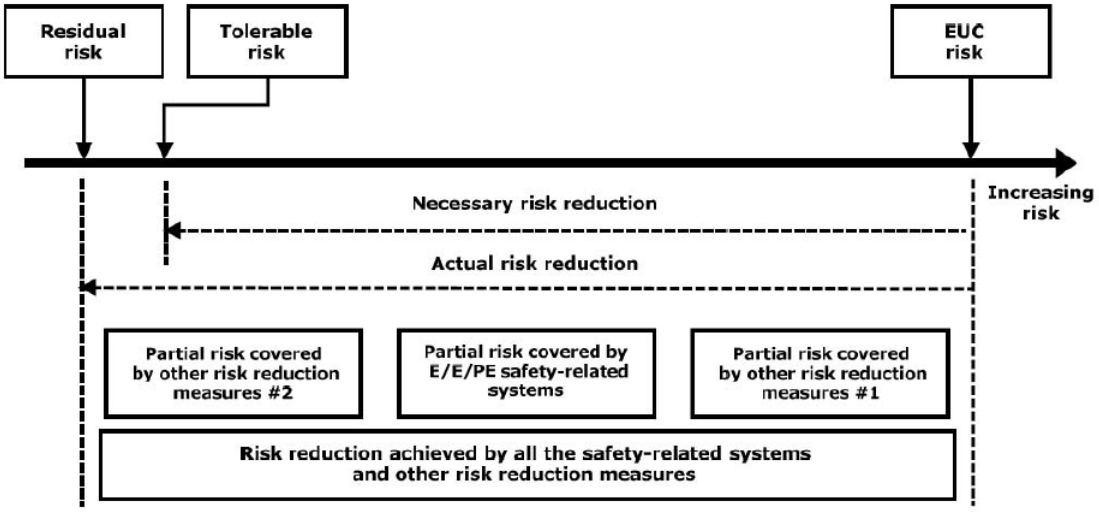

Figure 2.6: IEC 61508 - Risk reduction perspective (IEC, 2010)

identification of safety barriers is only one element of the strategy subprocess. Figure 2.7 depicts all the steps for each of the subprocesses in Figure 2.3. In addition to the safety barrier strategy also barrier performance must be defined, for example SIL-requirements, closing times for the safety valve etc.

This brings us to the technology qualification process, and in particular related to our example steps required to verify that the SIL-requirements are met, i.e., verifying the performance of the safety barriers.

2.4 Types of novelties and paths to safety demonstration

The path to safety demonstration consist of three basic steps as discussed earlier:

0. Determine what aspects of the solution are novel
1. Specify adequate safety criteria
2. Provide arguments and evidence that the specified safety criteria are met

The path to safety demonstration varies from project to project, depending on the type of novelties involved. However, a set of main paths is proposed by van der Meulen and Myhrvold (2022) and described below. The combination of operational environment and strategy is considered to constitute the application context of technologies.

Path A–Proven/qualified technology in a familiar application context There is nothing new about either the technologies used or any particular challenges related to the environment or strategies employed in the project. Compliance with existing recommended standards can be used to demonstrate safety.

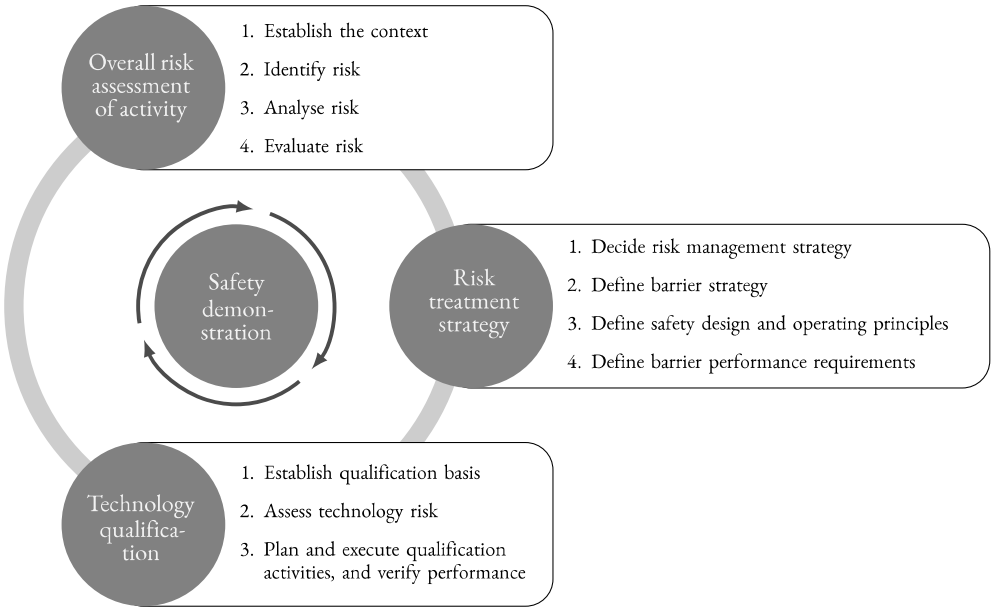

Figure 2.7: Overview of the safety demonstration process steps (van der Meulen and Myhrvold, 2022)

- *Example:* A train operator (Vy, Goahead, SJ) invests in new trains of similar types that have been available on the marked in recent years.

Path B–Proven/qualified technology in a new application context Technologies may need to be re-qualified against new requirements that follow from the new application context.

- In path B.1, the changes to requirements mainly result from the unconventional strategy.
  - *Example:* A train operator invests in new trains of similar types that have been available on the marked in recent years but train control (Bane NOR) is changed from using track circuit to axel counters, but where the environment is familiar and the technologies to be used for the barrier elements are proven/qualified, i.e., used in other countries.
- In path B.2, the changes to requirements may result both from an altered context at the activity level and from the unconventional risk treatment strategy.
  - *Example:* Same as B.1 but the operator company splits into two units, one focusing on maintenance and operation of the trains, and the other unit focusing on route planning and extending the services, i.e., bus service in addition to train operations.

Path C–Novel technology in familiar application context The novel technologies must be qualified against existing requirements corresponding to a standard environment and risk treatment

strategy.

- *Example:* A train operator invests in new trains on a non-electric line where the new trains have hydrogen as energy source. No changes neither to the operator nor the infrastructure manager.

Path D–Novel technology in new application context The novel technology must be qualified against the new requirements that follow from the new context.

- In path D.1, the changes to requirements mainly result from the unconventional strategy.
  - *Example:* The barrier strategy to control hydrogen is changed compared to how these barriers are implemented in trucks, i.e., 2oo2 to 1oo1 wrt pressure relief.
- In path D.2, the changes to requirements may result both from an altered context at the activity level and from the unconventional risk treatment strategy.
  - *Example:* Same ad D1, but in addition the Railway Inspectorate (SJT) also the Directorate for Civil Protection (DSB) are involved in the approval and safety follow-up processes.

In the Safety 4.0 textbook other examples are provided.

2.5 Types of failures

2.6 Overall risk assessment – Risk treatment – Technology qualification

These topics are discussed extensively in sections 2.2 – 2.4 in the Safety 4.0 textbook (van der Meulen and Myhrvold, 2022). With respect to barriers, typically Safety-Critical Systems as discussed in the SIS-book Rausand (2014) sections 2.3.2 (Barrier strategies) and 2.3.4 (Defining barrier performance requirements) are of interest.

With respect to technology qualification the Safety 4.0 textbook builds on the DNV recommended practice for technology qualification (DNV-RP-A203, 2021).

2.7 Types of failure

When it comes to risk treatment and technology qualification it is important to understand the various types of failures. The Safety 4.0 textbook proposes to classify failures as *random failures*, *systematic failures*, *systemic failures*, *operator failures*, *failures due to environmental causes* and *failures due to deliberate actions*.

2.7.1 Random failures

Random failures are linked to the physical properties of components. The term 'random' is used because the exact moment a specific component will fail is unknown and does not imply that the failure happens arbitrarily. Typical failure rates for a large group of the same component can be predicted through analysis of statistics from field experience, and this makes it possible to perform quantitative risk analysis (QRA) that takes into account the probability of failure for the different components in a system.

Note that the failure rate depends on operational loads and strategies, maintenance etc. In data handbooks such as OREDA and PDS handbook, the failure rate estimates are average values over all type of loads and strategies. Therefore it is not straight forward to find plant specific failure rates. This to be discussed later.

2.7.2 Systematic failures

Systematic-failure events are the consequence of inadequate work processes and may be introduced at all stages in the system life cycle. Some examples are incomplete risk analysis; inadequate development of barrier strategies; incomplete requirement specifications; weaknesses in software design; programming errors; quality problems in hardware production; and inadequate planning of maintenance. It is difficult to quantify the probability of systematic failure events as they typically will be present in a system from day one, or introduced through modification, but will be hidden until specific circumstances occur. In IEC (2010) it is stated that systematic failures shall not be quantified. However, in the PDS method systematic failures are quantified in terms of the test-independent failure probability, p_{TIF}.

2.7.3 Systemic failures

A systemic failure is an event which occurs even if no individual component in the system has failed. This may be caused, for example, by overlooked dependencies among the technical, operational, human and organizational elements of systems; specifications based on inadequate understanding of physical processes; or, by unexpected inputs for which no specific response has been specified. Increasing system complexity may also increase the risk of systemic failures, and this is particularly relevant for systems containing software functions. It can be related to intricate dependencies and feed-back mechanisms among system components leading to nonlinear and unpredictable system behaviour. In the famous book Normal Accident (1984) by Charles Perrow systemic failures are extensively discussed in light of e.g., the Three Mile Island accident.

2.7.4 Operator failures

Operator failures occur when an operator fails to perform appropriate actions, or performs an inappropriate action. Operator failures are often denoted human errors and usually dealt with under the heading human reliability analysis (HRA).

2.7.5 Failures due to environmental causes

Failures due to environmental causes are caused by physical processes having negative influence on the system. Some examples are: lightning strike, water ingress, fire, electrostatic discharge from personnel, sensors covered by salt, and electromagnetic interference affecting communications.

2.7.6 Failures due to deliberate actions

Failures due to deliberate actions may be caused, for example, by hacking, computer viruses, physical sabotage, deliberate jamming of radio signals, or spoofing (false signals).

2.8 Failure rate estimation

Vatn (2006) proposes a method for estimating failure rates for safety-critical systems. In the paper the following aspects are discussed:

- How to establish an initial failure rate when historical data for the actual system is not available? This involves a detail analysis of relevant generic data sources and how analysis of failure causes and mechanisms can be used to extrapolate the failure rate estimate for the given operational context.
- How to collect data in the operational phase and judge individual failure reports wrt failure cause and mitigating actions carried out in order to establish the statistical basis for updating the failure rate
- Bayesian statistical analysis to use for the update of the failure rate

See also Håbrekke et al. (2020) who extend the approach by Vatn (2006) also to include situation for equipment groups.

Bibliography

DNV-RP-A203 (2021). Technology qualification. recommended practice. Standard, DNV, Høvik, NO.

Håbrekke, S., Lundteigen, M. A., and Hauge, S. (2020). New method for updating failure rates and proof test intervals of equipment groups within safety instrumented systems. In *Proceedings of the 30th European Safety and Reliability Conference(ESREL)*, pages 797–804.

IEC (2010). Functional safety of electrical/electronic/programmable electronic safety-related systems (e/e/pe, or e/e/pes). Standard, International Electrotechnical Commision, Geneva.

Rausand, M. (2014). *Reliability of Safety-Critical Systems*. Wiley Series in Probability and Statistics. Wiley.

van der Meulen, M. and Myhrvold, T. (2022). *Demonstrating safety of software-dependant systems*. DNV AS.

Vatn, J. (2006). Procedures for updating test intervals based on experience data. *ESREDA seminar on Reliabilityi of Safety-Critical system*.