Improving risk and safety decision-making in the high-risk energy sector industries through cross-industry learning opportunities

by

Surbhi Bansal

Thesis submitted in fulfilment of
the requirements for the degree of
PHILOSOPHIAE DOCTOR
(PhD)

Faculty of Science and Technology
Department of Safety, Economics and Planning
2021

University of Stavanger
NO-4036 Stavanger
NORWAY
www.uis.no

ISBN:978-82-8439-011-6
ISSN:1890-1387
PhD: Thesis UiS No. 593

Preface

This thesis is submitted for partial fulfilment of the requirements for the degree of Philosophiae Doctor (PhD) at the University of Stavanger, Faculty of Science and Technology, Norway. The research presented was performed in the period from February 2018 to March 2021. This PhD thesis is funded by the Norwegian Ministry of Education and Research (Kunnskapsdepartementet). The financial support is gratefully acknowledged.

This PhD project has been a study of decision-making under uncertainty, with a focus on safety and risk management. The main goal was to contribute to new knowledge towards improving elements of decision-making under uncertainty for energy sector industries. The contributions made in this thesis are a result of an approximately three-year long process of research and discovery.

I would like to express my heartfelt gratitude to my supervisor and co-author, Associate Professor Jon Tømmerås Selvik. I want to thank you for your supervision, which consistently combined expertise, patience, positivity, and a consistent follow-up, throughout. Your support has been critical in helping me achieve my goals and complete this thesis successfully, especially towards the end.

Next, I would also like to thank Professor Eirik Bjorheim Abrahamsen. Your insights, support and encouragement were always available when required. Your constant push to take a broader perspective of things will always stay with me. I am grateful to you for making my journey enjoyable with your light-spirited attitude.

I received this exceptional opportunity to work under your guidance at a point in my career when it was much needed. I thank you both for this!

I present my heartfelt gratitude to another co-author, Nejm Saadallah at NORCE. Thank you for your support in work that has personal importance.

I also want to acknowledge the help received by Linda March (for her unparalleled skills at proof-reading) and all staff and coordinators at UiS. Your help was always quickly available whenever requested. To all my fellow colleagues in the C-gang, I am grateful for your interesting conversations that were stressbusters for me.

Finally, to my husband Abhishek and my parents, you have been a constant source of encouragement, right since the day I applied for the PhD. Thank you for entertaining my highs, as well as lows. This would not have been possible without your love, support, and sacrifice. I am truly blessed to have you in my life!

Surbhi Bansal

Stavanger, April 2021

Summary

The overall objective of this thesis is to contribute new knowledge to the applied area of decision-making under uncertainty. More specifically, this research relates to improving risk and safety decision-making in the high-risk energy sector industries, by exploring cross-industry learning opportunities.

The prevalence of common risk and safety issues faced by the energy sector industries presents opportunities for cross-industry knowledge transfer. Cross-industry learning requires fewer resources, to learn by experience. The commonality of accident causes, and high-level lessons make it a practical way to proceed towards achieving more effective safety management at the industrial level. In fact, these industries have adopted methods, principles, and tools from each other in the past. There is a trend towards developing more general holistic concepts for capturing the needs of assessing and managing decision problems in their industrial context. While the traditional safety and risk analysis tools and principles are still relevant for these industries, major learning opportunities that can prove useful for decision-support should not be left unexplored.

Observing and understanding the decision-making processes followed by industries in the energy sector (oil & gas, nuclear and chemical processing industries) reveals commonalities. All of them broadly involve decision problem identification and alternative description, decision-analysis, decision-makers' review and making the decision. A key feature of this process is the role of the stakeholder's inputs, i.e., his goals, criteria and preferences. Since they heavily influence all elements in the decision-making process, they need to be actively accounted for when evaluating the usefulness of an improvement opportunity.

Based on the evident commonality in risks and decision-making processes, several sources of learning opportunities for improving the

decision-making process emerge, some of which have already been adopted. However, identifying other potential improvement opportunities, assessing them and finding a suitable criterion to evaluate them is not so straightforward. Currently, there is a gap in this area that this thesis strives to fill.

During the research, several sources of learning became evident. Some of these learnings have been inspired by major accidents in the past. The accident mechanisms can reveal characteristics and conditions shared by other high-risk industries. Information on energy-related accident risks, such as containment barrier weaknesses, reliability of human and organisational barriers, weaknesses in safety performance systems, failure of monitoring and diagnosability systems, etc., can provide useful information to stakeholders with a critical decision-making role in the industry. A second source is the use of well-established assessment techniques for capturing risks in a difficult area (e.g., human performance). It can readily provide inspiration for adoption by other industries lacking it. Other areas to look for such learning opportunities are evident through the scientific works of the risk & safety community, tracking the developments in upcoming modern tools/techniques, etc.

The thesis makes use of logical frameworks, rationality criteria, scientific reasonings and case studies, to evaluate the actual usefulness of a learning opportunity, when needed by that industry. Certain cases of incompatibility, and alignment issues with the adopting industry, were discovered. Papers I & II demonstrate this. Here, the Return on Investment (ROI) tool and the Human Reliability Assessment (HRA) method were adopted from the financial and nuclear industries, respectively, for the purpose of decision-support within the oil and gas domain. In particular, the need to align the human reliability assessment method with the risk perspective of the adopting industry has been evident. Both the papers recommend ways to overcome their corresponding limitations in capturing the industry-specific uncertainties and risks.

Contributions have also been made regarding improving the analysis criteria for accepting/rejecting the adoption of safety principles that may prove useful for the decision-making process (Paper IV). This paper takes on a decision-maker's broader perspective on the usefulness of a safety indicator within a portfolio of other indicators, not just on a stand-alone basis. To this end, improving the existing SMART acronym ('specific', 'measurable & manageable', 'relevant' and 'timely') to STAR, for evaluating the usefulness of indicators measuring safety performance has been suggested. This will assist in evaluating and selecting safety indicators that provide the decision-makers with a more useful risk trend.

The thesis also found a case where a learning opportunity with limited usefulness was discovered. The Texas City accident highlighted limitations of the defence-in-depth safety principle. It was suggested that this principle should be used with another safety principle that advocates having superior monitoring and diagnosis (Paper V). While such a recommendation may seem to be useful immediately, on evaluation, such a recommendation did not seem to add significant value for decision-makers in the nuclear industry. The learning has also been in the direction of employing caution and determining a concrete rationale before adopting multiple safety principles. It is possible that just improving the implementation of existing safety principles may be sufficient. This means that, while there is a growing consciousness among energy sector industries regarding looking towards cross-industry learning opportunities, they also need to carefully consider gaps within their own systems and processes first.

Lastly (Paper III), the thesis inspires us to not limit the learning horizon to only across the industries but also look into the emerging techniques for more complex decision-making needs in a high-risk operating environment, where wrong decisions can prove to be costly in the long run. For this, a novel decision-support technique was developed, since offshore and other industries were just beginning to explore the

possibilities of modern data-based techniques for improving decision-support.

List of papers

I. Bansal, S., Selvik, J.T. and Abrahamsen, E.B. (2018) Return on Investment (ROI) for evaluating safety measures. Review and discussion. *The Business Review, Cambridge.* ISSN 1553-5827. Volume 26.

II. Bansal, S., Selvik, J.T. and Abrahamsen, E.B. (2019) Alignment of the Petro-HRA method with the risk perspectives in the Norwegian oil and gas industry. Proceedings of the 29th *European Safety and Reliability Conference* (ESREL). ISBN 978-981-11-2724-3.

III. Bansal, S., Saadallah, N., Selvik, J.T. and Abrahamsen, E.B. (2020) Development of a bivariate machine-learning approach for decision-support in offshore drilling operations. Proceedings of the 30th *European Safety and Reliability Conference* (ESREL2020), *15th Probabilistic Safety Assessment and Management Conference, (PSAM15) 15.* ISBN 978-981-14-8593-0

IV. Selvik, J.T., Bansal, S. and Abrahamsen, E.B. (2021) On the use of criteria based on the SMART acronym to assess quality of performance indicators for safety management in process industries. *Journal of Loss Prevention in the Process Industries*. ISSN 0950-4230, https://doi.org/10.1016/j.jlp.2021.104392

V. Bansal, S., Selvik, J.T. Investigating the implementation of safety diagnosability principle to support defense-in-depth in the nuclear industry: A Fukushima Daiichi accident case study. *Journal of Engineering Failure Analysis*. ISSN 1350-6307, https://doi.org/10.1016/j.engfailanal.2021.105315

Part I

Table of Contents

Preface i

Summary iii

List of papers vii

Part I viii

1 Introduction 1

1.1 Background 1

1.2 Objectives 4

1.3 Research approach 5

1.4 Thesis structure 6

2 Theoretical foundation 8

2.1 The concept of risk and uncertainty 8

2.1.1 Risk 8

2.1.2 Uncertainty 11

2.2 Decision-making under uncertainty 14

2.3 Model for decision-making process 18

2.3.1 Overview of decision-making in the energy sector industries 21

3 Research areas and problems 27

3.1 Improving the adoption and development of risk assessment approaches for decision-making 30

3.1.1 Economic evaluation of safety 31

3.1.2 Inter-industry adoption of risk assessment 33

3.1.3 Novel approach for supporting real-time decision-making 35

3.2 Capturing safety performance to safeguard decision-makers' preferences 37

3.3 Improving the use of safety principles for decision-makers' judgement 40

3.4 Discussion 43

4 Future work 48

References 50

Part II 65

Paper I 66

Paper II 79
Paper III 92
Paper IV 105
Paper V 138

1 Introduction

1.1 Background

The energy sector industries continuously face risk and safety management challenges. They are categorised as high-risk or safety-critical industries. High-risk industries have work processes that imply considerable risk for people and the environment, regarding large potential for either major accidents or smaller-scale incidents and occupational accidents (Grote, 2012). The procurement, production, distribution, and use of energy in its various forms have the potential to cause adverse effects on people and the environment (Rasmussen, 1981). The safety performance of energy systems, such as oil and gas, nuclear and chemical, can have important environmental, economic and social implications (Burgherr & Hirschberg, 2014). This has been realised through lessons learnt from devastating major accidents worldwide (e.g., the Three Mile Island nuclear incident, Piper Alpha accident, Chernobyl nuclear accident, Texas City Refinery explosion, etc.). Every accident has generated new learnings such as stricter industry-specific regulatory requirements, new safety principles, an emphasis on the human factor, etc. Inevitably, the energy sector industries have been emphasising the importance of accurately assessing accident risks since the 1980s (Burgherr & Hirschberg, 2014; Fritzsche, 1989; Inhaber, 2004; Rasmussen, 1981). Even today, these industries must continue to improve their risk analysis methodologies to account for uncertainties. In this respect, the industry-specific knowledge developed and accumulated over time, whether in the form of safety principles, barrier management or risk assessment tools, etc., presents a unique learning opportunity for these industries.

Safety-critical industries have been learning from each other for centuries, and this cooperation has extended nationally and internationally (Berg et al., 2015). The word ‘learning’ (n.d.) refers to

*knowledge or skill acquired by study, instruction, or experienc*e (Merriam-Webster). Cross-industry learning in the risk and safety context refers to acquiring knowledge about tools, methods or principles across industrial boundaries. Just as tools and technologies are adaptable from one industry to another with only little modification (Pearl, 2007), cross-industry learning in the high-risk energy sector, by either adopting or adjusting existing tools and methodologies, can play an important role in the risk and safety domain. For example, the basis of probabilistic risk assessment (PRA) originated in the aerospace industry in 1960 and later was extensively used by the nuclear industry for reactor safety study (Bedford & Cooke, 2001; Khan et al., 2015). Human reliability analysis (HRA), deriving its methods and guidance from the nuclear industry, is being adapted for applications in the oil and gas industry (Boring, 2015). The principle of defence-in-depth originally emerged as a military defence strategy (Parker, 1996); it is now a fundamental safety principle in the nuclear, oil and gas, cybersecurity, etc. fields.

In the safety management context, at different times, different industries were considered ahead of everybody else, e.g., the nuclear industry through the 1940-80s, the chemical industry for process safety management through the 1980s, etc. (Grote, 2012; Gu, 2018; Amyotte et al., 2007). These industries face ideas and challenges that are of a generic nature, i.e., common across sectors (Rosness et al., 2004). Consequently, those at the forefront of developing tools and methods for risk management implicitly expediated the adoption of some of these developments in other industries. Overall, the transfer of knowledge oscillated between “one size fits all” and “reinventing the wheel” (Amalberti et al., 2005; Grote, 2012; Hudson, 2003).

Adopting or adjusting novel risk assessment methods, tools, safety management principles, performance measures, etc. that have a strong basis of application in another industry can have several benefits. Apart from motivating continuous improvement in the risk assessment area, it can generate new insights for the adoptive industry. Learning new safety

management principles can present a way to challenge the validity of outdated assumptions. Adopting tools from an advanced industry can stimulate creative and modern solutions to support decision-making under uncertainty. For the adopting industry, there is the potential to discover best performance measurement practices from other high-performing industries and bridge its own safety gaps.

Cross-industry knowledge transfer also presents a good way to learn by employing fewer resources. This is due to the commonality across high-risk sectors in terms of both accident root causes and high-level cross-industry lessons, and it is important to consider those that can be distilled and learned (Gabor, 2020). Indeed, different types of applications need different sets of methods, procedures and models, but there is no reason why these areas should have completely different perspectives on how to think when approaching risk and uncertainty, since the basic problem is the same: to reflect our knowledge and lack of knowledge about the world (Aven, 2010b). The study of recent developments (see Aven, 2012b) shows a trend towards developing more general holistic concepts for capturing the needs of assessing and managing decision problems in these scientific environments. While the traditional safety and risk analysis techniques are still largely relevant to the modern world, it is important that all major learning opportunities are utilised to their full potential, regardless of the originating sector (Gabor, 2020).

While, on one hand, cross-industry learning should be encouraged, to improve safety performance, on the other hand, industries need to be conscious of potential implementation issues that may arise from direct adoption in an entirely different working domain. It has been seen that approaches developed in one industry are often advertised as being generalisable to other industries, without much empirical evidence and lacking systematic research on the inter-industry applicability of different safety and risk management methods (see Grote, 2012). Effective cross-industry learning in the energy sector will be decided by the differences and similarities among the industries. Some of the

significant attributes to be considered are: (1) the kind of safety to be managed, (2) the approach to managing uncertainty and (3) the regulatory regime (external vs self-regulatory) (Grote, 2012). To summarise, existing risk and safety knowledge across energy sector industries should be harvested for superior risk management. There is a need for cross-fertilisation across the boundaries separating different industries, disciplines and research traditions, to be able to deal with the increasing complexity of the threats and hazards to the functioning of society (Almklov, 2018), but these learning opportunities need to be carefully analysed for their usefulness, rationality and appropriateness to the adopting industry.

1.2 Objectives

The overall objective of this thesis is to explore the following research area, to make new contributions to the decision-making under uncertainty domain:

- Improving elements of the decision-making process under uncertainty for the high-risk industries in the energy sector, by evaluating the benefits and limitations associated with the cross-industry learning opportunities for the adopting industry.

To approach this research problem, the following two steps are followed:

1. Identify candidates where cross-industry learning opportunities can be adopted or have already been adopted. Also identify opportunities for adopting novel approaches.
2. Determine the benefits and limitations associated with all these opportunities regarding their ability to improve the elements of the decision-making process in that industry's context.

1.3 Research approach

Research refers to the contributions made to the existing stock of knowledge, using a systematic method of study, observation, comparison and experimentation (Kothari, 2004). Kothari (2004) organises the different research types into the following categorisations: descriptive vs analytical, applied vs fundamental, quantitative vs qualitative and conceptual vs empirical. Given the interdisciplinary scope of the risk field, the thesis is a combination of several research types, depending on the nature of the problem. The entire work (Papers I-V) falls into the applied category, since it is concerned with addressing practical decision-making problems faced by different industries and high-risk organisations. The work is also analytical in nature, given the use of available information (e.g., risk assessment tools, principles, the nature of the industry, the regulatory environment, etc.) to make critical evaluations and recommendations based on it. The analytical element is also visible through the use of a case study of past major accidents, operational examples and hypothetical scenarios to strengthen the scientific outcome. Some of the work (Papers I and III) including the reinterpretation of an existing concept and the development of a new technique, respectively, is associated with conceptual research. Lastly, a part of the work is proportionately fundamental in nature (i.e., Papers I, IV, V), where the generalisation of existing theories, fundamental safety principles and evaluation criteria is evaluated through systematic and logical reasonings. Overall, the thesis is a combination of conceptual, applied, analytical, fundamental and conceptual research, in which the mode of investigating the research problems is largely qualitative.

This PhD thesis follows the criteria for scientific quality laid out by the Norwegian Research Council (NRC) (2000). The research has been conducted to the best of the author's ability to emulate the criteria of originality, solidity and relevance (as per NRC, 2000).

Originality relates to the contribution of new knowledge to the existing academic literature. The work in this thesis maintains originality by developing new methods, improving existing concepts, and applying existing knowledge to new problem areas. Solidity refers to good substantiation of statements and conclusions in the research work. The use of good references, scientific methods, consistency of logic among statements, a critical mindset, and rigorous evaluation of results has been employed to satisfy solidity criteria. The research in this work is relevant both academically and society-wise. Its findings are applicable to different high-risk industries with a focus on filling the gap in cross-industry learning opportunities in the existing risk literature.

This is a two-part thesis that follows Day and Gastel's (2006) European PhD model. The first part of the thesis is an introduction to the research area. Through a review of the existing literature in this field, it narrows down what, why and how certain problems have been tackled. It also presents ideas for future work in this area. The second part comprises all the published scientific papers that constitute this thesis. The papers are a result of idea generation from conference participation, literature review, supervisory guidance and co-author discussions, coding and programming, general awareness about trends, rational and creative thinking for problem solving, peer review from journals, conference feedback, introspection, self-examination, proof-reading and continuous toiling at research.

1.4 Thesis structure

The thesis comprises two parts. Part I lays out the foundational principles and basic concepts that form the basis of the research. Part II contains five scientific research papers, whose work is associated with assessing appropriate cross-industry learning opportunities, as well as modern approaches for decision-support, in the risk and safety context.

Among the two parts of the thesis, Part I is organised into four sections. Section 2 lays out the foundational concepts relevant to the thesis's objectives. Section 3 discusses the research problems, while Section 4 presents a direction for future work. Lastly, Part II contains the scientific articles that are the main contribution of this thesis.

2 Theoretical foundation

Risk has a very long past but a very short history (Rosa, 1998). Risk assessment and management were only established as a scientific field a few decades ago, yet the principles and methods developed to conceptualise, assess and manage risk still to a large extent represent the foundation of this field today (Aven, 2016). This section presents the relevant theoretical concepts, industrial background and foundational principles for identifying the cross-industry learning opportunities. These will be useful for building an understanding of the research problems and their solutions in Section 3.

2.1 The concept of risk and uncertainty

2.1.1 Risk

Risk arises whenever some potential source of damage or loss to a target exists, for example, people, industrial assets, or environment (Aven et al., 2013). Aven (2014) outlines the various development paths of the risk concept and elaborates on how some of the risk definitions can be traced back to different environments – economics, engineering, social science, etc. The perspectives on risk vary among industries and disciplines. The risk literature consists of risk definitions that can be divided into two categories (Aven & Renn, 2009; Aven, 2014):

(1) Risk expressed by the means of probabilities and consequences (e.g., expected loss)
(2) Risk expressed through event/consequences and uncertainties.

In (1), probability and/or expected values form the basis of defining risk. For example, Kaplan and Garrick's (1981) definition of risk as a 'set of triplets' has been dominant in the nuclear industry over recent decades (Aven, 2014). It defines risk as a triplet of (s_i, p_i, c_i), where s_i is the ith

scenario, p_i is the probability of the scenario, and c_i is the consequence of the ith scenario, $i = 1, 2, \ldots N$ (Kaplan & Garrick, 1981). The chemical and processing industry, on similar lines, considers risk as the measurement of process safety which is a combination of "how bad an accident would be?" and "how often could it happen?", quantitatively expressed as a function of probability or frequency and their consequences (Centre for Chemical Process Safety, 2007). De Moivre defines the risk of losing any sum to be the product of the sum adventured multiplied by the probability of loss, i.e., expected loss (De Moivre, 1711). The expected value-based risk perspective is used by the insurance and finance industry, economists, portfolio managers, etc.

In category (2), the uncertainty aspect defines the risk concept. For example, risk is a two-dimensional combination of the consequences of an activity, C, and associated uncertainty, U, or (C, U) in short (Aven, 2007; Aven 2010c). In recent decades, the risk community has seen a shift from probability-based to uncertainty-based definitions, such as risk as the uncertainty about and the severity of the consequences of an activity or event with respect to something that humans value (IRGC, 2017), risk as the effect of uncertainties on objectives (International Organisation for Standardisation, 2018), etc. The Norwegian Petroleum Safety Authority (PSAN) also updated its risk concept from a more traditional probability/expected value-based definition to an uncertainty-based definition in 2015, to prevent oversimplification and loss of important information (PSAN, 2016). There have been extensive discussions discouraging the use of probability to define risk in favour of uncertainty, since it allows for a more pragmatic view that is appropriate for a general context and facilitates all types of uncertainty representations (including probability) (Aven, 2011; Aven, 2014; Flage et al., 2014; Askeland et al., 2017; Hillson & Hulett, 2004). In this thesis, the risk is understood as being based on the consequences and associated uncertainties, i.e., (C, U).

Safety is a disciplinary term (Selvik & Signoret, 2017) that can be seen as an attribute of risk. It refers to the absence of unwanted outcomes such as incidents or accidents, hence, a reference to a condition of being safe (Hollnagel, 2014). International organisations define it as the freedom from risk which is not tolerable (ISO/IEC, 2014) or being without unacceptable risk (SRA, 2018). It is commonly considered an antonym of risk, wherein a high safety level means a low risk level and vice versa (Aven, 2020; SRA, 2018). Just as no industrial activity has zero risk, there can be no absolute safety (Verma et al., 2010), irrespective of whether safety is interpreted as an acceptable level, state or absence of unwanted outcomes. Therefore, the term 'safety' is always associated with risk in such a manner that risk has to be assessed and eliminated and safety has to be assured (Chandrasekaran, 2016), where risk is the key concept and safety is defined based on it (Aven, 2020). Safety is also paraphrased as a dynamic non-event with the understanding that nothing untoward happening or the freedom from unacceptable risk is the non-event (Hollnagel, 2014). Within the energy sector, the type of risks facing the industry may determine the way safety is interpreted. For instance, in the nuclear context, 'safety' refers to the safety of nuclear installations, radiation safety, the safety of radioactive waste management and safety in the transport of radioactive material; it does not include non-radiation-related aspects of safety (International Atomic Energy Agency, 2006).

Möller (2012) presents a perspective on the potential complications that may arise when considering risk as an antonym of safety. He suggests that the safety concept must be distinguished from its absolute interpretation, i.e., the sense of there being no harm or an absence of accidents (as in Miller, 1988; Tench, 1985), and also from the notion of acceptable risk. This is because it is often reasonable to claim that even though an activity is not safe, its risk can be acceptable (Möller, 2012). While, from a broader perspective, these connotations may not have a significant impact, the importance of using the concepts of risk and

safety with their proper interpretations is acknowledged. In this thesis, the use of safety or its extended use, for example as 'safety system', is based on the understanding of acceptable risk, i.e., a system with the function of protecting from dangerous failures that can increase the risk to an unacceptable level (see Selvik & Signoret, 2017).

2.1.2 Uncertainty

Defining risk is distinct from describing risk. While risk defined as the pair (C, U) makes it easier to understand risk, it is not a sufficient means to evaluate or communicate it. A risk description serves to describe or measure risk (qualitatively or quantitatively) for performing risk assessments in decisions-making problems. Presenting a complete description of the risks of a future activity requires capturing several dimensions (see Aven (2014) for details). The most significant among these is the measure for uncertainty U (probability or others) that is based on some background knowledge K. The knowledge dimension enters the scene when we try to describe or measure risk, since the judgements about the specified consequences and uncertainties are always more or less conditional on the analyst's knowledge (Flage et al., 2014; Aven & Zio, 2018). The importance of reporting this knowledge for decision-making is discussed in the next section.

Risk analysts need to understand and predict technological systems' behaviour for their safety performance. This requires assessing the limited available information about the system, along with their own knowledge and expertise, which might be imperfect. This gives rise to a component of ignorance, known as uncertainty (Ayyub & Klir, 2006). The notion of risk differs from uncertainty, as it is associated with a rational decision based on the possibly limited knowledge of the states of the world, while uncertainty refers to the difficulty in describing, deciding or assessing the consequences of possible decisions (Emblemsvåg, 2012). It refers to the lack of knowledge about unknown quantities, i.e., about the occurrence of events (A) and what the

consequences or outcomes (C) will be if an activity is carried out or a system is put into operation (Flage and Aven, 2009).

In engineering risk assessments, a distinction is commonly made between two types of uncertainties – epistemic and aleatory uncertainty (e.g., Apostolakis, 1990; Helton & Burmaster, 1996; Aven et al., 2013). Epistemic uncertainty (or subjective uncertainty) refers to the lack of knowledge about a phenomenon and the latter (or stochastic uncertainty) refers to the uncertainty about a parameter due to variation in population (see Helton & Burmaster, 1996; Aven et al., 2013).

The representation and characterisation of uncertainties in risk assessment is a serious matter, as uncertainties feature strongly in the decision-making process involved in the risk management (Aven et al., 2013). The question of how to define and measure the different types of uncertainties is particularly critical in the analysis of high-consequence phenomena (e.g., failures of nuclear reactors) because of public sensitivity to the magnitude of the potential outcomes (Paté-Cornell, 1996). Engineering risk analysis, such as in the nuclear power industry, generally relies on the models of probabilistic risk analysis (PRA) to assess the risk of operations of nuclear power plants (United States Nuclear Regulatory Commission, 1975). It is also common to see probability and expected values being used to represent and describe uncertainty in high-risk energy sector industries. While alternate probability-based methods to describe uncertainty, such as probability bound analysis (Ferson & Ginzburg, 1996), imprecise probability (Walley, 1991), evidence theory (Dempster, 1967; Shafer, 1976), etc. exist, these have not been broadly accepted by the risk assessment community, since researchers are sceptical about their use for the representation and treatment of uncertainty in risk assessment for decision-making (see Aven et al., 2013; North, 2010).

Additionally, in a decision-making setting, the stakeholders may not be satisfied with a pure probability-based approach to risk analysis, as it can

involve subjective judgement made by a group of analysts; probabilities hiding the uncertainties of the assumptions they are based on; possibly weak or strong knowledge supporting probabilities; poor knowledge on the high-consequence risk problem, etc. (see Zio & Pedroni, 2012; Aven et al., 2013; Aven, 2014). In a risk/safety assessment context, whether the uncertainty of a quantity, model, phenomenon or future event needs to be represented, two main concerns should be balanced, as per Aven et al. (2013):

1. Knowledge supporting the representation should correspond to documented and approved evidence; the methods and models used to treat this model should neither add nor ignore information
2. Analysts' judgement ('degree of belief') should be clearly reflected ('judgements').

Both these concerns reflect the need to express the strength of background knowledge (K), along with the uncertainty representation. This is in support of the risk as (C,U) approach, holding uncertainty, not probability, as the main component of risk (Aven, 2008b) and regarding probability purely as an epistemically based expression of uncertainty (Flage & Aven, 2009). Since the probabilities (P| K) are conditioned on the background knowledge of the assessor, the decision-makers should be informed about how strongly this K supports the probability assignment. The thesis focuses on the need to highlight the strength of background knowledge when communicating the risk assessment, in order to capture the uncertainty holistically. The following approach can be used to assess this strength of knowledge as weak, in a probability-based analysis, if one or more of these conditions are true (refer to Askeland et al., 2017; Flage & Aven, 2009):

a) The assumptions made represent strong simplifications.
b) Data/information is non-existent or highly unreliable/irrelevant.
c) There is strong disagreement among experts.

d) The phenomenon involved is poorly understood, models are non-existent or known/believed to give poor predictions.

The knowledge is considered strong when the opposite of all the relevant conditions above are met. All the cases falling in between have medium strength of knowledge. Such a labelling of the knowledge aspect assists the decision-makers in judging the weight that needs to be placed on the probability and/or expected values in order to take risk-informed decisions.

2.2 Decision-making under uncertainty

Decision-making under uncertainty is closely related to risk/safety management. Both (1) support the decision-maker to take decisions that optimally balance risks and values, (2) involve similar process steps and (3) mostly use the same techniques for their analysis step. On one hand, appropriate decision-making is an important task in risk management implementation; on the other hand, a risk management process in decision-making is an important step for better decision-making (Lu et al., 2012). To this extent, this thesis considers them to be essentially the same. Thus, the understanding for risk and safety management presented in the following in this section stands to also contribute to understanding about decision-making.

Given the universal nature of risk being embedded in all industrial activities, the concept of risk is addressed in all fields, whether finance, engineering, health, transportation, security or supply chain management (Althaus, 2005; Aven, 2016). Its management involves all those activities that handle risk, such as prevention, mitigation, adaptation or sharing (Aven, 2014; Society of Risk Analysis, 2015). International Organisation for Standardization (ISO) defines the risk management process as the systematic and structured use of policies, procedures and practices for the task of establishing the context, and assessing, treating, communicating, consulting, monitoring and reviewing risk (ISO, 2018).

The aim is to strike the right balance between exploring opportunities on one hand and avoiding losses, accidents, and disasters on the other (Aven, 2008a; Society of Risk Analysis, 2015; PSAN, 2018). Establishing the context, risk assessment and risk treatment are the main steps of risk management (Aven et al., 2013). The critical risk assessment step should provide insights that support decision-making, such as choosing between alternatives, the implementation of risk-reducing measures, etc., so that the decision-making in the face of uncertainties is risk-informed, not risk-based (Apostolakis, 2004; Aven, 2010a).

Safety management has no clear-cut definition (Antonsen et al., 2012); safety is managed differently according to the industrial context. The earlier classical works define safety management as a systematic control of worker performance, machine performance, and the physical environment (Heinrich et al., 1980), while advocating that the basic safety management principles should be rooted in the general management of the organisation (Petersen, 1978; Antonsen et al., 2012). The idea that the management is responsible for the organisational safety is firmly rooted in the safety principles, even today. In broader terms, it is a process or a series of activities to realise certain safety functions (Li & Guldenmund, 2018). The safety managers evaluate the system's safety performance by producing frequency estimates of specific hazards, with a focus on the risk acceptance criteria (Abrahamsen et al., 2010).

The Oganisation for Economic Cooperation and Development (OECD) defines it as the organisational measures that seek to identify, assess and control *risks* in order to guarantee nuclear, personnel and environmental safety (OECD/NEA, 2006). The task of securing a good safety performance from a complex nuclear power plant system is challenging, because safety is an outcome of several organisational, individual, technical and environmental factors which also interact with each other (e.g., Rasmussen, 1997; Reason, 1995; Reiman & Oedewald, 2007; Kettunen et al., 2007). Similarly, CCPS (2007) guidelines define 'Process Safety Management' focused on the prevention of,

preparedness for, mitigation of, response to, or restoration from catastrophic releases of chemicals or energy from a process associated with a facility.

The disciplines of safety management and risk management are often thought to be independent, when they are essentially the same discipline working towards comparable goals of loss prevention or mitigation (Sloan, 2007). Depending on the industrial sector and professional field, one is preferred over the other or one is subsumed under the other, but it can be concluded that they basically mean the same thing (e.g., Harms-Ringdahl, 2004; Grote, 2012). For instance, safety management uses the same concepts, principles and techniques used in other areas of management (DNV, 2012), such as the ISO's risk management standards can be applied to both industrial safety and project risk management (Kontogiannis et al., 2017).

While safety management differs from risk management, in the sense that it does not concern itself with the cost and financing aspect (see Sloan, 2007; Kettunen et al., 2007), institutions and researchers emphasise the need to use risk management to demonstrate the business value of safety to organisations (Kontogiannis et al., 2017). For instance, safety managers can use risk assessments along with cost-benefit analysis (or return on investment analysis) to assess a safety barrier's economic efficiency. This symbolises the distinction between the two fields being blurred for practical purposes. In this thesis, safety management is included under the broad umbrella of risk management for providing an adequate basis for managing risk.

Within the area of safety, different perspectives exist on how to provide an adequate basis for managing risk (Engemann & Abrahamsen, 2020). There are several reasons for this. Firstly, multiple risk perspectives have developed among diverse disciplines. For instance, the scientific community views risk as a measurable objective reality, applying principles, assessments and knowledge to uncover facts and manage the

risks (Althaus, 2005; Breyer, 1993; Aven, 2010a). The industrial safety experts favour placing a stronger weight on uncertainties, often by nominating caution as the ruling principle when making safety decisions (Abrahamsen & Abrahamsen, 2015; Aven, 2014). On the other hand, the economists treat risk as a decisional phenomenon or a means to secure wealth/avoid losses, applying their knowledge of decision-making principles and postulates to understand the unknowns (Althaus, 2005). Decision-making guided by economic principles proposes the use of expected values, with the intention to optimise a criterion (Abrahamsen et al., 2004; Edwards, 1954; Simon, 1959). Aven (2014) points out that risk management is a balancing act between the pursuit of benefits from an activity/business that may increase risk over time. Then, logically, both the perspectives of risk management – economic as well as safety – need to be considered for decision-making. This is because generalising that every decision-making problem adopts a strict and extreme perspective (economic or safety) can be misleading. For example, Abrahamsen et al. (2018a) illustrate that, even if the cost-benefit (cost-effectiveness) analysis concludes upon no investments, high levels of uncertainty, among many other issues, can justify investments in a safety measure.

Clearly, different perspectives on risk can lead to different ways of assessing risk, which in turn may affect the risk management and decision-making in particular (Aven, 2009). Cost-benefit approaches and socio-economic profitability have been the guiding principles for the implementation of safety measures, through systematic analysis of the costs and benefits of various policy approaches (Adler, 2011). Other widely used risk management principles available to guide the decision-making from a safety management perspective are ALARP (As Low As Reasonably Practicable), the cautionary/precautionary principle, etc. (Baybutt, 2014; Abrahamsen et al., 2018a).

Choosing an unsuitable assessment tool or guiding principle can present significant challenges for managers of risk. These challenges can arise

either from not characterising uncertainties appropriately (e.g., basing decisions only on probability-based assessment such as expected values, not capturing the strength of knowledge, inappropriately capturing safety performance, etc.) or applying unsuitable principles (e.g., adopting redundant safety principles during decision review) to manage them. Both themes have been addressed in this thesis.

2.3 Model for decision-making process

Energy system studies include a wide range of issues from short-term (e.g., real-time, hourly, daily and weekly operating decisions) to long-term horizons (e.g., planning or policy making), where the decision-making chain is fed by input parameters which are usually subject to uncertainties (Soroudi & Amraee, 2013). The varying decision-making problems present require the decision-makers to maximise the value generated by their decision, while simultaneously satisfying business objectives related to attributes such as safety, cost minimisation, regulatory compliance, reputation, etc. Such decision problems are challenging because of the uncertainty associated with the input parameters and decision outcomes. Uncertainty results from incomplete and imprecise knowledge (epistemic uncertainty) or the intrinsic randomness of the world (aleatory uncertainty). Examples of such problems are deciding on the trade-off between risks and benefits, selecting the optimal alternative, real-time decision-making situations, safety barrier management, etc.

As per Aven & Kørte (2003), there are two schools of thought that can be adopted to reach a good decision:

(1) Decision-making as a modelling exercise of outcomes and alternatives to maximise/minimise certain criteria

(2) Decision-making as a process of risk and decision analysis, managerial judgement and review and, finally, a decision.

Approach (1) focuses on providing the decisions directly. This is unsuitable in the risk and safety management context, as the decision-makers require a full overview of the decision analysis to make the best decisions. For example, consider a decision-making model that analyses safety measures for minimising the investment cost. The model may not select the alternative with the highest safety performance, if minimising the cost is the model's only criterion. This means that approach (1) can strongly impose the decision-model's outcome while hiding uncertainties about the underlying assumptions, strength of background knowledge, model, input data, etc. Ignoring these aspects can have implications in the form of poor decisions with unintended consequences. Therefore, this thesis follows approach (2), as it supports the decision-makers through a well-structured process, rather than producing mechanical decisions. Approach (2) can be visualised using the figure below.

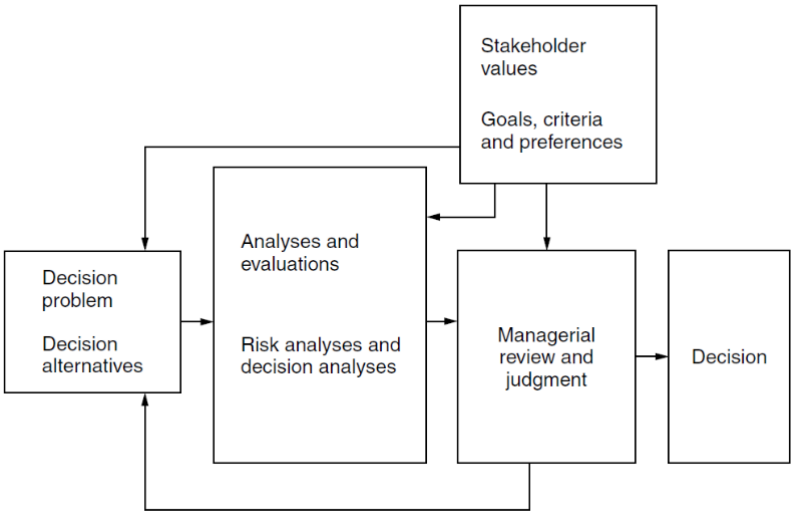

Figure 1 Basic structure of decision-making process (Aven, 2012a; Aven & Kørte, 2003)

This decision-making process depicted in Figure 1 is particularly useful for decision-making problems characterised by uncertainties (Aven,

2012a; Aven & Kørte, 2003). The process begins with defining the decision problem and listing the decision alternatives. The next step analyses and evaluates these decision alternatives, by selecting and applying the relevant analysis methods. The result of the analysis is presented for their managerial review and judgement to the decision-maker(s), who make the final decision. The stakeholder values, business goals, criteria and preferences are crucial at every step of this process. They are inputs at every stage and vary according to the industrial environment. From the large toolbox of risk analysis methods available, the risk analyst needs to select a method that will appropriately account for these inputs. This implies that different methodologies/approaches will capture and assess uncertainty to varying degrees affecting the decision.

Ensuring that the decision-making process supports value creation and protection, by adequately managing risks, making decisions and improving performance, requires following some fundamental decision-making principles (refer to ISO (2018), for the list of principles). These principles stress placing a greater emphasis on the iterative nature of risk management, wherein updated knowledge and analysis should be used to revise processes, actions and controls (Institute of Risk Management, 2018).

These principles list the characteristics that are the cornerstone of an efficient decision-making process. For example, for a fast-paced offshore drilling setting, these principles guide the decision-making process to be dynamic, utilise the best available information and strive to continuously improve the safety of drilling operations. Consequently, the drilling operator will receive risk-informed decision support to make optimal decisions, balancing production and risks. Similarly, all the other key principles considered together emphasise the need to identify uncertainties and account for their effects on the decision outcomes. Usually, the nature of the risks and operating environment can determine the weight for the relevant principles. Improving the implementation of

these principles in the industrial decision-making process is a natural implicit goal of this research.

2.3.1 Overview of decision-making in the energy sector industries

Risk management involves decision-making in situations involving high risks and large uncertainties, and such decision-making is difficult, as it is hard to predict what would be the consequences (outcomes) of the decisions (Aven & Vinnem, 2007). Most of the decisions to be made by energy sector decision-makers are subject to a significant level of uncertainty (Conejo et al., 2010). This uncertainty has often manifested itself as poor decisions that proved costly, in the form of several major disasters within the energy sector. Accidents in the energy sector have been shown to form the second largest group of man-made accidents, after transportation (Hirschberg et al., 1998; Burgherr & Hirschberg, 2008). The energy sector industries long ago acknowledged the deficiency in the decision-making process (or its elements), as the factor responsible for poor risk management, and continuously strive to improve. This section presents an overview of the decision-making needs and risk management practices in the energy sector industries that are under focus in this thesis.

Oil & gas industry

The oil and gas industry has come a long way from being reactive to having a proactive safety (or risk) management system (Hudson, 2003). This is a result of the decades of evolution of risk management practices, from risk-based to risk-informed decision-making (as in Figure 1). Yet decision-making in the safety context of the oil and gas industry remains a complex exercise, since it involves large numbers of variables, multiple disciplinary concerns and uncertainties arising from incomplete or unavailable information. In addition to this, the oil and gas industry is also capital oriented, with investment decisions being crucial (Deore,

2012). Most of the investment decisions require some form of trade-off between production and safety. The negative consequences of an inappropriate trade-off decision can be magnified by the declining oil and gas prices. So, the oil and gas companies emphasise the need to manage uncertainty through better decision-making. These challenging decision problems are handled by employing several tools and principles. The use of probabilistic decision-analysis methods, such as QRA (Quantitative Risk Assessment), HRA (Human Reliability Assessment), cost-benefit analysis using expected values, decision-trees, etc., is common.

Converting non-monetary criteria into monetary equivalents, cost-benefit analysis is a common way of addressing the challenge of analysing the decision alternatives against diverse criteria (Lev, 2007). However, this cost-benefit approach has several limitations regarding its use for assessing the investment benefits of safety measures. Despite the use of probabilistic methods (such as cost-benefit analysis) for uncertainty quantification having increased significantly over the years, it may not have translated into improved decision-making for the oil and gas industry (Bickel & Bratvold, 2008). Probability is often used to quantify the extent of knowledge about uncertainties (such as depositional environment, volume in place, production rate and oil price), but it merely captures the extent of our degree of belief in the possible outcomes of these events (Bratvold & Begg, 2009). Thus, probability may not sufficiently convey the uncertainties associated with risky operational decisions such as drilling or the reliability of an operator's performance.

To capture the risk of human performance, the oil and gas industry adopted learnings from the nuclear industry. HRA methods originated as a probabilistic risk assessment method for understanding and quantifying the risks of a serious accident at a nuclear power plant and are today also developed for or adapted to other industries, such as oil and gas, chemical, etc. (Massaiu & Paltrinieri, 2016). The Norwegian oil and gas

industry adapted the nuclear industry's HRA method into its Petro-HRA guideline in 2017. The new guideline is a definite step towards drawing the focus of the offshore industry towards capturing the likelihood of human error. However, the guidelines need to be adapted better to the oil and gas industry's risk perspective and operating environment.

Nuclear industry

Decision-making at a nuclear power plant consists of a wide spectrum of situations, from fast short-term decisions for operational transients to planning preventive maintenance and repair strategies in the long term (Vaurio, 1998). IAEA (2011) promotes the use of an integrated risk-informed decision-making (IRDM) process to ensure that decisions affecting nuclear safety are optimised without unduly limiting the conduct of operation of the nuclear power plant. IRDM has the basic decision-making process (Figure 1) at its core (see p. 9, IAEA, 2011).

It involves defining the problem, listing the solution alternatives, accounting for applicable requirements (i.e., mandatory, deterministic, probabilistic, organisational, etc.) and weighing the alternatives against these requirements, after which the decision-maker proceeds with his decision and its implementation (IAEA, 2011; Zio & Pedroni, 2012). The decision outcome's performance is finally monitored as feedback to the decision-making process. This decision-making process ensures that the outcome satisfies the safety principles of defence-in-depth (DID), safety margin maintenance, regulatory compliance, etc. The key to the risk-informed decision-making approach is that it is complementary to the defence-in-depth philosophy (Verma et al., 2011), which has established a strong position as a deterministic safety principle, not only in the nuclear industry but also in other high-risk industries (chemical & processing, petroleum, aviation, etc.). The defence-in-depth principle advocates redundancy, diversification and conservatism in system design (Niehaus & Szikszai, 2001), serving as a guiding principle for decision-making made difficult by the uncertainties involved.

DID has received regular criticisms for its limitations, along with suggestions for improvements by the adopting industries. However, its implementation within the nuclear industry has improved, due to the strict regulatory regime and conservative risk management practices – owing to public sentiment regarding nuclear power – which are largely a result of the lessons learnt from past major nuclear disasters. Suggestions emerging from other industries for the improvement of this principle should be carefully considered in this current light.

Chemical processing industry

The chemical processing industry has, unfortunately, witnessed several major accidents such as the Seveso dioxin release (1976), the Piper Alpha explosion (1976), the Bhopal gas release (1984), the Texas City Refinery explosion (2005), etc. These accidents have led to the development and revisions of several regulations worldwide for process safety (CCPS, 2019) that shape the process safety management process today. Consequently, the process safety strategies that govern the decision-making for loss prevention have evolved from a strict standard-based compliance to a risk-based strategic decision-making approach today (for details, refer to CCPS, 2010). Even then, the safety- and risk-related decisions in the chemical process industry, particularly within the European Union, are heavily subjected to multiple levels of legislation, standardisation and socioeconomic analysis (Kozine et al., 2001). The Texas City Refinery explosion in 2005 is of particular significance, since the Center for Chemical Process Safety (CCPS) proposed the risk-based process safety (RBPS) management approach to update the process safety management framework (CCPS, 2010; Chen, 2016). It is based on the rationale that a thorough understanding of the hazards of a process risk is fundamental to making good risk decisions involving competing alternatives with different risk reduction levels and costs (CCPS, 2019). The decision-making process consists of the following steps: defining the problem, evaluating the baseline risks, identifying alternatives,

screening the alternatives and making the decision (CCPS, 2019; Hammond et al., 1999; CCPS, 1995).

The industry uses conservative safety principles for guiding its risk-based decision-making. The high-hazard chemical industry uses principles such as ALARP, supported by cost-benefit analyses and the grossly disproportionate criterion, for decision-making in safety management, but often without paying the proper attention to the decision frame (e.g., level of uncertainty and knowledge of the chemical phenomena, the use of best available technologies, the potential of major losses due to the release of hazardous materials and other items) (Abrahamsen et al., 2018b). Such decisions can fail to generate the desired value for risk management.

The industrial organisations continue to evaluate their business/process decisions for their safety performance. Monitoring effectiveness and performance becomes an important step for optimisation (Deore, 2012), in which the safety barrier's performance, based on historical accident trends, near-misses and end business objectives, allows the identification of anomalies and continuously improved safety systems. The Seveso Directive III, responsible for controlling major accidents caused by industries dealing with hazardous substances, sets requirements related to performance control and checking the effectiveness of technical and organisational measures (Jovašević-Stojanovic, 2009). This establishes the need to evaluate the effectiveness of safety management within the process industry. The CCPS (2011) also sets out an extensive guideline for the use of process safety indicators, to measure the existing and future safety performance of the safety management system. While there is a whole spectrum of safety indicators in the safety literature, these metrics only convey a limited fraction of the input required for a decision-making process. The decision-makers often need to consider many such indicators and metrics during the review and judgement step, to understand the system's overall safety performance. So, there is a need

to evaluate the quality and relevance of information represented by safety indicators from an individual, as well as a portfolio, perspective.

3 Research areas and problems

Using learnings through cross-industry experience is a convenient resource for propelling continuous improvements in the elements of the decision-making process. For the energy industries, a complex industrial setup presents the common problem of better capturing high uncertainty and the decision-maker's preferences. In this respect, the thesis explores how the current risk assessment, principles and tools can be efficiently improved, by either adopting or adapting the existing knowledge already harnessed by other industries encountering similar challenges. Such adoptions lead to development within the risk domain but also require careful evaluation of the uncertainties associated with that adoptee's context. The improvements should meet the industry-specific decision-support needs. Technicalities may arise from diverging risk perspectives, regulatory environments, decision-maker's individual vs portfolio's perspective, etc. The need for a criterion to systematically evaluate the appropriateness of a particular tool/metric has also been addressed.

Improving an entire decision-making model that is suitable for all the energy sector industries may not be feasible. Instead, making contributions to improving selected elements within the decision-making framework of selected industries has been a target. On a broader level, assessing the suitability of such an improvement will require a focus on:

- How well the individual element captures and analyses its dedicated risk aspect
- Its compatibility with the existing portfolio of assessment tools or safety principles
- Its ability to adhere to the stakeholders' inputs, etc.

Incompatibility disconnect or failing to meet these needs would significantly reduce the usefulness of a learning opportunity for the

decision-maker who needs to consider the overall risk picture for making decisions.

This chapter presents the contributions of the individual papers and the research problems addressed by them under the light of a unifying research theme. The thesis makes contributions addressing the energy sector, in particular the oil and gas, nuclear power and chemical processing industries. The main research problem is to generate new knowledge for:

- Improving the decision-making process under uncertainty for the high-risk industries in the energy sector, by evaluating the benefits and limitations associated with the cross-industry learning opportunities for the adopting/adapting industry.

To approach this research problem, the thesis follows the following two steps:

1. Identify candidates where cross-industry learning opportunities can be adopted or have already been adopted. Also identify opportunities for adopting novel approaches.
2. Determine the benefits and limitations associated with all these opportunities regarding their ability to improve the elements of the decision-making process in that industry's context.

The first step selects the different candidates that have the potential for improving the decision-support process. These are selected by exploring tools, principles and methodologies in the energy sector industries that have either been adopted or have the potential to be adopted from other industries. The opportunity for incorporating modern decision-support techniques has also been explored. The next step assesses these opportunities for suitability to the adopting industry's decision-making context, thereby contributing specific knowledge to improve its risk management.

The main contribution of this thesis consists of five scientific papers. The research papers relate to improving certain elements of the decision-making process. This has been illustrated through Figure 2. While each paper is associated with an individual step in the decision-making process, its contributions result from consideration of the entire decision-making model. This is because the strongly connected stakeholder input (i.e., goals, preferences and criteria) proves a significant factor in determining the suitability of the learning opportunity for the adopting industry. Since this information flows to and through every step, no step can be considered individually for improvement in the decision-making process.

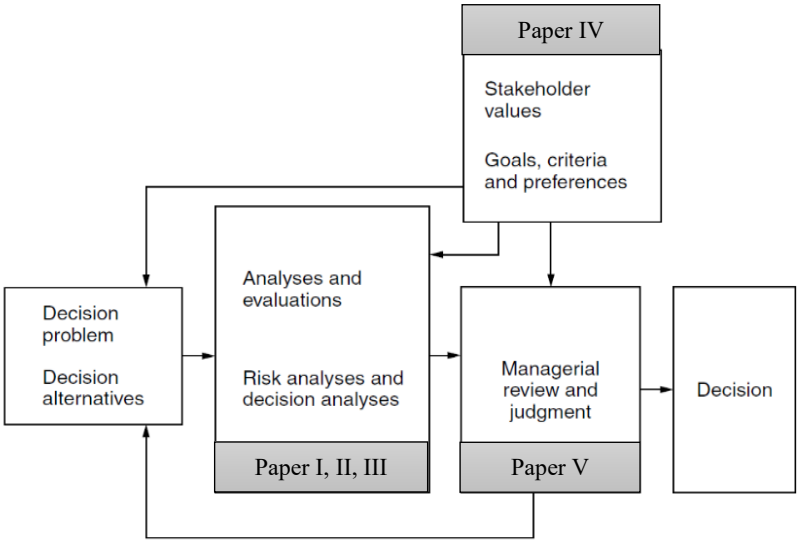

Figure 2 Link between research papers and the elements of decision-making under uncertainty

The thesis aims to contribute towards strengthening the scientific basis of cross-industry learning opportunities for decision-making in the energy sector. Considerable differences in the magnitude, timing and nature of associated risks can be expected among the various energy industries, allowing a degree of choice in the decision-making process, with regard to selecting alternatives, decisions on policies and achieving

safety goals (Burgherr & Hirschberg, 2008). The choice of decision-support approach suitable for that industrial context should be viewed with a focus on its uncertainty-capturing ability. For instance, industrial risk assessment methods, fundamental safety principles and other tools used by decision-makers should highlight the critical information/insights required to make better decisions for protecting business value. Current risk literature lacks investigation on improving the existing risk assessment toolbox by using cross-industry learning opportunities or determining their appropriateness for supporting managerial review and judgement in different industrial contexts of the energy sector. The following sections describe and discuss the contributions of the five papers, addressing the research problem at hand, which is also looked at from a broader decision-making context.

3.1 Improving the adoption and development of risk assessment approaches for decision-making

Decision analysis provides a formal methodology for the systematic examination of a complex and opaque decision situation, the formulation of alternative courses of action, the treatment of information, uncertainty and preferences, and the evaluation of supposedly the "best" alternative or course of action (Huang et al., 1995). In situations with lower or negligible uncertainty associated with the decision's outcome, a rule-based decision-support analysis can suffice. However, Hopkins (2011) reflects that not all decision-making can be procedural (or rule-based), as there will always be situations not covered by the rules or perhaps where quick decisions are needed, which require expertise to assess risks and act appropriately. This is particularly true for decision-making in certain energy sector applications characterised by both aleatory and epistemic uncertainties. Meeting the need for tools/methods for decision-support that are customised for that application, organisation or industry falls in this focus. The following sections discuss opportunities for

evaluating adapted, adopted or newly developed decision-supporting elements for the suitability of their purpose in that industry's context.

3.1.1 Economic evaluation of safety

Paper I: Return on Investment (ROI) for evaluating safety measures. Review and discussion.

Safety management searches for cost-effective solutions and to attain and maintain a safety level that conforms to defined policies, goals and requirements (Abrahamsen et al., 2004). In practice, while safety managers may not often be punished for putting safety first, they will quickly be penalised for not putting profits, market share or prestige first (Paltrinieri & Khan, 2016). This implies that prioritising safety measures with reference to both risk and socio-economic profitability is challenging (Engemann & Abrahamsen, 2020). For this, the most commonly used methodologies supporting safety investment decisions make use of classical approaches derived from the financial context, which are generally aimed at examining types of investments for their benefit to the company (Milazzo et al., 2020). Adopting a methodology from the financial context may not adequately capture the costs, benefits and associated uncertainties of investing in a safety measure. This is the main motivation for Paper I. It reviews and discusses a cost-benefit analysis-based performance measure, called return on investment (ROI), for its ability to prioritise safety measures by accounting for decision uncertainties appropriately. ROI is mathematically expressed as

$$\mathrm{ROI} = \frac{\mathrm{E[X]-E[C]}}{\mathrm{E[C]}}$$

where E[X] and E[C] are the corresponding safety measure's expected benefits and expected costs, respectively. The main contribution of the paper is a discussion on the usefulness of ROI for decision situations related to safety. It is found that ROI provides a rational framework for the evaluation of safety measures based on the link between the use of

expected values and the traditional portfolio theory (see Durbach & Stewart, 2009; Boardman et al., 2017; Zou et al., 2010).

Despite its obvious attractiveness for being simple to calculate and interpretable and for its widespread safety-related applicability, its major limitations emerge from its inability to express certain aspects related to the uncertainty of decision outcomes. Paper I identifies four key limitations that challenge the justification and usefulness of ROI for decision-making in the safety context, which often employs a cautionary mindset. These are as follows:

1. Extreme consequences may lead to poor predictions of a safety measure's ROI.
2. The challenge of transforming all attributes into monetary values.
3. The expected values, on which the ROI-value is based, are conditional on a number of assumptions and presuppositions.
4. The calculation of ROI is subject to corporate procedures, which may affect the value of the portfolio as a whole.

A review of the ROI measure reveals challenges that are significant for the different steps in the overall decision-making process (refer to Figure 1 in Section 2.3) and not just a single decision-analysis step.

To improve the use of ROI as a basis of evaluation for safety measures and to align itself with the cautionary mindset, an extended ROI framework is proposed. This is another main contribution of Paper I, in which the extended ROI framework proposes elements where due consideration is placed on improving the entire decision-making process. The extended ROI captures the net benefits of a safety measure given, conditional on a critical failure occurring. The findings in this paper are relevant for all the high-risk energy sector industries, since the extended ROI framework addresses the common challenge of sensitivity of ROI results to high-consequence, low-probability events. Overall, Paper I highlights that there is a need to see beyond the narrow quantitative and economic rationality of monetarisation of social benefits (Perrow, 1984)

when conducting a decision analysis to consider investments in safety measures. While a simplistic ROI analysis may be rational for a purely finance-driven sector, it can lead to poor decisions for industries prone to unsystematic (or undiversifiable) risks. Paper I also presents an oil and gas industry case study, in order to further strengthen the above argument.

3.1.2 Inter-industry adoption of risk assessment

Paper II: Alignment of the Petro-HRA method with the risk perspectives in the Norwegian oil and gas industry.

Human factors have been shown to play an important role in both the cause and the mitigation of major accidents in the petroleum industry (Norazahar et al., 2014). However, the focus of quantitative risk assessment for this industrial field has traditionally been on technical systems and capabilities (Zhen et al,, 2020). For this reason, the attention dedicated to human and organisational factors in the oil and gas industry is gradually increasing, following several related recommendations suggesting their inclusion (Skogdalen & Vinnem, 2011). In the Norwegian offshore oil and gas industry, human reliability assessment is being applied for the purpose of major accident risk analysis (Gould et al., 2012). To make further advancements in this area, Petro-HRA, i.e., a human reliability assessment (HRA) method for the Norwegian offshore industry, was developed by adapting the nuclear industry's SPAR-H method, in 2017 (Bye et al., 2017). The SPAR-H method was developed for the nuclear industry, utilising safety factors that were to some degree derived from nuclear research (Boring & Blackman, 2007).

The main problem addressed in Paper II is that the Petro-HRA method's output is an input to the offshore industry's quantitative assessment, which adopts a different perspective on risk, compared with that used in the nuclear industry. The challenges to decision-making that can arise from such a cross-industry learning opportunity are the main motivation

for Paper II. In this paper, the differences in the risk perspective of both industries are evaluated. The nuclear industry perceives risk as a combination of consequences and probability, while the petroleum industry sees risk, rather, as a combination of consequences and associated uncertainties. The main contribution of this paper is that, to ensure quality input to the QRA, the Petro-HRA method should be further aligned with the risk perspective of the Norwegian oil and gas industry.

This is based on the realisation that the nuclear industry's focus on probability to represent uncertainty is also reflected in its adapted Petro-HRA version. This raises practical challenges, since probability cannot capture the strength of knowledge aspect and insufficiently addresses the uncertainties hidden in background knowledge supporting it, ultimately influencing the quality of decision-making. Additionally, there are some characteristic differences in the two industries, because of which Paper II argues that there are practical limitations to whether uncertainty can be quantified using probability-based representations to support decision-making. The paper presents a case study of an offshore drilling unit applying Petro-HRA methodology, to discuss the implications for decision-making using a mis-aligned uncertainty treatment. The contribution of this case study is the finding that using a mis-aligned method for the Norwegian oil and gas industry may increase the risk of under-estimating human error probability or overestimating the system's recovery potential. Underestimation of the risk attributed to human performance can manifest itself in the form of poor decisions effectuated by decision-makers wrongly selecting inappropriate measures or strategies during risk treatment. This paper also identifies and attributes the two main sources of uncertainty (i.e., subjective judgement and quality of data) that the different steps of the HRA method may be prone to.

To improve the alignment of the Petro-HRA methodology, the paper makes several suggestions to improve its methodology, in order to meet

the specific needs of the offshore industry, more specifically its uncertainty-focused risk perspective. This is further followed by suggestions to improve its implementation. Together, the recommendations have a focus on improving the practical aspects of the overall decision-making process when employing the HRA methodology.

3.1.3 Novel approach for supporting real-time decision-making

Paper III: Development of a bivariate machine-learning approach for decision-support in offshore drilling operations

The operations involved in the upstream segment of the offshore industry's value chain are quite capital-intensive. Upstream oil and gas includes all the activities related to the exploration and extraction of crude oil and natural gas (e.g., drilling operations) which take place prior to shipping products to the refineries for processing (Shafiee et al., 2019). In the upstream segment, Paper III addresses the challenge faced by the drilling operators, who need to monitor the operation's physical variables continuously for their real-time decision-making needs. The decision-making is challenging, since several known and unknown factors can affect the input-output variable relationship (or the drilling principles) governing the dynamics of the operation. Paper III develops a novel decision-support approach for the offshore drilling operation, addressing the problem of uncertain (or non-deterministic) performance. It uses machine-learning (ML) to provide decision-support, by deriving knowledge from real-time data.

Decision-support systems (DSS) are interactive, flexible and adaptable computer-based systems, specially developed to support the solution of a particular management problem, aimed at improved decision-making (Sprague & Watson, 1993; Morton, 1971; Turban, 1995). Over the years, the use of decision-support systems based on computing methods to

account for uncertainties associated with subjective perception and experience in decision-making in this sector has become popular (refer to Shafiee et al., 2019). ML has resulted in prominent contributions across many industries facing similar challenges; yet its potential has not been fully tapped in the oil and gas industry (Noshi & Schubert, 2018).

Paper III is a step towards contributing to this area, as the use of ML is particularly suitable for this sector. The decision-making problems in the upstream segment are complex in nature, involving uncertainties and risks, requiring qualitative and quantitative decision-making support methods to assist stakeholders in understanding the reservoir characteristics, simulate field operations, make justifiable business decisions, etc. (see Shafiee et al., 2019; Mohaghegh & Khazaeni, 2011). Several applications of machine learning tools for drilling decision-making can be found in the literature (refer to Bello et al., 2016).

The drilling operation presents a process control problem. It requires sequential decision-making by the decision-maker, who is constantly interacting with an uncertain operating environment in real time. The ML approach developed in Paper III employs a reinforcement learning technique. This technique helps the decision-maker (autonomous agent, in this case) to learn to take optimal actions to approach their performance target, by learning through past rewards and punishments received for their actions. The agent's reward/punishment is evaluated based on whether their action results in a system state that is desirable or undesirable (refer to Mitchell, 1997). The agent remembers these learnings and uses them as a decision-making policy to guide their future actions, with a goal to maximise (or minimise) the reward (or penalty). The key contributions of this paper are that the developed approach ensures that the decision-support is dynamic in nature, constantly checks for changes in system-dynamics, by taking into account real-time data, and forecasts a series of input variable values that can assist the operator in achieving a desirable output drilling regime.

The paper demonstrates the application of the developed machine-learning approach, by testing it on a fluid circulation operation in a simulated offshore drilling scenario. It is concluded that the proposed approach satisfactorily predicts the input variable (flow rate of drilling fluid) to achieve a target output value (drilling well's pressure) in the simulated system. The decision-makers are also advised to complement this method's results with a careful consideration of the underlying assumptions and simplifications involved. The novel approach developed using a modern ML technique for the offshore drilling application can also be extended to support the real-time decision-making problems of complex operations in other energy sector industries.

Overall, Paper III is a step in the direction of acknowledging the energy sector's growing need for adaptive and dynamic decision-support tools for critical operations that were traditionally controlled based only on the operator's experience and knowledge. Developing such decision-support tools will require the help of modern ML techniques to some degree. This is because the concept of dynamicity has gone beyond time dependence and online monitoring; it now encompasses progressive calibration and the refinement of nonlinear repetitive processes, as well as reacting and adapting to changes and new information flows (Paltrinieri & Khan, 2016).

3.2 Capturing safety performance to safeguard decision-makers' preferences

Paper IV: On the use of criteria based on the SMART acronym to assess quality of performance indicators for safety management in process industries

Despite significant research to make the process industry safer, several catastrophic accidents have taken place in the last few decades (Knegtering & Pasman, 2013; Chen, 2016). Safety management is

particularly important in major hazard establishments (Abrahamsen et al., 2018a), in which are processed and stored large quantities of dangerous substances, whose release can give rise to severe accidental scenarios, such as fires, explosions and the dispersal of toxic substances (Palazzi et al., 2017). Chen (2016) identifies that, in the high-risk industries, there has been an increasing interest in knowledge management, which, particularly for process safety management, should be used to prevent chemical accidents and guarantee process safety. Among the knowledge used to support the safety decision-making, De Rademaeker et al. (2014) identify process safety performance indicators as a powerful pool of data on the functioning of an organisation.

While also used in other energy sector industries, the process industry places special importance on the use of knowledge from the safety indicators for assessing a barrier's safety performance. Collecting and monitoring the appropriate safety indicators that derive risk-relevant knowledge from the available data can provide a realistic and accurate risk trend to assist an organisation's decision-makers.

Given the critical consequence that can result from an inappropriately managed barrier system (refer to Section 2.3.1), the safety indicators should demonstrate adequate quality. The information achieved through the indicators should be able to help in identifying whether barrier- or safety-related actions are needed. The challenge lies in determining the extent to which quality information is being provided by the indicator. How can the appropriateness/quality of knowledge provided by the indicator, individually as well as within a portfolio, for the decision-making purpose, be assessed? Motivated by this, Paper IV examines the use of SMART criteria for assessing the quality of decision-making information provided by the safety indicators, in the process industry's context.

Selvik et al. (2020) discuss the use of the SMART criteria in a general business context, suggesting that the 'M' of 'measureability' be

swapped, instead, for an assessment of 'manageability', when using indicators for business goals. However, Paper IV finds that this is insufficient for the barrier safety context, as there are other important quality aspects that need to be considered. This paper's main contribution is towards improving the framework for performing assessments of safety indicators.

Knegtering & Pasman (2013) point out that several risk factors can fluctuate in magnitude and time, such that a combination of these factors may cause the actual safety level at a certain location to display dynamic behaviour. This affects the type of indicators required to monitor the safety level or barrier performance. CCPS (2009) provides a comprehensive list of hundreds of process safety metrics/indicators that it is practically infeasible to track and control. These indicators require a form of aggregation to a few significant ones, based on their relative importance (Hassan & Khan, 2012). The chosen safety indicators fall into a collection or portfolio. Despite several challenges associated with having a large Safety Performance Indicator (SPI) portfolio (see Parida & Chattopadhyay, 2007), there is no limit to the number of indicators selected for a portfolio, as long as the combined SPIs contribute valuable information for decision-makers.

The paper presents a discussion on the systematic process of constructing a portfolio of indicators with a main focus on combining individual SPIs' information. The goal of the portfolio is that this combined knowledge should prove useful for decision-making purposes. The discussion on the use of SMART criteria to evaluate an indicator's appropriateness within a portfolio results in a main finding. This is in the form of proposing a modified STAR criterion, wherein dropping the non-value adding 'M' component has been found to be better for the safety context.

To demonstrate the modified 'STAR' criteria in a refinery scenario, the Texas City Refinery accident (2005) caused by overfilling events is used. The usefulness of the 'dangerous fluid overfilling indicator event' is

evaluated for the quality of information it could have provided for superior barrier management, individually as well as within a portfolio. Consequently, the use of this indicator is challenged, as it is found to be lacking in the 'achievability' STAR criteria. The indicator is also applied to another application, i.e., tank storage scenarios, through the Buncefield oil storage depot's accident in 2005, to provide a broader context for its use. The STAR criteria again produce similar results regarding the usefulness of information provided by this overfilling indicator for managing the safety of tank storage operations.

Overall, Paper IV emphasises that the value derived through an indicator associated with a particular safety concern needs to be considered, based on the frame of the decision problem and the specific system. The decision-making frame is determined by the stakeholder's goals, objectives and preferences associated with the system he/she aims to protect. The STAR criteria capture those aspects that determine the suitability of the indicators, with respect to the stakeholder's input. This implies that, within the process industry, the usefulness of an indicator can vary from application to application, and a mechanical adoption of indicators should be carefully considered. However, the use of the suggested STAR criteria framework, for the suitability of indicators for the decision-maker's purpose, remains applicable to other high-risk industries in the energy sector.

3.3 Improving the use of safety principles for decision-makers' judgement

Paper V: Investigating the implementation of safety diagnosability principle to support defense-in-depth in the nuclear industry: A Fukushima Daiichi accident case study.

Defence-in-depth (DID) is a classical defensive concept, currently applied to a variety of technical fields, including nuclear, oil and gas, and many others (Chierici et al., 2016). This fundamental safety principle

aims to prevent major accidents, by promoting decisions in favour of a conservative design, installing safety systems, and incorporating other additional safety features (USNRC, 2016).

In the nuclear industry's context, the intention of fundamental safety principles is to convey the basis and rationale for safety standards to those at the decision-making levels, concerning the use of nuclear energy, since they may not be specialists in nuclear technology or its safety matters (IAEA, 2006b). DID is one of the ten fundamental safety principles charted by the International Atomic Energy Agency (refer to IAEA, 2006b) and has underlying scientific considerations that provide an objective basis for safety decisions. Despite its widespread application, DID is not without criticism. Accidents such as the Texas City Refinery explosion (2005) showed that major accidents can occur, even in a system designed according to this principle. Saleh et al. (2014), through an analysis of this oil and gas industry-related accident, suggest the oil and gas industry should pair DID with a new principle, called the Safety Diagnosability Principle (SDP). This new principle recommends setting up reliable detection and reporting capabilities for barrier failures and safety-degrading events. It advocates information availability, to support and promote safety-informed decision-making. Saleh et al. (2014) invite other high-risk industries (such as nuclear) already implementing DID to generate additional safety value by the use of the complementary SDP.

While the spreading of safety management across industries suggests that different high-risk industries can learn from each other, there are also limitations for generalising safety management methods within and across industries (Grote, 2012). Paper V questions and evaluates the basis of a similarly generalised learning recommendation suggested for the nuclear industry. This paper investigates SDP's value-adding capability or usefulness for the nuclear industry, by assessing its impact on improving the quality of safety management decisions, particularly during accidental situations.

To assess the value of information obtained through the combined SDP and DID implementation, the SMART criteria framework (refer to Section 3.2 for details) is used again. This is motivated by the work of Sørskår et al. (2019), in which it has been used for a consistent and transparent evaluation of a combination of two safety principles. To support this discussion, the paper takes up the Fukushima Daiichi nuclear disaster (2011), where DID failed to stop the accident's escalation. Paper V uses this case study to evaluate the 'achievability' aspect, i.e., whether it is practically possible to obtain the barrier information with high confidence. The discussion generates insights about uncertainties associated with achieving SDP's informational benefits during stressed scenarios, as was the case with the Fukushima accident. SDP also fails to satisfy the relevancy aspect when paired with DID. The relevancy discussion highlights weakness in the newly proposed principle that focuses heavily on monitoring technical barriers' failures alone. The Fukushima accident is a strong example that shows that extreme accidents are the result of a combination of events, ranging from repetitive technical barrier failure events to common non-technical barriers' (such as human and organisational barriers) failures (Paltrinieri et al., 2012). SDP's lack of guidance on tracking the non-technical barrier failure falls short of providing the complete diagnosis of the system when it is the most indispensable for the decision-makers, i.e., just before or during an accident.

The paper duly recognises that having reliable diagnosability, and thus implementing SDP, can be essential for the high-risk nuclear or other energy sector industries. However, determining its value-adding capability on a standalone basis is beyond the scope of this paper.

The paper's main contribution, i.e., presenting arguments challenging the benefits of adopting two principles instead of just using DID to guide decision-makers, rests on the need for better overall management of DID in the nuclear context. This is because SDP's ability to promote safer operations across this industry, while complementing DID, is limited by

its questionable achievability and relevancy that depend on factors such as its implementation, quality of information, operational practices, overlap with DID's diagnosis requirements and other factors discussed in the paper. The discussion on the retrospective application of SDP and, therefore, its potential to positively affect the operator's awareness in the Fukushima accident, infers a limited ability to significantly alter the outcomes of this incident. In other words, following SDP's guidance to complement the failing DID strategy would have had limited impact on improving the decision-support required for controlling/halting accident propagation. While generalising the conclusions through one incident alone may not be a sound practice, nevertheless, the SDP gaps highlighted are likely to be valid for a wide range of nuclear industry applications.

3.4 Discussion

The previous sections outlined the motivations, specific problem areas and the corresponding scientific contributions of the papers in this thesis. This section discusses their results in a broader context of how to improve the elements of decision-making under uncertainty for the high-risk energy sector industries by using cross-industry learning opportunities, wherever suitable.

The energy sector industries clearly involve significant risks. The sector has witnessed many major accidents in the past. The gaps in the sector's risk management practices produced learnings that were also generally applicable. For example, the chemical and processing industry's Texas City Refinery generated learnings about having a more integrated and comprehensive process safety management system (Broadribb & Flynn, 2009). Such accidents also questioned the industry's capability to apprehend the uncertainty of their business operations. Uncertainty handling has been one of the main concerns of the decision-makers (including governors, engineers, managers and scientists) for many years (Attoh-Okine & Ayyub, 2005). The energy sector has a growing

recognition of the need for better decision-support tools and principles that help them discern uncertainties and decide on treatment, based on the nature of the decision problem.

This PhD focuses on two categories of decision-making problems: firstly, non-operational decision-making (Papers I, II, IV, V), related to high-level decision problems that impact the planning, design, and investment aspects in an organisation (Hopkins, 2011). The second category is operational decision-making (see Paper III), which is about supporting quick operational decision-making by those directly responsible for managing operations (such as operators, operational managers, supervisors, etc.).

The decision-makers involved in investment and planning decisions are likely to be more powerful, have a greater impact on company profit and hence be more resistant to limitations on their decision-making freedom than the operators (Hopkins, 2011). The unrestrained decision-making freedom is accompanied by challenges and responsibilities to preserve business value. They are responsible for regularly making high-stake decisions that can potentially impact them, their colleagues and/or the general public's safety, employing both formal and informal decision-support means (Hayes, 2017). However, the operational decision-making scenario also holds safety challenges, albeit arising from different factors and reasons (Paper III).

Irrespective of the decision problem type, for the energy sector domain, risk management is indispensable for decision-support. Risk assessment uses different methods/tools to identify key contributors to risk and support the decision-making regarding which safety measures to implement (Aven & Krohn, 2014). Given the uncertainties and complexities involved, decision analysis tools/methods that use mechanistic rules to either directly recommend or help in quickly selecting the most suitable decision alternative (e.g., ROI, human error probability, expected net present value, etc.) might be appealing for their

simplicity. However, in selecting such tools/methods, the aspects related to uncertainty capturing, stakeholder preferences, risk perspectives, strength of knowledge, etc. (addressed in detail in Papers I & II) are either not fully accounted for or ill-represented. Instead, these should be carefully considered from the decision-maker's perspective.

Aven (2014) emphasises the managerial review and judgement step in the decision-making model. It is advocated for placing the results of the formal analysis in a broader context and taking its limitations and boundaries into account before a decision is made. This plays an equally important role as that of the decision-analysis itself. The decision-maker will utilise all the knowledge gathered from the assessment and place it within the broader frame of their own risk attitude, values, preferences, business goal, industrial requirements, etc. Additionally, fundamental safety principles may guide their decisions (e.g., defence-in-depth, discussed in Paper V, places requirements for a multiple-barrier safety approach for the energy industries applying it) on the premise of having a scientific basis. Since such principles may have their own limitations, it is common practice to recommend that they should be combined with other principles. Again, this should be carefully evaluated, especially when the finding results from a differently characterised industry. Avoiding a direct or mechanical adoption of learning opportunities, citing safety benefits, has been a common theme in this thesis. Paper IV goes in a related direction of improving the commonly used SMART criteria. While the SMART criteria framework is suitable for evaluating the usefulness of business metrics, it is suggested to improve it for the purpose of delivering the trends of a barrier's safety performance. This encourages a careful examination of safety indicators before adopting them from a different industry or even adopting an application from within the same industry.

There has been a common learning from evaluating the usefulness, appropriateness or implementation of adopted/adapted decision-support elements to the industry in question. It is that irrationally adopted or

missing decision-elements have the potential to bias the decision-maker's judgement about the risks of a business activity. This can push decisions in an undesirable direction. In particular, the thesis finds that,

- Deciding on a safety investment, based purely on a financial performance (ROI) metric, may lead to decisions that are misguided by monetary profitability alone.
- The risk perspective chosen strongly influences the way risk is analysed and may have serious implications for risk management and decision-making (Aven, 2014). A misaligned Petro-HRA methodology's results may fail to direct safety measures towards areas with higher uncertainty of human performance.
- The human limitation in accounting for the uncertainty of complex physical principles that govern operations in real time can increase the likelihood of poorly managed operations, due to a lack of adequate decision-support.
- Using a poor-quality safety indicator may not provide a useful risk trend to the decision-maker. He might receive insufficient decision-support for improving the future barrier performance. Using a specific, achievable, relevant and timely-informing indicator is, thus, critical to enable decisions focused on achieving safety goals.
- Where improving the implementation of the existing safety principle would have been sufficient, implementing an additional safety principle advocating monitoring and diagnosis can misguide the resource allocation or the general focus of the decision-makers.

Evidently, papers in this research project adhere to the theme of the 'suitability', 'value-adding capability' or 'rationality' of such an element in the industry being considered. The recommendation for and against adopting such learnings depends on a systematic evaluation of the learning opportunity's benefits and limitations. This is followed by its

compatibility with the adopting industry's existing practices, working environment, risk perspective, regulatory requirements, etc.

In the course of this PhD research, it was recognised that finding cross-industry learning opportunities was challenging, due to characteristic differences among the industries. Despite these differences, for example between the nuclear and the chemical process sectors, a continuous exchange of knowledge and methods from one to the other has led to huge improvements in the chemical process industry (Paltrinieri et al., 2012; Paltrinieri & Khan, 2016). A similar knowledge transfer between the nuclear and oil and gas industry has led to significant developments in the risk assessment domain.

There are several opportunities that the industries can use to learn from each other. But there is also a need to make the best use of modern techniques (e.g., machine learning) where the emerging data science field can drive development of decision-support solutions. Overall, this PhD thesis encourages work in both these directions, where such adoptions/developments can benefit the overall industrial risk domain, while simultaneously recommending a careful evaluation of the suitability to the adoptee's context.

4 Future work

In order to inspire future research in the scientific risk and safety field, the papers in this PhD thesis identify some areas for future work. The recommendations presented below can be used as general guidance by this thesis's papers in areas requiring a deeper scientific work.

- In Paper I, an extended ROI framework has been suggested, in which the traditional ROI metric can be complemented with a conditional ROI calculation, to highlight the un-averaged safety returns of a safety measure. While the extended ROI framework addresses several key challenges of solely using ROI for safety decision-making, the challenge of transforming attributes to monetary values remains. This challenge relates to the intangible costs and benefits associated with the safety investment. Making scientific contributions in this area will be especially useful for the decision-makers in the risk context.

- Given the problem of lack of data- and knowledge-sharing in the petroleum industry, it is suggested to initiate a joint effort in this area, by building a common database for capturing human-error-related events by the industrial stakeholders (Paper II). This will promote the standardisation of human error data collection and sharing within the petroleum industry. A systematically maintained and reported industrial 'near-misses' accidental database will also be useful for decision-makers to employ more targeted measures to reduce the risk from identified causes of human errors.

- Paper II identifies the need for future research on capturing the uncertainty of human error probability (HEP) that is calculated as per Petro-HRA methodology. A suggested way it to estimate the distribution of HEP vs performance-shaping factors' assigned

weightage, by simulating the effect of Performance Shaping Factors (PSF) on human performance.

- The range of context of PSFs for the petroleum industry is quite large. This implies that the nuclear-derived PSF definitions, number of PSFs, nominal values and PSF rating scheme need to be revised, to more closely reflect the petroleum industry. The nuclear industry conducted its own research over the years; it is available today in the form of the SPAR-H methodology. Paper II identifies and recommends similar research to construct the petroleum industry's own PSF rating scheme and embed guidance in the Petro-HRA methodology, to capture and communicate the strength-of-knowledge aspect.

- The machine-learning technique developed in Paper III is an initial step in the direction of automated decision-support for the operational environment. There is a need to extend it to other more complex operations, involving several physical variables, by experimentation with the neural network's structure. Further, the use of a simple threshold value to minimise the model's uncertainty or prediction error can be improvised, using other approaches, metrics, indicators (e.g., slope, variance, etc.).

- Paper V identified that the value-adding capability of the safety diagnosability principle was limited. One of the factors was its lack of guidance in capturing the failure of human and organisational barriers. Although unaddressed in Paper V, it is suggested that the new safety principle be improved in this way through further research.

References

Abrahamsen, E. B., Abrahamsen, H. B., Milazzo, M. F., & Selvik, J. T. (2018a). Using the ALARP principle for safety management in the energy production sector of chemical industry. Reliability Engineering & System Safety, 169, 160-165.

Abrahamsen, E. B., Aven, T., & Iversen, R. S. (2010). Integrated framework for safety management and uncertainty management. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 224(2), 97-103.

Abrahamsen, E. B., Aven, T., Vinnem, J. E., & Wiencke, H. (2004). Safety management and the use of expected values. Risk, Decision and Policy, 9(4), 347-357. Doi: 10.1080/14664530490896645

Abrahamsen, E. B., Moharamzadeh, A., Abrahamsen, H. B., Asche, F., Heide, B., & Milazzo, M. F. (2018b). Are too many safety measures crowding each other out? Reliability Engineering and System Safety, 174, 108-113.

Abrahamsen, H. B., & Abrahamsen, E. B. (2015). On the appropriateness of using the ALARP principle in safety management. Safety and Reliability of Complex Engineered Systems: ESREL 2017, 773-777. London, UK: CRC Press

Adler, M. (2012). Well-being and fair distribution: beyond cost-benefit analysis. Oxford University Press, Oxford.

Almklov, P. G., Antonsen, S., Størkersen, K. V., & Roe, E. (2018). Safer societies. Safety Science.

Althaus, C. E. (2005). A disciplinary perspective on the epistemological status of risk. Risk Analysis: An International Journal, 25(3), 567-588.

Amalberti, R., Auroy, Y., Berwick, D., & Barach, P. (2005). Five system barriers to achieving ultrasafe health care. Annals of Internal Medicine, 142, 756-764.

Amyotte, P. R., Goraya, A. U., Hendershot, D. C., & Khan, F. I. (2007). Incorporation of inherent safety principles in Process Safety Management. Process Safety Progress, 26, 333-346.

Antonsen, S., Skarholt, K., & Ringstad, A. J. (2012). The role of standardization in safety management–A case study of a major oil & gas company. Safety Science, 50(10), 2001-2009.

Apostolakis, G. (1990). The concept of probability in safety assessments of technological systems. Science, 250(4986), 1359-1364.

Apostolakis, G. E. (2004). How useful is quantitative risk assessment?. Risk Analysis: An International Journal, 24(3), 515-520.

Askeland, T., Flage, R., & Aven, T. (2017). Moving beyond probabilities–strength of knowledge characterisations applied to security. Reliability Engineering & System Safety, 159, 196-205.

Attoh-Okine, N., & Ayyub, B. (2005). Applied Research in Uncertainty Modeling and Analysis, Vol. 20. Springer Verlag.

Aven, T. (2007). A unified framework for risk and vulnerability analysis covering both safety and security. Reliability Engineering & System Safety, 92(6), 745-754.

Aven, T. (2008a). Risk Analysis. Chichester: John Wiley & Sons Ltd.

Aven, T. (2008b). Risk Analysis – Assessing Uncertainties beyond Expected Values and Probabilities. New York: Wiley.

Aven, T. (2009). Perspectives on risk in a decision-making context–review and discussion. Safety Science, 47(6), 798-806.

Aven, T. (2010a). Misconceptions of Risk. John Wiley & Sons Incorporated. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/uisbib/detail.action?docID=477869.

Aven T. (2010b). What is risk?: towards a unifying approach. Riskbooks, London; 437-455.

Aven, T. (2010c). On the need for restricting the probabilistic analysis in risk assessments to variability. Risk Analysis: An International Journal, 30(3), 354-360.

Aven, T. (2011). A risk concept applicable for both probabilistic and non-probabilistic perspectives. Safety Science, 49(8-9), 1080-1086.

Aven, T. (2012a). Foundations of Risk Analysis. John Wiley & Sons.

Aven, T. (2012b). The risk concept—historical and recent development trends. Reliability Engineering & System Safety, 99, 33-44.

Aven, T. (2014). Risk, Surprises and Black Swans: Fundamental Ideas and Concepts in Risk Assessment and Risk Management. Routledge.

Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. European Journal of Operational Research, 253(1), 1-13.

Aven, T. (2020). The Science of Risk Analysis: Foundation and practice. 1st ed., Routledge. Routledge. https://doi-org.ezproxy.uis.no/10.4324/9780429029189

Aven, T., Baraldi, P., Flage, R., & Zio, E. (2013). Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods. John Wiley & Sons.

Aven, T., & Krohn, B. S. (2014). A new perspective on how to understand, assess and manage risk and the unforeseen. Reliability Engineering & System Safety, 121, 1-10.

Aven, T., & Kørte, J. (2003). On the use of risk and decision analysis to support decision-making. Reliability Engineering & System Safety, 79(3), 289-299.

Aven, T., & Renn, O. (2009). On risk defined as an event where the outcome is uncertain. Journal of Risk Research, 12(1), 1-11.

Aven, T. and Vinnem, J.E. (2007). Risk Management: with Applications from the Offshore Petroleum Industry. New York: Springer Verlag

Aven, T., & Zio, E. (2018). Knowledge in Risk Assessment and Management. John Wiley & Sons.

Ayyub, B. M., & Klir, G. J. (2006). Uncertainty Modeling and Analysis in Engineering and the Sciences. CRC Press.

Baybutt, P. (2014). The ALARP principle in process safety. Process Safety Progress, 33(1), 36-40.

Bedford, T., & Cooke, R. (2001). Probabilistic risk analysis: foundations and methods. Cambridge, UK: Cambridge University Press.

Bello, O., Teodoriu, C., Yaqoob, T., Oppelt, J., Holzmann, J., & Obiwanne, A. (2016). Application of artificial intelligence techniques in drilling system design and operations: a state of the art review and future research pathways. In SPE Nigeria Annual International Conference and Exhibition. Society of Petroleum Engineers.

Berg, H. P., Griebel, S., & Milius, B. (2015). Comparing operations in nuclear and railways based on a socio-technical system model. In Safety and Reliability of Complex Engineered Systems – Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015, pp. 4359-65.

Bickel, J. E., & Bratvold, R. B. (2008). From uncertainty quantification to decision making in the oil and gas industry. Energy Exploration & Exploitation, 26(5), 311-325.

Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2017). Cost-Benefit Analysis: Concepts and Practice. Cambridge University Press.

Boring, R. L. (2015). Adapting human reliability analysis from nuclear power to oil and gas applications. No. INL/CON-15-35411. Idaho National Lab. (INL), Idaho Falls, ID, United States.

Boring, R. L., & Blackman, H. S. (2007). The origins of the SPAR-H method's performance shaping factor multipliers. IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting, (177-184), doi: 10.1109/HFPP.2007.4413202.

Bratvold, R. B., & Begg, S. (2009). Making Good Decisions. ProQuest Ebook Central https://ebookcentral-proquest-com.ezproxy.uis.no.

Broadribb, M., & Flynn, S. A. (2009). Years on from Texas City. In AIChE Global Congress on Process Safety.

Breyer, S. (1993). Breaking the Vicious Circle: Toward Effective Risk Regulation. Massachusetts: Harvard University Press; British Medical Association.

Burgherr, P., & Hirschberg, S. (2008). A comparative analysis of accidentrisks in fossil, hydro, and nuclear energy chains. Human and Ecological Risk Assessment, 14(5), 947-973, DOI: 10.1080/10807030802387556

Burgherr, P., & Hirschberg, S. (2014). Comparative risk assessment of severe accidents in the energy sector. Energy Policy, 1(74), S45-56.

Bye, A., Laumann, K., Taylor, C., Rasmussen, M., Øie, S., Van de Merwe, K., Øien, K., Boring, R., Paltrinieri, N., Wærø, I., Massaiu, S., & Gould, K. (2017). The Petro-HRA Guideline. Report no. IFE/HR/E-2017/001. Halden, Norway

Center for Chemical Process Safety (CCPS). (1995). Tools for Making Acute Risk Decisions with Chemical Process Safety Applications, American Institute of Chemical Engineers, Center for Chemical Process Safety, New York, NY.

Center for Chemical Process Safety (CCPS) & American Institute of Chemical Engineers. (2007). Guidelines for Risk Based Process Safety. Wiley-AIChE.

Center for Chemical Process Safety (CCPS), et al. (2009). Guidelines for Process Safety Metrics, American Institute of Chemical Engineers, 2009. ProQuest Ebook Central, https://ebookcentral-proquest.com.ezproxy.uis.no/lib/uisbib/detail.action?docID=477845.

Center for Chemical Process Safety (CCPS). (2010). Guidelines for Risk Based Process Safety. John Wiley & Sons.

Center for Chemical Process Safety (CCPS). (2011). Process Safety Leading and Lagging Metrics. Center for Chemical Process Safety. AIChE, New York, http://www. aiche. org/sites/default/files/docs/pages/CCPS. ProcessSafety-Lagging_2011-2-24. pdf (accessed 07.06.15).

Center for Chemical Process Safety (CCPS). (2019). Guide for Making Acute Risk Decisions. American Institute of Chemical Engineers, 2019.

Chandrasekaran, S. (2016). Health, Safety, and Environmental Management in Offshore and Petroleum Engineering. John Wiley & Sons.

Chen, M. (2016). Process safety knowledge management in the chemical process industry. American Journal of Chemical Engineering, 4(5), 131-138.

Chierici, L., Fiorini, G. L., La Rovere, S., & Vestrucci, P. (2016). The evolution of defense in depth approach: A cross sectorial analysis. Open Journal of Safety Science and Technology, 6(2), 35-54.

Conejo, A. J., Carrión, M., & Morales, J. M. (2010). Decision Making Under Uncertainty in Electricity Markets (Vol. 1). New York: Springer.

Day, R. A., & Gastel, B. (2006). How to Write and Publish a Scientific Paper. Cambridge University Press.

De Moivre, A. D. (1711). De mensura sortis, seu, de probabilitate eventuum in ludis a casu fortuito pendentibus. Philosophical Transactions of the Royal Society of London, 27(329), 213-264.

Dempster, A. P. (1967). Upper and Lower Probabilities Induced by a Multivalued Mapping. Annals of Mathematical Statistics. Vol. 38, 325-339.

Deore, P. (2012, March). Decision making in upstream oil and gas industry-an integrated approach. In SPE Oil and Gas India Conference and Exhibition. OnePetro.

De Rademaeker, E., Suter, G., Pasman, H. J., & Fabiano, B. (2014). A review of the past, present and future of the European loss prevention and safety promotion in the process industries. Process Safety and Environmental Protection, 92(4), 280-291.

Det Norske Veritas (DNV), (2012). Safety & Loss Control and the International Safety Rating System (ISRS).

Durbach, I. N., & Stewart, T. J. (2009). Using expected values to simplify decision making under uncertainty. Omega, 37(2), 312-330. doi:10.1016/j.omega.2007.02.001

Edwards, W. (1954 July). The theory of decision making. Psychological Bulletin, 51(4), 380.

Emblemsvåg, J. (2012). Risk Management for the Future: Theory and Cases. BoD–Books on Demand.

Engemann, K. J., & Abrahamsen, E. B. (2020). Advances in safety risk management. In Safety Risk Management. Berlin, Boston: De Gruyter Oldenbourg, doi: https://doi.org/10.1515/9783110638189-202

Ferson, S., & Ginzburg, L. R. (1996). Different methods are needed to propagate ignorance and variability. Reliability Engineering & System Safety, 54(2-3), 133-144.

Flage, R., & Aven, T. (2009). Expressing and communicating uncertainty in relation to quantitative risk analysis. Reliability: Theory & Applications, 4(2-1), 13.

Flage, R., Aven, T., Zio, E., & Baraldi, P. (2014). Concerns, challenges, and directions of development for the issue of representing uncertainty in risk assessment. Risk Analysis, 34(7), 1196-1207.

Fritzsche, A. F. (1989). The health risks of energy production. Risk Analysis, 9(4), 565-77.

Gabor, P. (2020 March). Cross-industry learning from high hazard sectors. The Chemical Engineer, 26 March 2020.

Gould, K. S., Ringstad, A. J., & van de Merwe, K. (2012, September). Human reliability analysis in major accident risk analyses in the Norwegian petroleum industry. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 56, No. 1, pp. 2016-2020). Los Angeles, CA: Sage Publications.

Grote, G. (2012). Safety management in different high-risk domains–all the same? Safety Science, 50(10), 1983-92.

Gu, Z. (2018). History review of nuclear reactor safety. Annals of Nuclear Energy, 1(120), 682-90.

Hammond, J. S., Keeney R. L., & Raiffa, H. (1999). Smart Choices: A Practical Guide to Making Better Life Decisions. Boston: Harvard Business School Press.

Harms-Ringdahl, L. (2004). Relationships between accident investigations, risk analysis, and safety management. Journal of Hazardous Materials, 111, 13-19.

Hassan, J., & Khan, F. (2012). Risk based asset integrity indicators. Journal of Loss Prevention in the Process Industries, 25.

Hayes, J. (2017). Operational Decision-Making in High-Hazard Organizations: Drawing a Line in the Sand. CRC Press.

Heinrich, H. W., Petersen, D., & Roos, N. (1980). Industrial accident prevention: a safety management approach. (5th ed.). New York: McGraw-Hill.

Helton, J. C., & Burmaster, D. E. (1996). Guest editorial: treatment of aleatory and epistemic uncertainty in performance assessments for complex systems. Reliability Engineering and System Safety, 54, 91-94.

Hillson, D. A., & Hulett, D. T. (2004). Assessing risk probability: Alternative approaches. In Proceedings of PMI Global Congress (pp. 1-7).

Hirschberg, S., Spiekerman, G., & Dones, R. (1998). Severe Accidents in the Energy Sector—First Edition. PSI Report No. 98-16. Paul Scherrer Institut, Villigen PSI, Switzerland

Hollnagel, E. (2014). Safety-I and Safety-II: The Past and Future of Safety Management. Taylor & Francis Group.

Hopkins, A. (2011). Risk-management and rule-compliance: Decision-making in hazardous industries. Safety Science, 49(2), 110-120.

Huang, J. P., Poh, K. L., & Ang, B. W. (1995). Decision analysis in energy and environmental modeling. Energy, 20(9), 843-855.

Hudson, P., (2003). Applying the lessons of high-risk industries to health care. Quality & Safety in Health Care 12, i7-i12.

Inhaber, H. (2004). Risk analysis applied to energy systems. Encyclopedia of Energy, ISBN 9780121764807

International Atomic Energy Agency (IAEA) (2006). Safety Standards for Protecting People and the Environment. Fundamental Safety Principles, No SF-1.

International Atomic Energy Agency (IAEA), (2006b). Fundamental Safety Principles, SF-1 (2006b)

International Atomic Energy Agency (IAEA) (Corporate Author), & IAEA (Corporate Editor). (2011). Framework for an Integrated Risk Informed Decision Making Process, A.

International Risk Governance Center (IRGC). (2017). Introduction to the IRGC Risk Governance Framework. Authors: Florin, Marie-

Valentine, and Marcel Thomas Bürkler. (No. REP_WORK), 2017.

Institute of Risk Management (IRM). (2018). A Risk Practitioners Guide to ISO 31000 – 2018.

ISO/IEC. (2014). Safety Aspects – Guidelines for their inclusion in standards. ISO/IEC Guide, 51.

International Organization for Standardization. (ISO). (2018). ISO 31000:2018. Risk Management – Guidelines.

Jovašević-Stojanović, M., & Stojanovic, B. (2009). Performance indicators for monitoring safety management systems in chemical industry. Chemical Industry and Chemical Engineering Quarterly/CICEQ, 15(1), 5-8.

Kaplan, S., & Garrick, B. (1981). On the quantitative definition of risk. Risk Analysis, 1(1), 11-27.

Kettunen, J., Reiman, T., & Wahlström, B. (2007). Safety management challenges and tensions in the European nuclear power industry. Scandinavian Journal of Management, 23(4), 424-444.

Khan, F., Rathnayaka, S., & Ahmed, S. (2015). Methods and models in process safety and risk management: Past, present and future. Process Safety and Environmental Protection, 98, 116-147.

Knegtering, B., & Pasman, H. (2013). The safety barometer: How safe is my plant today? Is instantaneously measuring safety level utopia or realizable? Journal of Loss Prevention in the Process Industries, 26(4), 821-829.

Kontogiannis, T., Leva, M. C., & Balfe, N. (2017). Total safety management: principles, processes and methods. Safety Science, 100, 128-142.

Kothari, C. R. (2004). Research Methodology: Methods and Techniques. New Age International.

Kozine, I., Duijm, N. J., & Lauridsen, K. (2001). Safety-and Risk Analysis Activities in Chemical Industry in Europe. Proceedings of Values in decisions on risk (VALDOR). (No. NEI-SE--436).

Learning. (n.d.). In Merriam-Webster's collegiate dictionary. https://www.merriam-webster.com/dictionary/learning.

Lev, V. (2007). Analysis of multi-criteria decision-making methodologies for the petroleum industry. In International Petroleum Technology Conference.

Li, Y., & Guldenmund, F. W. (2018). Safety management systems: A broad overview of the literature. Safety Science, 103, 94-123.

Lu, J., Jain, L., & Zhang, G. (2012). Handbook on Decision Making, Vol 2: Risk Management in Decision Making (1st ed., Vol. 33, Intelligent Systems Reference Library). Berlin, Heidelberg: Springer Berlin Heidelberg; Imprint: Springer.

Massaiu, S., & Paltrinieri, N. (2016). Human reliability analysis: From the nuclear to the petroleum sector. In Dynamic Risk Analysis in the Chemical and Petroleum Industry (pp. 171-179). Butterworth-Heinemann.

Milazzo, M. F., Abrahamsen, E. B., Selvik, J. T., & Abrahamsen, H. B. (2020). Safety investment decisions in the chemical industry: Common approaches and case studies. In Safety Risk Management. Berlin, Boston: De Gruyter Oldenbourg, doi: https://doi.org/10.1515/9783110638189-012

Miller, C. O. (1988). System safety. In Wiener, E. L., & Nagel, D. C. (eds.) Human Factors in Aviation, pp. 53-80. San Diego: Academic.

Mitchell, T. M. (1997). Machine Learning, pp. 367-369. McGraw-Hill Science/ Engineering/ Math.

Mohaghegh, S. D., & Khazaeni, Y. (2011). Application of artificial intelligence in the upstream oil and gas industry. International Journal of Computer Research, 18(3/4), 231.

Morton, M. S. S. (1971). Management Decision Systems: Computer-Based Support for Decision Making. Division of Research, Graduate School of Business Administration, Harvard University.

Möller, N. (2012). The concepts of risk and safety. Handbook of Risk Theory: Epistemology, Decision Theory, Ethics, and Social Implications of Risk. p Springer Netherlands, Dordrecht. p.55-85.

Niehaus, F., & Szikszai, T. (2001). Risk informed decision making. Topical Issues paper no. 1.

Norazahar, N., Khan, F., Veitch, B., & MacKinnon, S. (2014). Human and organizational factors assessment of the evacuation operation of BP Deepwater Horizon accident. Safety Science, 70, 41-49.

North, D. W. (2010). Probability theory and consistent reasoning. Risk Analysis: An International Journal, 30(3), 377-380.

Noshi, C. I., and Schubert, J. J. (2018, October). The role of machine learning in drilling operations; A review. SPE/AAPG Eastern Regional Meeting. Society of Petroleum Engineers. https://doi.org/10.2118/191823-18ERM-MS

Norwegian Research Council (2000). Kvalitet i norsk forskning: En oversikt over begreper, metoder og virkemilder.

Organisation for Economic Co-operation and Development/Nuclear Energy Agency OECD/NEA, (2006). State-of-the-art report on systematic approaches to safety management (Report No. NEA/CSNI/R(2006)1). Issy-les-Moulineaux: OECD Nuclear Energy Agency.

Palazzi, E., Caviglione, C., Reverberi, A. P., & Fabiano, B. (2017). A short-cut analytical model of hydrocarbon pool fire of different geometries, with enhanced view factor evaluation. Process Safety and Environmental Protection, 110, 89-101.

Paltrinieri, N., Dechy, N., Salzano, E., Wardman, M., & Cozzani, V. (2012). Lessons Learned from Toulouse and Buncefield Disasters: From Risk Analysis Failures to the Identification of Atypical Scenarios Through a Better Knowledge Management. Risk Analysis: An International Journal, 32(8), 1404-1419.

Paltrinieri, N., & Khan, F. (Eds.). (2016). Dynamic Risk Analysis in the Chemical and Petroleum Industry: Evolution and Interaction with Parallel Disciplines in the Perspective of Industrial Application. Butterworth-Heinemann.

Parida, A., & Chattopadhyay, G. (2007). Development of a multi-criteria hierarchal framework for maintenance performance measurement (MPM). Journal of Quality in Maintenance Engineering, 13(3), 241-258.

Parker, G. (1996). The Military Revolution: Military Innovation and the Rise of the West, 1500-1800. Cambridge University Press.

Pasman, H. J., Jung, S., Prem, K., Rogers, W. J., & Yang, X. (2009). Is risk analysis a useful tool for improving process safety? Journal of Loss Prevention in the Process Industries, 22(6), 769-777.

Paté-Cornell, M. E. (1996). Uncertainties in risk analysis: Six levels of treatment. Reliability Engineering & System Safety, 54(2-3), 95-111.

Pearl, M. (2007). Creating a Competitive Edge: The Value of Cross-Industry Knowledge. Business Strategy Series, 8(2), 142-147.

Perrow, C. (1984). Normal Accidents: Living with High Risk Technologies. Princeton University Press.

Petersen, D., 1978. Techniques of Safety Management. McGraw-Hill, New York.

Petroleum Safety Authority Norway (PSAN). (2016). The Concept of Risk in the Petroleum Business.

Petroleum Safety Authority Norway (PSAN). (2018). Integrated and Unified Risk Management in the Petroleum Industry. Retrieved from https://www.ptil.no/contentassets/8d93722526cb4c57a5068e680be90a7b/risikostyring-2018-engelsk.pdf.

Rausand, M. (2013). Risk Assessment: Theory, Methods, and Applications. Vol. 115. John Wiley & Sons.

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. Safety Science, 27, 183-213

Rasmussen, N. C. (1981). The application of probabilistic risk assessment techniques to energy technologies. Annual Review of Energy, 6(1),123-38.

Reason, J. (1995). A systems approach to organizational error. Ergonomics, 38(8), 1708-1721.

Reiman, T., & Oedewald, P. (2007). Assessment of complex sociotechnical systems–Theoretical issues concerning the use of organizational culture and organizational core task concepts. Safety Science, 45(7), 745-768.

Rosa, E. A. (1998). Metatheoretical foundations for post-normal risk. Journal of Risk Research, 1(1), 15-44.

Rosness, R., Guttormsen, G., Steiro, T., Tinmannsvik, R. K., & Herrera, I. A. (2004). Organisational Accidents and Resilient Organisations: Five Perspectives Revision. Trondheim: SINTEF Industrial Management.

Saleh, J., Haga, R. A., Favarò, F. M., & Bakolas, E. (2014). Texas City refinery accident: Case study in breakdown of defense-in-depth and violation of the safety–diagnosability principle in design. Engineering Failure Analysis, 36, 121-133.

Selvik, J. T., & Signoret, J. P. (2017). How to interpret safety critical failures in risk and reliability assessments. Reliability Engineering & System Safety, 161, 61-68.

Selvik, J. T., Stanley, I., & Abrahamsen, E. B. (2020). SMART criteria for quality assessment of key performance indicators used in the oil and gas industry. International Journal of Performability Engineering, 16(7), 999-1007.

Shafer, G. (1976). A Mathematical Theory of Evidence (Vol. 42). Princeton University Press.

Shafiee, M., Animah, I., Alkali, B., & Baglee, D. (2019). Decision support methods and applications in the upstream oil and gas sector. Journal of Petroleum Science and Engineering, 173, 1173-1186.

Simon, H. A. (1959 June). Theories of decision-making in economics and behavioral science. The American Economic Review, 49(3), 253-83.

Skogdalen, J. E., & Vinnem, J. E. (2011). Quantitative risk analysis offshore—human and organizational factors. Reliability Engineering & System Safety, 96(4), 468-479.

Sloan, S. (2007 June). Risk Management vs. Safety Management: Can't we all just get along? Paper presented at the ASSE Professional Development Conference, Orlando, Florida.

Society of Risk Analysis (SRA). (2015). Risk Analysis: Fundamental Principles. Retrieved from https://www.sra.org/wp-content/uploads/2020/04/SRA-Fundamental-Principles-R2.pdf

Society for Risk Analysis (SRA). (2018). Society for Risk Analysis Glossary. Retrieved from https://www.sra.org/wp-content/uploads/2020/04/SRA-Glossary-FINAL.pdf

Soroudi, A., & Amraee, T. (2013). Decision making under uncertainty in energy systems: State of the art. Renewable and Sustainable Energy Reviews, 28, 376-384.

Sprague Jr., R. H., & Watson, H. J. (Eds.). (1993). Decision Support Systems Putting Theory into Practice. Prentice-Hall, Inc.

Sørskår, L. I. K., Selvik, J. T., & Abrahamsen, E. B. (2019). On the use of the vision zero principle and the ALARP principle for production loss in the oil and gas industry. Reliability Engineering & System Safety, 191, 106541.

Tench, W. (1985). Safety is No Accident. London: Collins.

Turban, E. (1995). Decision Support and Expert Systems: Management Support Systems. Prentice-Hall, Inc.

United States Nuclear Regulatory Commission (USNRC). (1975). Reactor Safety Study. An assessment of accident risks in US commercial nuclear power plants. WASH-1400 (NUREG-75/OI4), Washington DC.

United States Nuclear Regulatory Commission (USNRC). (2016). Historical Review and Observations of Defense-in-Depth (NUREG/KM-0009), Washington DC.

Vaurio, J. K. (1998). Safety-related decision making at a nuclear power plant. Nuclear Engineering and Design, 185(2-3), 335-345.

Verma, A. K., Ajit, S., & Karanki, D. R. (2010). Reliability and Safety Engineering, Vol. 43, pp. 373-392. London: Springer.

Verma, A. K., Srividya, A., Gopika, V., & Rao, K. D. (2011). Risk-informed decision making in nuclear power plants. In Safety and Risk Modeling and its Applications, pp. 325-363. Springer: London.

Walley, P. (1991). Statistical reasoning with imprecise probabilities. Chapman and Hall/CRC Monographs on Statistics & Applied Probability, ISBN 13: 978-1489934734

Zhen, Xingwei, Vinnem, Jan Erik, Yang, Xue, & Huang, Yi. (2020). Quantitative risk modelling in the offshore petroleum industry: Integration of human and organizational factors. Ships and Offshore Structures, 15(1), 1-18.

Zio, E., & Pedroni, N. (2012). Overview of risk-informed decision-making processes. FonCSI.

Zou, P., Sun, A., Long, B., & Marix-Evans, P. (2010). Return on investment of safety risk management system in construction. Paper presented at the Proc., CIB World Congress.

Part II

Paper I

Return on Investment (ROI) for evaluating safety measures. Review and discussion

Authors: Surbhi Bansal, Jon Tømmerås Selvik, Eirik Bjorheim Abrahamsen

Published in *The Business Review*, Cambridge, ISSN 1553-5827, Volume 26.

This paper is not available in Brage due to copyright restrictions.

Paper II

Alignment of the Petro-HRA method with the risk perspectives in the Norwegian oil and gas industry

Authors: Surbhi Bansal, Jon Tømmerås Selvik, Eirik Bjorheim Abrahamsen

Published in *Proceedings of the 29th European Safety and Reliability Conference (ESREL)*. ISBN 978-981-11-2724-3.

Paper III

Development of a bivariate machine-learning approach for decision-support in offshore drilling operations.

Authors: Surbhi Bansal, Nejm Saadallah, Jon Tømmerås Selvik, Eirik Bjorheim Abrahamsen

Published in *Proceedings of the 30th European Safety and Reliability Conference (ESREL2020), 15th Probabilistic Safety Assessment and Management Conference*, (PSAM15) 15, ISBN 978-981-14-8593-0

Paper IV

On the use of criteria based on the SMART acronym to assess quality of performance indicators for safety management in process industries.

Authors: Jon Tømmerås Selvik, Surbhi Bansal, Eirik Bjorheim Abrahamsen

Published in the *Journal of Loss Prevention in the Process Industries*. 2020, p.104392. ISSN 0950-4230, https://doi.org/10.1016/j.jlp.2021.104392

On the use of criteria based on the SMART acronym to assess quality of performance indicators for safety management in process industries

Jon Tømmerås Selvik[a,b*], Surbhi Bansal[a] and Eirik Bjorheim Abrahamsen[a]

[a]*University of Stavanger, 4036 Stavanger, Norway*
[b]*NORCE Norwegian Research Centre AS, 4068 Stavanger, Norway*

Abstract

Management of safety, and barriers in particular, includes using information expressing performance, i.e. use of safety performance indicators. For this information to be useful, the indicators should demonstrate adequate quality. In other words, they should satisfy some predefined set of quality criteria. Without showing adequate quality, the indicators are generally unable to provide sufficient support for barrier management, which could result in poor decisions. In this article, the use of the SMART criteria is considered to assess the quality of safety performance indicators in process industries. SMART being an acronym for 'specificity', 'measurability' or 'manageability', 'achievability', 'relevancy' and 'time-based', covering five key aspects and criteria for assessing the quality of an indicator. A discussion on whether the indicators are able to demonstrate adequate quality by satisfying these criteria has been conducted. The finding is that all of the SMART criteria should be satisfied for a safety performance indicator to demonstrate acceptable quality and to be regarded as useful to support barrier management decision-making. However, it has also been observed that including the 'M' criterion in the assessment of quality is not needed. When all the other criteria are satisfied there is no way the conclusions could be misleading as a result of measurability or manageability aspects. Hence, for safety performance indicator quality, only four of the criteria are assessed and suggested for such situations to shorten the acronym to 'STAR'. A key safety indicator used in downstream process facilities, i.e. 'dangerous fluid overfilling events', motivated from the 2005 Texas City refinery accident, is used to illustrate the situation. The indicator is also applied to another incident, the Buncefield oil storage depot's accident in 2005, to provide a broader context for using it. The findings in this article could also be applied beyond the context studied. This means that, despite focusing on safety indicators in the process industries, the findings are considered as relevant and applicable to other types of performance indicators and to other energy industries.

Keywords: Performance indicators, safety, barrier management, SMART, criteria, quality, process industries

1. Introduction

In this article, the focus is on achieving useful performance indicators to support decision-making related to safety and barrier management in the process industries. For example, when adopting the "safety diagnosable principle" or "defence in depth" it is essential to have appropriate indicators measuring barrier conditions; see Saleh et al. (2014a; 2014b). A variety of safety performance indicators (SPI) are used for this purpose and included in indicator portfolios to provide a sufficiently broad information basis. However, the usefulness is challenged by quality, as information from some indicators might be misleading or totally disregarded in practise but nevertheless be associated with costs. Consequently, assessment of SPI quality is an important activity related to the construction and use of the performance indicator portfolio. Adequate quality links to the ability to meet safety target and business goals, and visions.

One common and in principle simple way to assess the quality of performance indicators is by using the SMART criteria, referring to five standard criteria covering main quality aspects (Badawy et al. 2016; Parida and Kumar 2006; Doran 1981). Basically, by verifying that the indicators satisfy the criteria, one avoids spending resources on collecting and analysing information not contributing with any or with poor business value. SMART being an acronym for:

- Specificity
- Measurability
- Achievability
- Relevancy
- Timeliness

These are further described in Section 3 and in Table 1.

Despite being commonly used, and quite intuitive in their relation to assessment of quality, it is not obvious that these criteria meet the objective of demonstrating SPIs with high quality, despite there being extensive literature available on different benefits and challenges related to performance indicators. In this article we focus on the SPI quality and relation to the SMART criteria, aiming to provide some clarification regarding how suited the SMART criteria are for the safety and barrier context. For this, we question whether these five criteria are appropriate for assessment of quality, or whether some adjustments are called for. There could be a need to add other criteria or reject some of those already present. Relevant criteria could perhaps be left out due to poor incentives, for example to keep the nice acronym created or simply be context related.

In a previous study by Selvik et al. (2020) discussing the use of these criteria in a general business context, it is suggested an 'M' swap, i.e. to include an assessment of 'manageability' instead of 'measureability'. This latter criterion is considered to make more sense when dealing with key performance indicators compared with business goals. Making the swap should make the SMART criteria better suited for assessing quality. See also discussion in Section 4. However, this is not necessarily the situation when studying indicators in a safety context, as there could then be other quality aspects being relevant. Particularly, it is not obvious that a 'manageability' criterion is needed,

as there should in principle always be possible to perform some safety-related action to improve current situation, otherwise it challenges the need of the 'relevancy'.

Regarding the assessment of SPI quality in the process industries, we believe it is important to consider the appropriateness of the SMART criteria as basis for demonstrating SPI quality. An objective of the article is thus to contribute to an improved framework for performing the assessments. As a basis for the discussions, we include also consideration of other criteria that could be applicable for the assessment of quality, being suggested in literature, such as e.g. adding 'explainability' and 'relativity' to extend the acronym into 'SMARTER' (Better Regulation Task Force 2000). There are also several other alternatives as presented by the overview in Section 3.

The article is structured as follows. Section 2 gives a brief introduction to barrier management and use of safety performance indicators, where different types of indicators can be combined into portfolios. Then Section 3 summarises the five SMART criteria. In Section 4, we discuss whether these five criteria in themselves are appropriate to use for the assessment of quality for a selected performance indicator. We also point to other criteria suggested in literature that could be considered. Then, in Section 5, we discuss how to combine the individual SPIs into a portfolio useful for decision-making purposes. In Section 6, we consider the overall perspective and discuss the use of the SMART criteria from a portfolio perspective, and how the indicator in focus influences the safety targets and overall business goals and visions. In Section 7, we refer to the 2005 Texas City refinery accident, and use this to illustrate the main points from the previous discussions. A main reason for referring to this specific accident, is the importance it illustrated for having quality performance indicators for process safety in the refinery and petrochemical industries, for example the developments of API 754 (API Recommended practise 754: 2010; 2016). A performance indicators program provides useful information for driving improvement and when acted upon, contributes to reducing risks of major hazards by identifying the underlying causes and taking action to prevent recurrence. In Section 8, the SPI is assessed by referring to another incident, the 2005 Buncefield oil storage depot accident, to illustrate its usefulness in a broader context. Finally, in Section 9, we give some conclusions, including recommendations regarding the appropriateness of using the SMART criteria in the context considered.

Measurement of performance in safety management

SPIs are used to provide insights into safety performance, something that is conceptually difficult to measure directly. The indicators are measures that express the level of safety performance achieved for a given system, particularly barriers, and representing a type of key performance indicator allowing for measurable results linked to both quantitative and qualitative findings (ISO 41011:2017). A safety indicator covers any indicator giving relevant information about the state of equipment, organization or human activity related to safety, for example the number of hydrocarbon leakages, which are type of events linked to higher risk for major accidents (Vinnem 2012). Another key indicator measuring barrier safety performance, is the

'failure fraction', which is e.g. used by the Petroleum Safety Authority Norway in their analysis of the risk level on the Norwegian Continental Shelf. It gives the ratio between number of failures and the corresponding number of tests performed (Selvik and Abrahamsen 2015). In general, the information achieved through the indicators should be able to help identify whether barrier- or safety-related actions are needed. As such, the use of such indicators are in line with the suggestions from particularly Saleh et al. (2014a; 2014b), pointing to the importance of the "safety-diagnosability principle", where focus is on the ability to identify dangerous states in the operations through observability. A key is to achieve reliable information about the barrier safety performance, where the selected indicator is suitable for the application and can be used for a meaningful evaluation of the performance.

Barrier management is a core part of the safety management, which in the process industries is about establishing and maintaining layers of protection against hazardous events to achieve specified safety objectives, as part of overall safety management. According to the Norwegian Petroleum Safety Authority, the purpose is "to establish and maintain barriers so that the risk faced at any given time can be handled by preventing an undesirable incident from occurring or by limiting the consequences should such an incident occur" (PSA 2013). It concerns having barriers, i.e. "functional grouping of safeguards or controls selected to prevent major accident or limit the consequences" (ISO 17776:2016), which could be of either technological, organizational or human character. For the technological barriers, terms such as 'hardware', 'process', 'process safety' or 'process-related' are often used to label the type of barrier. For the different types there exist also several sub-categorisations. Refer to e.g. NORSOK D-010 (2013), for operations on the Norwegian Continental Shelf, giving guidance for barriers in drilling and well systems; and, reports from the International Association of Oil and Gas Producers (IOGP 2016; IOGP 2018a), giving general categorisations and description of the "hardware" and "human" barrier types.

The barrier management and use of SPI, are similar to general use of key performance indicators, where the information acquired allows for informed decisions by evaluating the level of past, current or future performance. To support barrier management, multiple indicators (an indicator portfolio) are tracked, as the performance cannot usually be described from only one indicator. For example, regarding the quality of a barrier element, both reliability and maintenance information could be relevant and are normally evaluated. A list of relevant indicators from the reliability and maintenance field are given in Annex E of ISO 14224 (2016), which includes common measures such as the 'mean time to failure' (MTTR), the 'mean overall repair time' (MRT), and also 'technical availability' and 'operational availability'; see also EN 15341 (2017) guiding the use of maintenance indicators. Such measures are widely used across process industries and the combining of different SPI are important for the overall monitoring of barrier performance and safety management, but also for general business management though the link to safety objectives or goals.

OECD (2008) separates between 'activities' and 'outcome' indicators, in the context of chemical process barriers. Activities indicators are proactive, meaning that they provide information about ongoing activities and conditions, and/or development of these, expressing the potential of barrier failure or accidents. This type is often called 'leading' as the information is supposed to help predicting or giving some expectation

about future safety, before anything critical occurs. It is giving answers to 'why' safety performance is going in some direction. Outcome indicators, on the other side, are reactive. These intended to provide information about the effects of operations and actions taken, having then instead focus on observable events occurring. It addresses the current or past performance, thus giving answers to 'what occurred'. Often this latter type is labelled as 'lagging indicators'; see Kongsvik et al. (2011), Payne et al. (2009), Tamim et al. (2017), Smith and Mobley (2008) and IOGP (2018b). There is also a type called 'diagnostic indicators', used for performance indicators that are signal the health of processes or activities (Badawy et al. 2016; Peng et al. 2007). These are not directly linked to potential for safety events occurring, but rather focusing on the general safety culture level.

API Recommended Practise 754 (2016), strongly motivated by the 2005 Texas refinery accident (see Section 7), focus on both activities and outcome indicators. And both types should follow the same basic principles for quality:

- Indicators should drive process safety performance improvement and learning
- Indicators should be relatively easy to implement and easily understood by all stakeholders (e.g. workers and the public)
- Indicators should be statistically valid at one or more of the following levels: industry, company, and site
- Indicators should be appropriate for industry, company or site level benchmarking

It is clearly relevant to capture both activities and outcome indicators when evaluating safety performance. Further, as there are multiple indicators providing input, some structured approach, for dealing with them and combining the information, is required, for example, using balanced scorecards (Kaplan and Norton 1996; Vukomanovic and Radujkovic 2013). The scorecards allow for easier overview of the aspects measured and what tolerance levels the measures are tested against. The evaluation depends on what is the motivation of the indicator(s), beyond having a safety relation. There could be motivations such as:

- Evaluating the ability to meet objectives and safety targets
- Identifying focus and improvement areas
- Monitoring quantitative effect of actions taken
- Demonstrating that some benchmark level is satisfied

The SPIs provide key safety information, which gives them a role also in overall business management. A main task is to establish a link between the information achieved through the set of indicators selected, covering then a portfolio of SPI, and their ability to create overall value and quality in decision-making, where the quality of the SPIs obviously plays an important role.

2. SMART criteria overview

The SMART criteria have a broad application area, and are used for various key performance indicators, not only safety or barrier indicators. The reference to these criteria in relation to assessment of quality allows for a transparent process, where each of the criteria needs to be assessed and satisfied. It is a common way of considering quality aspects of information potentially having business value. This because the information links to decisions that influence achievement of goal, targets and visions (Parida and Kumar 2006; Kaganski and Toompalu 2017). By satisfying all five of the SMART criteria, the information provided by the indicator demonstrates usefulness as well as adequate quality. See also Doran (1981), which is often cited in relation to quality, goals and business objectives. For the history of the development of 'SMART', we refer to for example Lawlor and Hornyak (2012).

The five criteria are listed in Table 1 along with a brief description on what is covered. When all five criteria are satisfied, then the SPI in principle is having adequate quality to inform decision-making in barrier management.

Table 1. SMART criteria for assessment of performance indicator quality

Criterion	Description
Specificity	Precision; the indicator should be sufficiently precise. It should be clear what the indicator expresses (measures); the parameters of the measure should be unambiguous; and the numbers should not depend on who is producing them and who is interpreting them (i.e. consistent interpretation).
Measurability	Comparability; it should be possible to quantify and compare to other data, e.g. progress towards the attainment of the objectives, where it should reflect the level of general development in a certain aspect. The data on the parameters defining the indicator measure should be collectable and available in sufficiently high quality.
Achievability	Attainability; it should be possible (realistic) to achieve the objectives on which the indicator is based. The indicator should provide adequate information, with respect to confirming attainment of the objective.
Relevancy	The indicator should provide essential information for business management and improvement (i.e. aligned with business objectives). The indicator should thus be important for business performance.
Time-based	The indicator value should cover an appropriate period (a predefined and relevant time-frame period). Too short a period provides limited knowledge about the aspects studied.

Above SMART is presented as being "one" specific set of criteria. But in fact, there are different versions of the SMART acronym being used, where the letters could refer to other aspects or criteria. As one example, the letter 'S' sometimes refers to 'sustainable', 'A' sometimes refers to 'attainable', the 'R' to 'realistic' and the 'T' to 'traceable', but typically, the combination of alternatives suggested in literature covers more or less similar meaning. The following overview gives examples of possible alternatives for the letters used in and applicable for the SMART acronym:

S: Short; Sensible; Simple; Significant; Strategic; Stretching; Sustainable
M: Maintainable; Manageable; Meaningful; Motivating
A: Acceptable; Adjustable; Adaptable; Action-oriented; Agreeable; Aligned; Appropriate; Attainable

R: Relative; Results-oriented; Rewarding; Reviewable; Robust
T: Trackable; Traceable; Tangible; Time- (bound; constrained; constricted; related; specific)

There are also acronym variations, such as for example 'SMARTER', which extends with two additional letters and criteria. This is suggested by several, for example Vukomanovic and Radujkovic (2013), Kaufman et al. (2003) and Galligan et al. (2000). The common meaning of the new letters 'E' and 'R' then being (Better Regulation Task Force 2000) 'explainability', meaning that the indicator is simple to understand and communicate; and 'relativity', meaning that the indicator is still considered as useful or applicable if business conditions change (for example if production volume increases), respectively. Regarding the new 'E', 'explainability', it might be argued that this is similar to the criteria 'specificity' used for the letter 'S'. And, the new 'R', relativity', is to some extent is already covered by the criteria 'relevancy' being already used for the letter 'R', which expresses relevancy in changed business conditions. Hence, there are reasons to question whether the added letters add much or whether these additions are more motivated by a motivation to come up with something new or design a catchy acronym.

There are also other acronyms that extends 'SMART' by adding just one letter, such as for example 'SMAART', where the letter 'A's could refer to 'attainable' and 'action-oriented'. There is also 'C-SMART', attained by adding the letter 'C' for 'challenging' or 'controllable'. In addition to 'SMARTER', there are also other two-letter suggestions such as for example 'SMARTIE', adding 'I' and 'E', for 'inspiring' and 'enthusiasm'. It is another example of an acronym created to achieve a nice acronym, where the letter 'M' for 'motivating' could have been used instead but would not produce such a catchy acronym. Then, we have the double-layer 'SMART' variants, the 'SMART2' and 'SMART^2 ', meaning that each of the letters in the acronym is considered twice (RapidBI 2016; Kavanagh 2013).

For the discussions in the following sections, we will focus the criteria listed and described in Table 1. However, several of the other criteria mentioned above as potential candidates and possibly relevant quality aspects will to some extent be part of the discussion on whether a sufficiently broad quality picture is achieved by using the SMART criteria.

3. Use of the SMART criteria to assess quality of a SPI

In this section we address the assessment of SPI quality when disregarding the portfolio influence. We do not yet assess the influence from other indicators in the portfolio, and we only assess the quality of an individual and isolated SPI. It also means that we are not considering the managerial context and influence in the assessments, and thus fail to consider the broader picture. This simplifies the quality assessments, as there is then no need to cover the portfolio management and possible duplicity or conflict of interest between the indicators included. We leave to Section 5 the discussion related to the quality influence from the way the indicator portfolio is composed. The role of the SPI from a portfolio perspective is obviously important and relevant, but is for now ignored, meaning that a quality SPI, individually, does not depend on how it is used and balanced

with other indicators. Hence, we have the situation where a SPI could be acceptable, while the portfolio of SPIs, of which it is part of, could have low quality.

The value of the information provided by the indicator needs to be seen in relation to the decision-making where it applied. However, at the time when the indicator is selected, it might not be clear exactly how it will be used. Understanding how it will be used, makes it possible to consider the value it might have in barrier management. It is about usefulness. A main characteristic of quality in relation to quality decision-making is that the information is useful. According to Matheson and Matheson (1998), as one out of six dimensions characterising quality decision-making:

- Helpful frame (what is it that I am deciding?)
- Creative alternatives (what are my choices?)
- Useful information (what do I know?)
- Clear values (what consequences do I care about?)
- Sound reasoning (am I thinking straight about this?)
- Commitment to follow through (will I really take action?)

Being useful is about applicability for its area of use, but also means that it should compatible with the data handling tools being used, which is becoming important when dealing with software products, big data, etc. The combination of information provided by the SPI and applicability influences decision quality, and then also influences how such data can create business value. Further, Bratvold and Begg (2010) state that the two aspects 'reliable' and 'relevant' are part of the 'information usefulness'. 'Reliable' referring to both the source, how it is collected, and the content of the information provided. For the information to be 'reliable', it should be unbiased, representative and verifiable, such that the numbers give a correct representation of the situation. These aspects are to some extent already covered by the 'achievability' and 'relevancy' components in SMART, as then appropriate information is provided, the SPI is of interest to the context considered, and it has the ability or characteristics to influence the associated barrier management decision-making. What it means in practice, is that any indicator that is ambiguous, complicated, difficult to analyse, vague, analyst-dependent, or not linked to business objectives is obviously characterised as of poor quality, and thus not very useful or valuable.

The main question, then, is whether usefulness is adequately covered by the SMART criteria. If not, there is a strong argument for claiming that the criteria cannot be used to demonstrate acceptable quality. We will go through the five criteria and discuss this below.

We start with 'S' for 'specificity'. For the information to be useful, it is difficult to argue against the claim that it should be understandable and clearly expressed. There should not be any room for misinterpreting the meaning or definition of the indicator but be clear what kind of information it provides such that it is interpreted consistently. This relates also to the 'time-based' (T) criteria. There is no point in measuring the performance if the period considered is off. Overall, it is a matter of having precise knowledge. Implicitly then, aspects such as 'consistency', 'explainability', and 'transparency' are also covered.

Moving to ‘M’ for ‘measurability’, it could be questioned whether there is substantial need for this criterion, as any SPI, being a measure by definition, is measurable per se. Basically, the safety aspect addressed must be possible to measure. But, except the point that the indicator must be “qualified” as a safety-related measure, we do not see there is a need to include this criterion. Also, the criterion relates somewhat to whether the information needed to perform the calculations are possible to collect or produce with quality, but this is already covered by the following criterion, ‘achievability’ (A). This one is assumed very important, as it should be possible to produce the numbers with acceptable quality, which is evaluated from this criterion. For example, the calculation should not be overly complex. It could perhaps be better to use the term ‘producibility’, where it not for it starting with the “wrong” letter and would not give such a catchy acronym. Nevertheless, this criterion opens a way for capturing uncertainty. When including it, it comprises some evaluation of uncertainty regarding the numbers produced.

Finally, we have the ‘relevancy’ (R) criterion, on whether the indicator information matters to the management of safety performance. It would obviously be possible that it adds value beyond safety, for example provides general business value, but it would then not contribute to the barrier or safety management, which here is the focus and objective. There is a need to state whether the measurement reflects safety or barrier performance, not only measure some changing conditions, i.e. measure according to intention.

For the SPI to have safety-value, it should also be considered so-called ‘safety-sensitive’, which relates to the ‘relevancy’ criterion. One could maybe discuss exactly how sensitive the indicator needs to be; however, we find it here sufficient that there is such a relationship and will not pursue further discussion about the strength here. While we not yet will consider the situation from a portfolio perspective, ‘relevant’ also means that the particular safety-aspect measured is not already covered by other indicators used. Although there could for some situation be reasonable to include information from two or more indicators on similar aspects, it does not add much value except confirming the results or observations to be correct. Also, it is challenging to conclude on the usefulness of the SPI without considering the other SPIs used. Relevancy is to a large degree a managerial review activity, which cannot be disregarded when evaluating the usefulness of the SPI. For example, it depends on which safety or decision-making principles are adopted and how these are used. This activity involves assessing the whole portfolio, although it is clearly possible to make some decisions based on results from individual SPIs. But, the particular role of the SPI within the portfolio is an issue that is then not covered by the ‘relevancy’ criterion. Regarding the alternatives for the letter ‘R’, as indicated in Section 3, some suggest a ‘relativity’ criterion. We assume that this criterion is already covered by ‘specificity’ as the situation for which the SPI applies should be precisely described.

To summarise the discussion above (see Table 2), we conclude that all criteria deal with relevant quality aspects. The letters ‘S’ and ‘M’, and some degree ‘T’, refer to ‘what we know’ aspects, the letter ‘A’ refers to ‘how to use it’ aspects, the letter ‘R’ focus on ‘why’ aspects, and the letter ‘T’ refers to aspects related to ‘when or which period to consider’. For ‘measurability’ it may be questioned whether perhaps this criterion could be removed being implicitly already covered, as any SPI by definition

qualifies as a measure. In principle, there is not a problem keeping it, but it adds limited value. By including it we just achieve an assessment of whether the safety or barrier phenomena considered, is possible to measure, which is basically the same being assessed by the achievability criterion. However, we await the discussion from a portfolio perspective before making any conclusions on this issue.

Table 2. Overview of which of the SMART criteria are covered by 'usefulness'

Criterion	**Covered by usefulness?**	**Comment**
Specificity	Yes	Should be in place for clear understanding of the SPI, and for consistent use
Measurability	Yes, but not needed	It should be possible to compare the SPI numbers scientifically. However, this criterion is already covered as the indicator necessarily is a measure.
Achievability	Yes	The SPI must be producible in a consistent way
Relevancy	Yes	The aspect covered by the SPI should matter to safety or barrier management
Time-based	Yes	The SPI is of limited value if the time aspect is poorly covered

The five criteria discussed above seem all relevant to some degree, but there also other candidates that could be considered, to complement the aspects already covered. Neither of the letters links specifically to the aspect of 'how to use it', although, it is part of the 'relevancy' aspect, as it measures safety or barrier performance and implicitly assumes that a safety or a barrier action is required if performance is for example poor. But it is not fully covered by this. Say, for example, that we consider 'extreme weather events' as a basis for a SPI. Would such a measure satisfy all five criteria discussed above? For overall business performance, it might be the situation. But not for safety performance. Clearly, it would not be very useful as a SPI. Yes, extreme weather may have a safety impact, but it will be possible to takes precautionary or consequence-reducing measures. For safety and barrier management, any SPI that are checked as 'relevant', are implicitly associated with a possibility to make decisions influencing or controlling future outcomes recorded by the measure. Selvik et al. (2020) claim that one key quality characteristics, related to a discussion on key performance indicator quality in general, is that they are controllable. In a safety context, it means that that appropriate safety-related actions might have an effect and could improve SPI results, but as that is assumed to always be the situation, we cannot see a need for this criterion. In a safety or barrier context, if we are not able to improve safety with respect to the aspect considered, the indicator is not 'relevant' and of minimal usefulness. Hence, as for M in 'SMART', we cannot see that it matters much whether 'manageable' or 'measurable' is selected in the SMART acronym, both add aspects already covered by the other criteria.

An example of the use of the SMART criteria is given in Section 7, where the criteria are discussed with basis in the 2005 accident that occurred in a petroleum refinery in Texas after critical barrier failures. However, we should also consider the role of the SPI portfolio as part of the overall SPI quality assessment in situations where several indicators are tracked. As already stated, we find it insufficient to consider quality without making assessments on what influence the other portfolio indicators

have. This is a main aspect of 'relevancy'. In the discussion in Section 6, we address how the inclusion of other indicators matters for the SPI usefulness. But first we present and discuss fundamentally how to develop the SPI portfolio.

4. How to build an indicator portfolio with adequate quality

The management of SPIs involved understanding the results collected from the individual indicators. This requires some structured way that allows the decision-makers to achieve appropriate balance of the indicators included. The use of balanced scorecards is one way. When establishing this structure, again, focus should be away from the distinction and variety of aspects (spread) covered by the indicators, and rather on, as Øien et al. (2011) also argue, in a safety context, how to achieve a useful collection or portfolio of indicators. We refer also to the discussion on the use of leading safety indicators in Leveson (2015).

5.1 Identifying candidates for the SPI portfolio

The starting point for selecting SPIs, is to clarify the safety targets and objectives beyond the barrier requirements. The targets and objectives should be framed for the relevant context, such that the appropriate level of detail and information support needs for decision-making is reflected. The aim is to achieve a set of SPI that can express a broad spectrum of performance, for management to make safety-informed decisions. The SPI candidates are typically referring to failure information, and many are linked to barrier reliability and maintenance area. Such information is typically business sensitive in general, as having barrier failures can have a significant effect on business value. Hence, the indicators are sometimes labelled as key performance indicators or safety key performance indicators, as in e.g. Bellamy and Sol (2012). Several of these are described in the ISO standard on reliability and maintenance data collection and exchange, ISO 14224 (2016), which recommends that the key performance indicators are aligned to the organisation's objectives for the facility (or operations), and that improvements are identified and implemented in order to achieve the organisation's planned objectives. It is then appropriate that the indicator portfolio reflect targets and objectives at different levels, such that they cover various levels of the organisation when aligned with other performance indicator selected for different groups of equipment, systems or personnel. This is not an activity driven by the analyst or decision-maker but rather a coordinated activity of stakeholders, including managers and discipline experts, whose opinions all in some way should be captured in the assessment of the alternative measures and their effects and importance.

The task of selecting amongst SPI candidates involves a structured prioritization of which are the important performance aspects. When focus is on barrier performance, there is usually not many failure events occurring. Hence, hence it is clearly fruitful to map also other candidate types. The candidates normally cover a range of both leading (activities) and lagging (outcome) indicators, and diagnostic indicators. The above-mentioned ISO 14224 (2016) provides a list of 34 key performance indicators which are applicable within the reliability and maintenance area. Bellamy and Sol (2012) present an extensive review on SPIs related to barrier management, and in the review

go through relevant candidates. Beyond the typical candidates, where in addition, companies also develop specific candidates suited to their needs. It is a quite complex landscape. However, a key is to identify how the safety or barrier performance may be expressed and to link it to the use of the information. There is overall a large amount of literature discussing the appropriateness of performance indicators, particularly the leading ones (Badawy et al. 2016; Swuste et al. 2016). It illustrates how challenging it can be to select amongst the leading indicator candidates. See also discussion in Bellamy and Sol (2012).

A characteristic of the SPIs is the explicit link to safety performance. Many would perhaps characterise them as 'appealing' due to the understandable, simple, and compressed way key safety information is communicated. The SPIs comprise key safety information. Hence, it is not surprising that there is a strong link to the use of risk acceptance criteria (RAC; see e.g. Hokstad et al. (2004) and Aven and Vinnem (2005)). These may also be labelled as safety acceptance criteria, but risk being the broader umbrella. The RACs indicate some aspect of performance related to risk. The different measures used in the process industries for comparison against some RAC can then be considered as a larger set compared with the safety acceptance criteria, which for example does not cover possible cost consequences. Nevertheless, the use and definition of these criteria as part of the objectives and safety targets, is often found as the basis for the selection of appropriate SPIs. For example, an indicator may be selected to assess and evaluate against some defined acceptable criteria.

Focus when addressing the quality of a specific indicator part of a SPI portfolio, is on its value. Without adding value, the information has minimal contribution or is misleading in decision-making and is obviously not considered very useful. For example, SPIs measuring 'wrong things', such as indicators with no 'path' to credible accident events, or is having significant uncertainty, should be avoided. The consideration is closely linked with traditional value of information assessment (Bratvold et al. 2007; Bjørnsen et al. 2019), analysing and evaluating to what degree the information (here the indicator information as part of the SPI portfolio), has a significant influence on the decision-making. In practise, this is achieved by the indicator having a safety role not already covered by other indicators in the SPI portfolio, for example, by identifying safety or barrier status and trends, and calling for actions. It can be claimed that the indicators should be 'action-guiding'.

As mentioned, selecting amongst SPI candidates involves a structured prioritization. One alternative, which may be used as basis for ranking the candidates for evaluation of which is the more useful, is the use of a multi-criteria analysis. An example is the traditional 'analytical hierarchy process' (Saaty 1980). Such an analysis is presented by Elhuni and Ahmad (2017) and used to assess 14 different key performance indicators considered for an oil and gas company in Libya. Such a prioritisation can be fruitful to identify whether there are candidates with low value. However, despite there are several challenges associated with having large SPI portfolios, as discussed in Parida and Chattopadhyay (2007), there could be good reasons for including many indicators. For example, the operations having many safety facets. In principle, there are no restrictions regarding how many SPIs should be included, as long as the contribution is good. Companies should select the set of SPI candidates that are best suited to their safety objectives and targets. The main principle

is that the SPIs combined are contributing with useful information. Obviously then, companies need flexibility as there is not a one solution that fits all. For example, there could be different designs making equipment failures more or less severe, making a big difference for management of the barrier elements across the companies. Target and objectives may be different, as well as digital tools for handling the SPI portfolio; all influencing the portfolio setup. Besides, inside the company there are likely to be sub-organisations with different safety targets and objectives. This giving root to sub-organisations selecting a set of indicators best suited to their needs.

5.2 Combining information from the selected SPIs

After identifying individual SPIs with adequate quality, next step is to combine these into an appropriate SPI portfolio, i.e. selecting candidates for a new portfolio or considering candidates to complement an existing one. The challenge is to develop a quality portfolio that is aligned with intended or planned use, as well as targets and objectives. However, this is far from a simple task. A set of SPI candidates are identified, but it is not obvious how to then identify combinations of these giving basis for good decisions, or whether the possible combinations are able to completely cover the safety information needs with respect to the company's safety and barrier management. There is a need to see beyond the individual indicators and understand how they work together, i.e. 'coherence'.

As indicated already, there are different ways of combining the SPIs, but also different ways to visualise or communicate the portfolio. There has also been some development over time, where digital tools are increasingly important for the portfolio management. The typically tools are digital scorecards, dashboards, and analytic reports.

The digital tools allow for presentation of multiple attributes, where the digitalisation could make it simpler to identify scores for attributes linked specifically to safety. For example, it is possible to add colour coding (e.g. red, yellow and green) to highlight the ones having or should be given higher priority, and also adding information about uncertainty related to the individual attributes. These basically list the scores given for each attribute. But there are also other ways. The information could, as some prefer, into one score, making it easier to conclude based on the results. Another way is to restrict the portfolio to a minimum and low number of SPIs. The challenge is then to select the few ones that can present the key safety information needed. This makes it again difficult to achieve the bigger safety picture and could provide misleading information. To some degree, it depends on the type of business and company considered. But, overall, the practise of having a portfolio with one or only a few SPIs, would not have the simplicity and communicative abilities typically characterising the use of SPIs and key performance indicators in general.

To achieve a SPI portfolio with quality, several aspects should be taken into consideration. Despite having clarified safety target and objectives, and selected indicators according to these, everything is not in place. For example, there is the always reoccurring issue of cost versus benefits. There is usually a cost of acquiring the SPI information, which should be seen in relation the benefits. There is also an issue of uncertainty, i.e. to what extent the information provided is credible. Further, the

portfolio should cover a broad spectre of performance aspects but without repeating information for similar aspects. Obviously, key aspects considered as useful to have information about, should be included. However, the challenge is often to make sure that key ones are not missed or which ones to leave out.

Above mentioned the possibility of sub-organisations having different safety targets and objectives. Quite often, this is the situation, where there could be conflicting drivers across the organisation. For example, there could be parts of the organisation focusing on solely on maintenance activities, where safety focus and use of various performance indicators, including general key performance indicators, relates to maintenance activities. These could be contradictory when compared with parts dealing with for example on-site process safety. However, for the company overall, assuming the SPIs being consistent with the business and safety strategy of the company, they could both be appropriate. For example, the indicator 'total maintenance cost' (for a given period) is from the maintenance part's side obviously a number that should be minimised. Seen from an overall company perspective, however, also other aspects that should be part of the consideration. It might be unreasonable to lower the maintenance costs if this leads to significant reduction in reliability and thus higher accident risk. The decision on whether to increase maintenance costs, depends on the reliability and overall safety benefits.

There is an increasing use of digital tools in safety management. There are extensive software applications assisting the analytic tasks and presentation of results. Some of these allow for user friendly interfaces and simplified understandings of safety, however, there is also the challenge that these become sort of 'black boxes' hiding key information, particularly when automated techniques are applied. Nevertheless, such tools allow for also use of machine learning techniques that can be used to identify risk and safety trends (see. e.g. Bansal et al. 2020), making it possible to identify patterns not else recognisable. Another point is that the use of digital tools makes it possible to reach out and spread information, make it available and useful, in a more effective way. For example, an automated dashboard for SPI tracking could allow for 'real-time' updates. Related to the maintenance activities, such use is associated to 'maintenance excellence' status, meaning that reliability and maintenance performance should be aligned at a strategic level and the performance should be communicated in an appropriate way. An industrial example is Maersk Oil Qatar's efforts to achieve such status, where the use of effective communication means to present performance aspects is seen as very important (Smart and Blakey 2014). Another example is the 'maintenance excellence' programme built in Shell, for which Jansen (2015) claims that: A "computerized maintenance management system (CMMS) should be the backbone for work management and performance improvement", stating the importance of bridging performance indicators and the digital tools.

Finally, before turning to a discussion on use of the revised SMART criteria, we acknowledge that the safety situation and associated targets and objectives are not a static matter. This is something that could change, for example due to measures implemented or requirement for more robust designs. The indicators should reflect a situation of targets and objectives being dynamically redefined. There is a need to continuously review whether the basis for the SPI construction holds, and if needed, to

update the SPI portfolio and reconsider how to use the information, as argued in Øien et al. (2011).

5. Using the SMART criteria to assess the indicator quality from a portfolio perspective

Including an assessment of the SPI portfolio complicates the quality assessment. It becomes more complex in nature, partly because the other SPIs might not be sufficiently clear on the spectrum of use (decision-making situations) and usefulness. It is challenging when having to capture a mix of attributes. There is also aspects of confidence and resources needed to perform the quality assessment, not always in place. These are typical challenges, when using the information in safety or barrier decision-making, addressed in the 'managerial review and judgement'. There are likely to be situations where the benefits or usefulness of the SPIs can be questioned, for example because there is not collected a specific type of data or there not being enough history to conclude with certainty. It is not the intention that the SPI should support all types of safety or barrier decisions. The SPIs provide information giving insights into safety or barrier 'performance' and business 'health.' They should not be seen as available 'decision-making instruments.' A fundamental principle of the 'managerial review and judgement' activity is that it is the responsibility of the decision-makers to consider what information is appropriate and how to use this in decision-making situations. It is an activity where management considers and weights the different concerns, including interests from various stakeholders (internal and external). Again, the use of the SPI portfolio is a dynamic process; being strongly influenced by the context and stakeholders involved. As such, quality is interpreted as a relative matter. It is a result of those involved, which obviously could make it challenging to assess the SPI usefulness.

In the same way as for the assessment of individual SPIs (outlined and discussed in Section 4), the assessment should be performed with respect to safety targets, objectives, and usefulness, also when taking a portfolio perspective. Focus is still on achieving or contributing to improved decision quality. However, this requires the safety targets and objectives to be clearly defined. Otherwise it is difficult to evaluate whether the SPIs are useful or needed. Next, we will discuss the use of the modified SMART criteria for the quality assessment.

As in Section 4, we start with the 'specificity' (S) criterion. There is no doubt in this quality aspect being relevant. But focus is slightly different. When considering this aspect from a portfolio perspective, 'specificity' extends beyond the specific SPI in focus and covers also the other SPIs in the portfolio. Hence, for this criterion to be satisfied, there should be precise information on which other SPIs are included, besides, it should be clearly stated how the SPIs are combined in the portfolio and how the information is expressed (pictured). For example, information on SPI ranking or priority should be available, to define clearly the SPI roles in the portfolio and how they compare for decision-making purposes. Such specificity makes it simple to understand the purpose of the SPI amongst the other SPIs, and how it can be used in barrier and safety management.

Continuing with the next 'SMART' criteria, we have then 'manageability' (M). The point of this criterion is to assess whether, when combined with the full portfolio, there are challenges restricting management of the safety aspects addressed by the SPI in focus. For example, there could be a situation where real safety benefits cannot be achieved as this would 'steal' resources from other and more critical safety activities. In other words, it means that it is in principle manageable, but not in practise. Assessment of the specific SPI as part of a defined portfolio addresses the ability to manage the SPI in focus seen from a systems perspective. The point is not to find a suitable way of managing the portfolio but, rather, to identify what is the room for improvement of the considered safety aspect, given a more relevant context of the current situation. Prioritisation of resources and the SPI role could clearly make a difference for this ability. However, this would be a managerial task and for the quality assessment, the conclusion would always be that it is possible to manage safety or barrier performance in some way. As for the conclusion that a relevant SPI is always manageable from an individual SPI perspective, although the actions are not identified specifically, this will also be the situation when taking account also the other indicators part of the portfolio. As the 'M' criterion adds no value to the quality assessment, it would be better, for the safety indictor context, to shorten the acronym to 'STAR'.

The 'achievability' (A) criterion follows up on the managerial (the decision-maker's) ability to take actions. Again, there is a need to consider that the management could be facing several conflicting safety targets and objectives being addressed by different SPIs in the portfolio. Basically, what we need to assess is, whether it is possible to achieve SPI results with adequate quality when combined with the portfolio of SPIs. This implicitly relates to the way the results are integrated in the format used to compile the SPI results, for example using digital scorecards. As for the 'manageability' criterion, the conclusion reached for the 'achievability', is likely to be the same for both the individual SPI quality assessment and for the portfolio SPI assessment. Not necessarily, but usually this will be the situation.

The 'relevancy' (R) criterion is perhaps the one attracting most attention. At least in literature because of the strong link to 'why' the company should spend resources on it. The assessment of this covers the ability to make good safety decision and take appropriate actions using the information from the available multi-attribute indicator portfolio (Wood 2016; Longhi et al. 2015). Quality, then, comes from whether the decision-makers are able to make safety-informed decisions showing a positive effect on the performance aspect considered, which are based on the information provided by the SPI(s), and would not have been made otherwise. From an individual SPI perspective, this criterion is already considered; however, there is again the possibility that changes to the specific indicator, could have an overall negative effect on safety performance when also other SPIs are considered, for instance, a conflict of interest could exist between the SPIs. Hence, we could have a situation where it is possible to manage the SPI results over time, but where the benefits of the specific indicator are marginal or disproportionate compared with the benefits obtained from the portfolio. For example, it could be that the safety aspect in focus is already covered, or partly covered, by another SPI.

Related to information needs in various management situations, there is often assumed a relationship between management and measurement in line with the saying,

that: “you cannot manage if you don’t measure”. It is about having enough information to make good decisions and to have some level of control over the situation. However, related to performance measurements, associated analysis and decision-making, we often find the opposite to be just as relevant: “what you measure is what you manage”. The information and knowledge obtained from the SPIs could assist in establishing a safety picture describing the current situation, but clearly this information may also have strong influence on which safety aspects are given priority. Say the company has adopted a vision zero principle, i.e. defining a safety target and vision of zero critical personnel injuries and fatalities. Then, based on this, SPI could be developed to track the number of events occurring and use this information to guide further improvements. However, management guided from this SPI, despite being suited to this objective, could fail to be rational if it is compared with traditional cost-benefit principles and overall safety benefits, i.e. seen from a system perspective.

‘Time-based’ (T), being the final criterion, considers whether the defined measurement period is appropriate, when used in combination with the other SPIs. An argument for considering a different period, is that similar information is already provided by another SPI. It could be appropriate to make changes, to make the portfolio cover the complete range of past, present and future performance. In a similar way, the portfolio should cover target and objectives of both operational and strategic character, i.e. short-term and long-term, respectively.

6. Use of the modified criteria (STAR) to assess a safety performance indicator in a refinery scenario

In this section, we will consider a safety performance indicator called ‘Dangerous fluid overfilling events.’ This indicator could be attractive to process industries and is obviously related to safety. Monitoring of trends and level of occurrence can potentially add value by identifying undesired safety and business performance. According to Chang et al. (2006), overfilling events cause a loss of containment and claim it to be the most frequent cause of operational error for tank accidents. Overfill hazard also depends on the type of vessel and associated upstream/downstream equipment (Summers and Hearn 2010). There are differences in the fluid overfilling for a process vessel vs. storage tank. The distinction between the two types of equipment is clarified e.g. in ISO 14224 (2016), which details taxonomy classification for reliability data collection within the process industries. Both are listed as a mechanical equipment category and show that storage tanks and pressure vessels contain similar subunits. Further, this international standard clarifies that storage tanks include atmospheric tank and low-pressure tanks, while the pressure vessels could handle gas or other fluids with higher pressure.

When a process vessel starts overfilling, usually the fluid outlet of the vessel (e.g. relief system, control valves, etc.) is blocked during the fluid inflow. In a storage tank, an unchecked rate of inflow accumulates large amount of fluid such that it exceeds the tank’s maximum holding capacity. After a processing vessel is overfilled, the excess liquid unintentionally enters the outlets designed for gas phase or is passed to the downstream equipment that is not designed to receive it (Summers and Hearn 2010). An overfilled storage tank releases excess liquid through its vents or fails under excess

structural pressure (Waite 2013). While overfilling may materialize somewhat differently in both vessel types, the overfilling event equally threatens the operations' safety in both. We will investigate the SPI's usefulness in tracking both of these two different conditions.

A main example of an overfilled process vessel is the major accident that occurred at a refinery in Texas City March 2005, where 15 people were killed, 180 was injured, in addition to major structural and financial consequences, from fires and explosions caused by overfilling.

We will use the Texas City refinery scenario, and more specifically the 'Isomerization unit' (ISOM), which was the source of the accident, as basis for the discussion regarding the quality of the overfilling indicator for process vessel. The refinery had previously ignored a past trend of minor-overfilling events assuming it not to pose any hazard, but by that repeatedly removing a key safety barrier. This allows for a discussion on the indicator usefulness from a realistic safety management view, both from individual and from a portfolio perspective. This accident is particularly relevant to assess if the information conveyed by the chosen indicator can help in determining why the combination of safety barriers did not function properly. But before discussing this, we will give a brief and simplified description of the system and what happened. For a more detailed description, we refer to e.g. Saleh et al. (2014b), Hopkins (2008) and CSB (2005).

7.1 Key barriers related to operation of the Texas City refinery ISOM - and what went wrong

Figure 1 shows a simplified layout of the main components of the ISOM unit at the refinery. Liquid raffinate flows into a tank or vessel called the 'raffinate splitter tower', being the centre of the unit. The vessel is a about 50 meters high and is where heaver raffinate is separated, sending parts of the raffinate to storage. The tower has sight glass and a level transmitter (sensor) measuring the fluid level in the range 1.5-2.7m above bottom, In addition two separate level alarms are installed to indicate high liquid level. The first alarm is programmed to sound when the transmitter's reading reached 2.3m in the tower. The second alarm is a redundant high-level switch that sounds at 2.4m fluid level, independent of the level transmitter. The 'level alarm low' is another low-level redundant alarm. From the top of the tower, lighter raffinate flows out and into an air-cooled condenser, from where it is sent either for storage or routed back to the tower. To effectively deal with potential high level or over-pressurisation, upset operations or shutdowns, three parallel safety relief valves are installed. The outlet of this line leads to the disposal system, i.e. 'blowdown drum and stack' and 'sewer'. Liquids will then end up at the bottom while, the gases escapes to air through the vent stack on the top. The liquids then discharge into the unit's sewer by opening a manual block valve. The blowdown drum had level sight glass for level monitoring and a high-level alarm to alert operators when liquid was close to flowing above a certain level (i.e. seal leg of the gooseneck pipe opening to the drain).

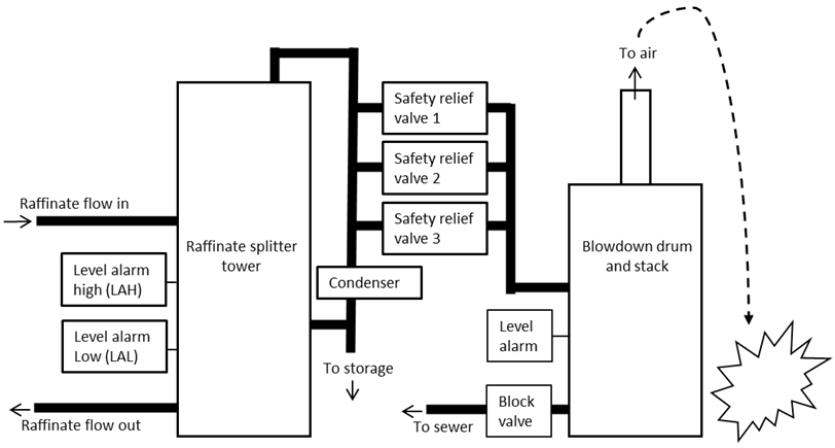

Figure 1 ISOM unit – Simplified layout

On the morning of the accident, when starting up, the lead operator as usual started pumping raffinate into the splitter tower. According to plant operators' common practise, although a violation of formal start-up procedure that calls to maintain 50 percent transmitter reading level, the raffinate was pumped in to a 99 percent transmitter level. As the tower was filling up beyond the set point of the high-level alarms, only one high-level alarm triggered but was ignored. The redundant high alarm did not sound. The level sight glass was not readable and not used. The operator was unaware and interpreted the transmitter's 99% (maximum) reading as the correct level measurement. In reality, the tower had filled 1.2m above the top level of the transmitter's range. After the raffinate section equipment were filled up, the start-up procedure and raffinate feed were suspended. Against procedure, the operator also closed a control valve instead of leaving this in 'automatic' mode. Before, leaving, the night shift operator left incomplete information in the logbook about what steps were taken and what was to be done in the next shift.

Consequently, the next shiftoperator did not receive proper information about the unit's status. Due to the miscommunication, the new operator was unaware that the raffinate equipment was filled during the previous shift. The unit supervisors were also unaware of these conditions. Next morning, due to miscommunication, the supervisors instructed the operations crew to restart the raffinate feed into the tower. The operators controlling the heavy and light raffinate products were uncoordinated. They did not receive clear instructions about the feed and product routing prior to start-up. They made false assumptions about the conditions and ended up closing both the level control valves (outlets) while the tower was continuously being fed. The splitter tower was unknowingly being overfilled now as it had no output discharge or real level monitoring. At the time when the operator raised the temperature of raffinate in the splitter tower, the level transmitter falsely displayed 2.6m fluid level (investigation reports indicate the level was in fact around 20m and increasing). Some hours later, the overfilling was still unknown to the operators, who still misinterpreted the system

behaviour. It ultimately led to raffinate liquid overflowing to the overhead line, through the safety relief valves and into the blowdown drum. And, without the operators knowing it, the blowdown drum filled up (the level alarm was out) and raffinate was shot out through the vent stack into the air. At the ground, vapor ignited, most likely from a nearby idling pickup truck, causing a massive explosion. Clearly, a series of safety barriers for preventing dangerous fluid overfilling failed on the way; see below.

Organisational safety barriers:
Operators and staff controlling the ISOM unit, was inadequate. They were overworked and poorly trained to handle the abnormal start-up conditions leading to fluid overfilling. The control room was ill-equipped to display the net fluid flow rate or to detect overfilling events. There were insufficient instructions to the operators regarding how to consider the incoming-outgoing raffinate flow readings being essential for overfilling situations, and particularly relevant during start-ups. The company to large extent failed in enforcing formal procedure (e.g. inadequate shift handover, poor recording quality in logbooks, lack of technical supervision, no instrumentation checks pre-start-up). There was also a history of budget restrictions delaying maintenance activity. Overall, the organisational barriers of promoting a safety culture, providing adequate safety preparedness and operator training were largely failing.

Human safety barriers:
The operators frequently ignored alarms at the unit and violated start-up procedures. Besides, there was a lack of communication among the shift operators and management in conveying critical decisions, such as the decision not to follow formal start-up procedure. The human barriers of skill, training and experience failed to detect the overfilling incident and containing it early.

Technical safety barriers:
The instruments were poorly calibrated or not designed to detect the actual fluid level. The sight glass needed replacement, and the high-level alarms failed to activate, both at the tower and at the blowdown drum. The failure of the level alarms meant that the operator received no warning about the critical fluid level nor that it was exceeding detectable level. The sight glasses were both able to only display fluid level in a small range and was poorly designed. The tower's level transmitter was unreliable (e.g. it wrongly displayed fluid below 100% level (2.4m) when the fluid was overfilling in the tower). Since, the operators trusted this instrument's reliability, they could not detect that the fluid had surpassed the transmitter's recognisable range and was escalating into an overfilling event. The ISOM unit discharged the flammable raffinate into a sewer, however, as per the industry guidelines this was an unsafe practice to prevent blowdown drum overfilling. The system lacked screening points of fluid flow in and out of the equipment. These weak barriers of instrumentations and alarm systems in combination failed to detect the overfilling incident, making the overfilling go undetected, up to the explosion.

7.2 Quality assessment of the safety performance indicator: Dangerous fluid level events

The event described above represents only one event. What we are questioning is, whether it is useful to record the number of such events as a key indicator of safety performance. Below, we will assess the quality of the dangerous fluid overfilling event indicator using the modified SMART criteria, now referred to as the 'STAR criteria'. We will do this both individually and at a portfolio level. For the portfolio level, we adopt relevant SPIs suggested by the CSB accident investigation report (CSB 2007). Note that the adopted list of indicators is selected for the purpose of the discussion in this article, is not meant to be neither exhaustive nor fully representative of any real portfolio of SPIs tracked by the current facility management. There are obviously other relevant candidates not included. The portfolio consists of the following six indicators:

1. Personal fatality and injury rate
2. Days away from work
3. Hazardous material release events
4. Dangerous fluid overfilling events
5. Raffinate pressure indicator
6. Raffinate level indicator

We maintain that this portfolio is dedicated to managing the overall safety performance at the ISOM unit of the refinery. The aim is to use information from these SPIs to manage safety performance and avoid accidents is the future. The discussion regarding the usefulness of the indicator, i.e. 'dangerous fluid level events', is given within this frame.

S - Specificity

To satisfy for 'specificity' the indicator should be defined appropriately. In process industries, vessel 'overfill' is given a comprehensible and specific definition in API 2350 (2012), as the point when the product inside a tank rises to the critical high level i.e. the highest level in the tank that the product can reach without detrimental impact, e.g. product overflow or tank damage (Roos and Myers 2015). The important term being 'critical high level'. The API 2350 (2012) calls this the 'overfill level', which is the maximum fill-level of a product within a tank measured from the gauging reference point, above which level any additional product will overfill and spill out of the tank. Staying consistent with the standard, all combustible and flammable liquids are under focus because their mismanagement poses a higher safety risk. We refer to these as 'dangerous fluids' or simply 'fluids' in this context.

An overfilling event is thus an event where some vessel is filled with a fluid quantity that is more than the maximum capacity. All situations where the vessel is over-filled, or the operator losing fluid level control to cause spillage or tank damage, should be recorded for the indicator in focus. This allows for making trends over fixed intervals, e.g. annually.

Considering the other SPIs in the portfolio, none of these conflict with the dangerous fluid level indicator. These specifically addresses other safety aspects. Indicators 1 and 2 are mainly concerned with on-site personnel. They are type of indicators tracking occupational safety and standards of the working environment. They provide limited information relevant for process safety. Indicators 3 and 4 are both

lagging indicators, recording past safety performance. As material release is not seen as relevant to the ISOM unit, the two should not be overlapping. Indicators 5 and 6 are leading SPI related to process health in the splitter tower, managed in real-time. These two reflect the current system state (pressure and fluid level) and are used by operators to make short-term control decisions. From a portfolio perspective, the SPIs are sufficiently specific on which of the SPIs that are to be prioritized for short-term vs. long-term decisions and to be used to track business and safety goal achievement. We conclude that the 'S' criterion is sufficiently satisfied from an individual and portfolio perspective.

T - Time-based

The indicators should show trend for reasonable timeframe. Hale (2009) claims this motivates appropriate safety actions. The overfill indicator counts the events occurring during the period. The question is whether, for the period considered, there are enough events to produce a meaningful rate (Hopkins 2009). If this period is too short, a lack of events could be mistaken for a sound barrier performance. On the other hand, if long time goes by without any evet being recorded, Hopkins (2009) argues that it is not possible to compute a meaningful annual rate, nor is it possible to conclude from one occurrence that safety is deteriorating. The time interval considered should be sufficiently long to capture the system's safety status before and after a safety barrier is deployed so that performance comparison is meaningful. According to API Recommended Practise 754 (2016), recommends reporting indicators by current year process-safety-event count, and a 5-year rolling average on a company and industry level. A 5-year rolling average may perhaps capture a broader spectre of events. Although, by producing the overfilling events with an annual rate, it should be easier to identify outliers, and it should be sufficient to capture a trend.

The SPI portfolio covers a combination of short- and long-term focus. Indicators 3 and 4 is to some extent long-term oriented, by considering achievement of objectives through annual (un-averaged 5-year trend can also be relevant) observation periods, while short-term policy goals are more relevant from indicators 1 and 2. The current system state is observed by indicators 5 and 6, although this information could be of interest also for longer terms, and vice versa for the other indicators. The combination of indicators in the portfolio facilitates observing the operational (process safety) objective achievement and the effect of strategic changes in safety and business policies. From a portfolio perspective, the measurement period is quite flexible and can be changed if required. Overall, when the overfilling indicator is recorded for an annual interval, it sufficiently satisfies the 'T' criteria.

A - Achievability

Achievability refers to the ability to produce accurate information. Which can be challenged by uncertainty regarding the number of events recorded. Basically, the number of events come from recording the instances when the level transmitter show 'overfill/high-level' or by other observation or alarm. However, identifying and segregating an overfill-event is not that straightforward. There are several reasons. According to Summer and Hearn (2010), operators rarely track the fluid levels directly because a 'high-level' event is an overfilling hazard only when the liquid begins

flowing to equipment such not designed to receive it. This is when the overfill event can cause loss of containment, as in the Texas City refinery accident. An overfill may occur in a few minutes or may take several hours. As the event propagation time can vary significantly, its classification becomes uncertain, raising data credibility issues. Besides, the cause of a fluid 'high-level' event depends on the operation mode (i.e. start-up, normal or abnormal) as it can influence the amount of fluid accumulated (Summers and Hearn 2010). For example, a higher level under abnormal conditions could be intentional and necessary to prevent equipment stresses. Making it unclear whether the overfilling event is to be recorded if it is assumed as non-hazardous.

The indicator does not separate between hazardous events and inconsequential overfilling events. Although it may be relevant to analysis, information about operating levels, operational modes, safe-fill levels, etc. are ignored when collecting data for the indicator. In the Texas City accident, the operators accepted a high-level against the prescribed start-up procedure. This was due to a lack of information on the safe-fill limit and the level transmitter displaying a limited operating fluid-level range. But assuming the raffinate level in the vessel to be below the high level. Ignoring the role of the measuring device, crucial for this indicator, may produce uncertain and misleading results. A limited-range or unreliable transmitter can result in failure to identify overfilling events in some situations, and perhaps include non-events in others.

The key is to collect credible information about the barrier performance. But as claimed in Saleh et al. (2014), the design configuration and equipment limitations, challenge the ability to collect such information with high credibility. Basically, the uncertainty is significant, making the indicator subject to phenomenon understanding, as it is necessary to assess this uncertainty. A peer-group trend comparison would be risking using misrepresented data. Such an indicator could motivate mistargeted actions, clearly not being in line with the safety objectives. Consequently, on an individual basis, the indicator does not satisfy the 'A' criterion within the current design solution.

From a portfolio perspective, it can be discussed how the overfilling indicator is linked to the collection of data to the 'raffinate level' indicator (Indicator 6). If the quality of any of these are good, then it can be assumed that an overfilling will be detected. However, for the system considered this is not fully the situation. However, this relates also to budget restrictions and priorities, as it is possible to implement a way more credible level monitoring system for the vessel. Which would make the overfilling indicator satisfying the 'A' criterion on a portfolio level.

R - Relevancy

Relevancy is perhaps the most important criterion, indicating why to use the indicator. Fluid overfilling is one of the most commonly occurring instances causing near-misses and loss of containment accidents. In the chemical and petrochemical industries, the loss of containment of a hazardous substance has been the main factor in several major incidents (Collins and Keeley, 2003). It is acknowledged that fluid overflow events pose risk and should be given attention sooner. It is a way of measuring the effectiveness of the control upon which the risk control system relies, which is a key according to Hopkins (2008). A high fraction of historic overfilling events analysed by Chang and Lin (2006) ended with fires and explosions, potentially causing major

accidents. Chuka et al. (2016) presents a variety of consequences related to containment loss in the process industries.

Dangerous overfilling events as a lagging indicator can be criticised for not giving early warnings, requiring looking further back in the causal chain, at the underlying causes and the condition of the factors that leads to accidents (Øien et al, 2011). However, Hopkins (2009) argues that in situations when hazardous events are occurring frequently enough to produce a meaningful rate, the rate can be used to measure and manage safety. If the events are rare, it is not that relevant, and we must look to more frequently occurring precursor events to be able to measure safety (Hopkins, 2009). For the refinery scenario we assume there is a significant number of events. Historic data showed that the processing tower experienced dramatic swings in liquid level during 18 of the 19 previous start-ups and had numerous tower overfilling incidents (CSB, 2005). Between 1995 and 2005, the refinery had four other serious releases from the ISOM unit blowdown drum that were unignited ground-level vapor clouds (Baker Panel report 2007).

Overfilling events typically follow a complex escalation path, aided by hidden latent failures at different operational stages, which is only implicitly revealed by the overfilling indicator. It does not give the analyst any information about what, where and how the overfilling took place. He must find this out by collecting supporting information (or other SPIs) that underlying conditions and safety gaps. In practise, a variety of safety barriers (e.g. human, technical, organisational) can play a role in preventing such events occurring.

From the portfolio perspective, as Indicator 6 provide a different type of information, i.e. on the current condition, the overfill indicator is complementing the portfolio. Neither the hazardous material release indicator provides conflicting information, as the overfilling refers specifically to the vessel safety performance. This makes the two indicators even more relevant when considered together.

The Baker Panel report (2007) concluded that the operating company in a way placed more attention on personal safety compared with process safety; mistakenly seeing improvement of personal injury rates as an indication of acceptable process safety performance at the refinery. From a portfolio perspective, we can assume that the delay in maintenance actions can be attributed to prioritisation of personal safety, promoted by Indicators 1 and 2. It suggests that resources, investments and attention were ‘stolen’ away from maintaining the overfill-prevention barriers, e.g. installing reliable fluid level transmitters and adequate operator training. The potential for overfilling, on a portfolio level, clearly ranked behind personal safety-targets for the management, as visible in the maintenance budget-cuts, degrading infrastructure, and under-staffed operations (Baker Panel report 2007). This in practise challenges the benefits at the portfolio level, but also shows why it is important to include such an indicator.

7.3 Refinery scenario findings

To summarize the overall results of the above STAR criteria quality assessment of the dangerous-fluid overfilling indicator, we find that there is only one criterion that is not satisfied. The assessment and associated discussion conclude that the criteria specificity, time-based and relevancy are all satisfied. Both for the individual and the

portfolio perspectives. However, not the achievability criterion, which fails on both perspectives. Hence, we overall conclude that the indicator in focus is not having adequate quality. This is not to say it cannot be useful, but the achievability obviously challenges this.

7. Assessing the safety performance indicator in a storage tank scenario

Above we discuss the overfilling indicator in relation to the Texas City refinery accident, i.e. for a process vessel context. In this section, a similar event i.e. the Buncefield depots' tank overfilling accident is considered. In this accident, the level measurement device did not display the changing level even though the tank's fluid level was rising. This presents a different use case that can be tracked using the indicator. We will re-assess the SPI using the Buncefield case to determine whether it produces similar results on STAR criteria when focusing on storage tanks. It will provide a broader understanding related to the use of this indicator within the process industry.

8.1 Key barriers related to operation of the Buncefield depot - and what went wrong

Buncefield oil storage & transport depot is a farm of several tanks serving areas in UK, including London. The operating site stored hydrocarbon fuel received via a complex network of three pipelines. It had experienced a devastating explosion and fire in 2005 due to failure of its overfilling protection system.

There were two main safety barriers against tank overfilling. First, an automatic tank gauging system (ATG) that displayed the fuel level on control room screen for the operators to monitor. The ATG also had alarms at 3 succeeding levels (1) 'user high'- set by the supervisor indicating the need for intervention, (2) 'high level'- at a level below the tank's maximum working level, (3) 'high-high level'- at a level above 'high-high' but below IHLS (COMAH 2011). Independent high-level switch (IHLS) was the second barrier set above the ATG alarm levels. Its function was to raise audible alarms when the fuel reached an unintended high level and automatically operated the shut-off valves to stop the fuel supply. IHLS and ATG operated independently of each other to safeguard against tank overfilling. The barriers are illustrated in Figure 2.

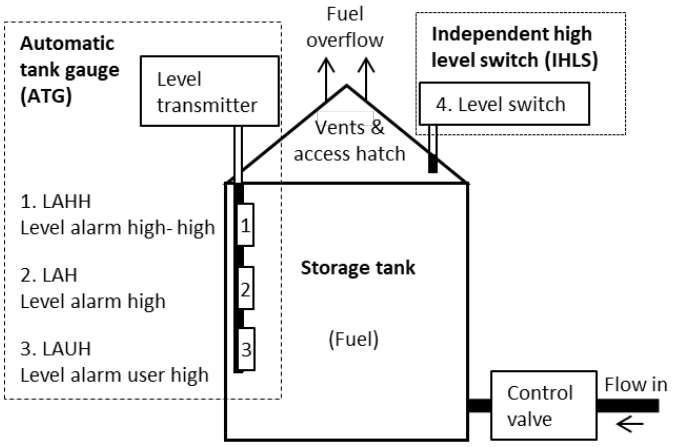

Figure 2 Buncefield storage tank – Simplified layout

On 10 December 2005, a pipeline started delivering fuel to a storage tank at the depot. But unknowingly the level monitoring instrument of the ATG stopped registering the rising fuel level midway of the delivery. The monitor erroneously displayed a 'flatline' (indicating that the tank was no longer filling up) while the fluid continued to be delivered. The ATG alarms, dependent on this level monitor, could not operate since the level reading remained below their corresponding set levels. The tank's first safety barrier against overfilling had failed. The second barrier, IHLS, was also ineffective because those who installed and operated the switch did not fully understand its working; such that the switch was left effectively inoperable after a previous test (COMAH 2011). The inoperable IHLS meant that neither the final alarm alerted the operators about overfilling nor the automatic fuel supply shutdown activated. Tank's maximum fuel capacity was soon exploited, thereafter the excess fuel started spilling from vents in its roof. This exposed fuel formed a white flammable vapor cloud at the site. After an employee noticed the cloud, he raised an alarm and the firewater pumps got initiated. Almost immediately, the vapor cloud ignited with an explosion of high over-pressure. The explosion was followed by a five-day long fire that injured forty people, engulfed twenty fuel tanks, and had widespread environmental consequences. The overfilling incident was important in causing a complete loss of primary containment (i.e. the tank unit). The failure of ATG recognizing the hazard i.e. misleading level monitors and inoperable IHLS were the main cause for the fuel tank overfilling.

8.2 Quality assessment of safety performance indicator: Dangerous fluid level events

The event described above presents a case study of storage tank accident to evaluate the usefulness of recording the overfilling incidents to improve the safety performance. The SPI is already assessed in section 7.2 for its usefulness on the STAR criteria for the case of Texas City refinery's process vessel. Using the results derived from the previous discussion, we reinstate that the overfilling SPI satisfies the

'specificity' and 'time-based' criteria since these qualities are independent of its application.

Next, the 'achievability' of the indicator needs to be examined for storage tanks. As discussed in 7.2, there are uncertainties associated with investigating if a 'high-level' reading indicates an actual overfilling event in the process vessel case. This applies to the storage tanks as well. In the Buncefield case, the tanks had three alarm levels starting from the lowest 'user-level' alarm, raised the need for human intervention incrementally. However, given the poorly specified filling procedures, the Buncefield operators used these alarms subjectively. They underestimated the likelihood of overfilling event by allowing the 'high level' and even 'high-high' level alarms to pass unchecked sometimes (COMAH 2011). The ATG barrier alarms were not being used for performing the intended safety function. The shutdown IHLS barrier was neither properly maintained nor understood clearly. Investigation from the past storage tank accidents commonly point to factors such as poorly maintained hazard measuring devices (alarms and sensors), inconsistently used reporting (logging) system for overfilling incidents, over-worked staff, and lack of data with quality. These factors along with the system complexity and equipment's limitations (refer to section 7.2) add significant uncertainty about the indicator's trend. Therefore, on an individual basis the SPI fails to satisfy the 'A' criterion even for the storage tank application. On a portfolio basis, tank overfilling events can be detected and recorded with the help of other quality indicators.

For the SPI's 'relevance', the consequences of the operation being tracked is important. Fluid filling is the primary operation conducted on a storage tank, often several times every day. Frequent transfer of dangerous fluids warrants monitoring the overfilling events and consequently, its safety barriers' performance. This makes the overfilling SPI particularly relevant for tracking the trend of poorly performed filling operations. Storage tanks are also vulnerable to similar negative consequences of fluid overfilling as discussed for process vessels in 7.2. At the portfolio level, the indicator may receive less or more resource prioritization depending on the management's decision-making principles and risk appetite. In the Buncefield accident, the indicator was ignored by the management and operators alike, as is evident from the investigation report (COMAH 2011). It states that the defect in the tank's level monitor, that had stuck 14 times within three months before the accident, was treated with quick fixes only. The management and staff had underplayed the importance of monitoring key safety trends and later faced the consequences. So, on a standalone as well as portfolio basis, the SPI satisfies the 'R' criterion.

8.3 Fuel tank depot findings

To summarize the STAR criteria assessment of dangerous-fluid overfilling indicator, again only the 'A' criterion is unsatisfied for the storage tank application. The indicator is specific, timely and relevant from individual and portfolio perspectives. This case study provides a broader context for SPI's usefulness in a different context. While the indicator's usefulness can be challenged from the aspect of 'achievability', all the other criterion, especially 'relevance', stands in support for the value it can generate for safety barrier's performance management.

9. Conclusions

The quality of a safety performance indicator relates to the potential use of this to identify safety challenges for the system considered. This by providing information not already being produced by other indicators, and as such it complements the SPI portfolio. Properly defined and understood indicators can give companies confidence that the right things are being managed and tracked (API Recommended Practise 754:2010).

In this article, we discuss the use of SMART criterion for the quality assessment. This covers five basic criteria assumed to be fruitful for a general key performance indicator context. The SMART criteria cover a range of aspects, which we have considered; one by one. Both individually and from a systems (portfolio) perspective. Overall, we find the criteria to be applicable, and should be included for a general assessment of SPI quality, except for the 'M' aspect. This, regardless of whether the letter 'M' refers to 'measurability' or ''manageability'. In either of the criterion is assumed to be covered by the other four. We claim that the 'M' can be effectively removed, for both individual and portfolio assessments. Thus, we suggest to instead, when dealing with indicators related to safety business objectives, to rather adopt the following acronym:

'STAR': 'Specificity' – 'Time-based' – 'Achievability' – 'Relevancy'.

The criteria represented by these four letters are suggested as the basis for assessing SPI quality. To demonstrate the use, we have assessed a potential indicator called: Dangerous fluid overfilling events. The assessment identifies significant uncertainty related to producing accurate SPI numbers, and the SPI thus fails for the 'achievability' criterion. The uncertainty, although the indicator is found to be both specific, time-based and highly relevant, challenges the usefulness. Without providing sufficient accuracy it is difficult to use it for informed decision-making and safety business management. However, by using such an indicator there is a chance that one could have seen the 'top of the iceberg' and acted on that to improve the barriers. Besides, as the indicator is seen as highly relevant, this could motivate actions to make it achievable. Overfilling clearly represents a risk, as demonstrated by the 2005 Texas City refinery and the Buncefield depot accident.

The dangerous fluid overfilling indicator assessed is associated with a common safety concern among petroleum but also petrochemical and natural gas industries, as well as nuclear, basically any industry that handles hazardous fluids, i.e. the risk of loss of containment. However, the discussion about quality and usefulness is restricted to the frame and specific system considered and is thus not automatically transferrable to any other process system. Even for other refineries the conclusion could be different. Nevertheless, the use of the STAR criteria is applicable to basically any industry and system being safety oriented.

Acknowledgements

The authors are grateful to the anonymous reviewers for their useful comments and suggestions to the original version of this article.

References

1. API Recommended Practise 754 (2010). Process Safety Performance Indicators for the Refining and Petrochemical Industries, First Edition.
2. API Recommended Practise 754 (2016). Process Safety Performance Indicators for the Refining and Petrochemical Industries. Second Edition.
3. API 2350 (2012). Overfill Protection for Storage Tanks in Petroleum Facilities, Fourth edition. Petroleum Institute (API).
4. Aven, T. and Vinnem, J. E. (2005). On the use of risk acceptance criteria in the offshore oil and gas industry. *Reliability Engineering and System Safety*; 90: 15-24.
5. Badawy, M., Abd El-Aziz, A. A., Idress, A. M., Hefny, H. and Hossam, S. (2016), A survey on exploring key performance indicators. *Future Computing and Informatics Journal*; 1(1-2): 47-52.
6. Bansal, S., Saadalla, N., Selvik, J.T. and Abrahamsen, E.B. (2020). Development of a bivariate machine-learning approach for decision-support in offshore drilling operations. In: Proceeding of the 2020 European Safety and Reliability and Conference, Venice, Italy, November 1-6, 2020.
7. Bellamy, L. and Sol, V.M. (2012). A literature review on safety performance indicators supporting the control of major hazards". RIVM Report 620089001/2012. National Institute for Public health and the Environment.
8. Better Regulation Task Force. (2000). *Principles of Good Regulation*. London: Cabinet Office.
9. Bjørnsen, K., Selvik, J.T. and Aven, T. (2019). A semi-quantitative assessment process for improved use of the expected value of information measure in safety management. *Reliability Engineering and System Safety;* 188: 494-502.
10. Bratvold, R. B., and Begg, S. H. (2010). *Making Good Decisions.* Society of Petroleum Engineers, Richardson, TX, USA.
11. Bratvold, R. B., Bickel, J. E. and Lohne, H. P. (2007). Value of information in the oil and gas industry: Past, present and future. SPE paper 110378, presented at the SPE Annual Technical Conference, Anaheim, CA, USA.
12. Chang, J. and Lin, C.-C. (2006). A study of storage tank accidents. *Journal of Loss Prevention in the Process Industries;* 19: 51-59.
13. Chuka, C.E., Freedom I.H., Anthony, U. (2016). Risk Assessment and Consequence Evaluation of Loss of Containment in Process industries. *International Journal of Modern Studies in Mechanical Engineering*; 2(1): 14-28.
14. Collins, A. and Keeley, D. (2003). Loss of containment incident analysis. Health & Safety Laboratory.
15. CSB – Chemical Safety and hazard investigation Board, U. S. (2007). Investigation Report, Refinery Explosion and Fire, BP-Texas City, Texas, March 23, 2005. Report no. 2005-04-I-TX. CSB.
16. COMAH- Control of Major Accident Hazards. (2011). Buncefield: Why did it happen? Health and Safety Executive UK. [https://www.hse.gov.uk/comah/buncefield/buncefield-report.pdf]
17. Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review*; 70: 35-36.
18. Elhuni, R. M. and Ahmad M. M. (2017). Key Performance Indicators for sustainable production evaluation in oil and gas sector. *Procedia Manufacturing*; 11: 718-724.

19. EN 15341. (2007). Maintenance - Maintenance Key Performance Indicators. European Standards (EN).
20. Galligan, F., Barry, T., Crawford, D., Howe, D., Maskery, C., Ruston, A. and Spence, J. (2000). Advanced PE for Edexcel Student Book. Oxford, UK: Heinemann Educational Publishers.
21. Hale, A. (2009). Editorial special issue on process safety indicators. *Safety Science*; 47: 459.
22. Hokstad, P., Vatn, J., Aven, T. and Sørum, M. (2004). Use of risk acceptance criteria in Norwegian offshore industry: Dilemmas and challenges. *Risk Decision and Policy*; July.
23. Hopkins, A. (2008). *Failure to Learn: The BP Texas City Refinery Disaster*. CCH Australia; Reprint edition, July 1.
24. Hopkins, A. (2009). Thinking about process safety indicators. *Safety Science;* 47(4): 460-465.
25. IOGP – The International Association of Oil and Gas Producers. (2016). Standardization of barrier definitions. Report 544, Supplement to Report 415.
26. IOGP – The International Association of Oil and Gas Producers. (2018a). Asset integrity – the key to managing major incident risks. Report 415.
27. IOGP – The International Association of Oil and Gas Producers. (2018b). Process safety – Recommended practise on key performance indicators. Report 456.
28. ISO 14224. (2016). Petroleum, petrochemical and natural gas industries — Collection and exchange of reliability and maintenance data for equipment.
29. ISO 41011. (2017). Facility management — Vocabulary.
30. ISO 17776. (2016). Petroleum and natural gas industries — Offshore production installations — Major Accident hazard management during the design of new installations.
31. Jansen, M. (2015). Shell maintenance excellence. *SPE-177964*. Presented at the Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, UAE, 9-12 November 2015.
32. Kaganski, S. and Toompalu, S. (2017). Development of key performance selection index model. *Journal of Achievements in Materials and Manufacturing Engineering*; 82 (1): 33-40.
33. Kaplan, R. S. and Norton, D. P. (1996). The Balanced Scorecard: Translating Strategy into Action. 1st. edition Boston, MA, USA: Harvard Business Review Press.
34. Kaufman, R., Oakley-Browne, H., Watkins, R. and Leigh, D. (2003). Strategic Planning for Success: Aligning People, Performance, and Payoffs," San Francisco, CA: Jossey-Bass/Pfeiffer (by John Wiley & Sons, Inc.).
35. Kavanagh, D. (2013). Advantage – A Roadmap for Entrepreneurs and Leaders in The Digital Age. Kindle edition. Published by Declan Kavanagh. [http://intelligentorg.com/wp-content/uploads/2013/11/@Note-10-SMART-squared.pdf].
36. Kongsvik, T., Johnsen, S. Å. K. and Sklet, S. (2011). Safety climate and hydrocarbon leaks: An empirical contribution to the leading-lagging indicator discussion. *Journal of Loss Prevention in the Process Industries*; 24(4): 405-411.
37. Lawlor, K. B. and Hornyak, M. J. (2012). SMART goals: How the application of SMART goals can contribute to achievement of student leaning outcomes. *Developments in Business Simulation and Experiential Learning*; 39: 259-267.
38. Leveson, N. (2015). A systems approach to risk management through leading safety indicators. *Reliability Engineering and System Safety*; 136: 17-34.
39. Longhi, A. E. B., Pessoa, A. A. and de Almada Garcia, P. A. (2015). Multiobjective optimization of strategies for operation and testing of low-demand safety instrumented systems using a generic algorithm and fault trees. *Reliability Engineering and System Safety*; 142: 525-538.
40. Matheson, D., Matheson, J. (1998). The Smart Organization – Creating Value through Strategic R&D. Boston, MA, USA: Harvard Business School Press.
41. NORSOK D-010. (2013). Well integrity in drilling and well operations. Edition 4.
42. OECD – Organisation for Economic Cooperation and Development. (2008). Guidance on

developing safety performance indicators related to chemical accident prevention, preparedness and response. Series on chemical accidents no. 19. Second edition.

43. Parida, A. and Chattopadhyay, G. (2007). Development of a multi-criteria hierarchal framework for maintenance performance measurement (MPM). *Journal of Quality in Maintenance Engineering*; 13 (3): 241-258.
44. Parida A. and Kumar, U. (2006). Maintenance performance measurement (MPM): Issues and challenges. *Journal of Quality in Maintenance Engineering*; 12(3): 239-251.
45. Payne, S. C., Bergman, M. E., Beus, J. M., Rodríguez, J. M. and Henning, J. B. (2009). Safety climate: Leading or lagging indicator of safety outcomes? *Journal of Loss Prevention in the Process Industries*; 22(6): 735-739.
46. Peng, W., Sun, T., Rose, P. and Li, T. (2007). A semi-automatic system with an iterative learning method for discovering the leading indicators in business processes. Proceedings of the 2007 International Workshop on Domain Driven Data Mining, ACM. San Jose, CA, USA — August 12-15.
47. PSA – Petroleum Safety Authority Norway. (2013). Principles for barrier management in the process industries.
48. RapidBI. (2016). Writing SMARTer objectives and goals. In: https://rapidbi.com/writesmartobjectives/ [Uploaded 2 May 2016].
49. Roos, C. J., and Myers, P. E. (2015). The Engineer's Guide to Overfill Prevention. *Emerson Process Management*.
50. Saaty, T. L. (1980). The Analytic Hierarchy Process. McGraw-Hill Inc.
51. Saleh, J.H, Marais, K.B and Favaro, F.M. (2014a). System safety principles: A multidisciplinary engineering perspective. *Journal of Loss Prevention in the Process Industries;* 29: 283-294.
52. Saleh, J.H., Haga, R.A., Favaro, F.M. and Bakolas, E. (2014b). Texas City refinery accident: Case study in breakdown of defense-in-depth and violation of the safety-diagnosability principle in design. *Engineering Failure Analysis;* 36: 121-133.
53. Selvik, J.T. and Abrahamsen, E.B. (2015). A review of safety valve reliability using failure fraction information. In: Proceedings of the 25th European Safety and Reliability Conference, Zurich, Switzerland. 7-10 September 2015.
54. Selvik, J.T., Stanley, I. and Abrahamsen, E.B. (2020). SMART criteria for quality assessment of key performance indicators used in the oil and gas Industry. *International journal of Performability Engineering*; 16(7): 999-1007.
55. Smart, K. and Blakey, K. (2014). Achieving maintenance excellence in Maersk Oil Qatar. IPTC 17623. In: proceedings of the International Petroleum Technology Conference (IPTC), Doha, Qatar, 20-22 January 2014.
56. Smith, R. and Mobley, R. K. (2008). Chapter 6 – Key Performance Indicators. *Rules of Thumb for Maintenance and Reliability Engineers*; pp. 89-106. Butterworth-Heinemann.
57. Summers, A.E. and Hearn, W. (2010). Overfill protective systems—Complex problem, simple solution. *Process Safety Progress;* 29(4): 283-287.
58. Swuste, P., Theunissen, J., Schmitz, P., Reniers, G. and Blokland, P. (2016). Process safety indicators, a review of literature. *Journal of Loss Prevention in the Process Industries*; 40: 162-173.
59. Tamim, N., Laboureur, D. M., Mentzer, R. A., Hasan, A. R. and Mannan, M. S. (2017). A framework for developing leading indicators for offshore drillwell blowout incidents. *Process Safety and Environmental Protection*; 106: 256-262.
60. The BP US Refineries Independent Safety Review Panel. (2007). The report of the BP US refineries independent safety review panel. Downloaded 7 August 2020, from: http://sunnyday.mit.edu/Baker-panel-report.pdf
61. Vinnem, J.E. (2012). On the analysis of hydrocarbon leaks in the Norwegian offshore industry. *Journal of Loss Prevention in the Process Industries;* 25(4): 709–717.

62. Vukomanovic, M. and Radujkovic, M. (2013). The balanced scorecard and EFQM working together in a performance management framework in construction industry. *Journal of Civil Engineering and Management*; 19(5): 683-695.

63. Waite, P. (2013). Recurring accidents: Overfilling vessels. *Chemical Engineer*, 861: 40-44.

64. Wood, D. A. (2016). Asset portfolio multi-objective optimization tools provide insight to value, risk and strategy for gas and oil decision makers. *Journal of Natural Gas Science and Engineering*; 33: 196-216.

65. Øien, K., Utne, I. B. and Herrera, I. A. (2011). Building safety indicators: Part 1 – Theoretical foundation. *Safety Science* 2011; 49 (2): 148-161.

Paper V

Investigating the implementation of safety diagnosability principle to support defense-in-depth in the nuclear industry: A Fukushima Daiichi accident case study.

Authors: Surbhi Bansal, Jon Tømmerås Selvik

Published in the *Journal of Engineering Failure Analysis.*
ISSN 1350-6307,
https://doi.org/10.1016/j.engfailanal.2021.105315

Investigating the implementation of the safety-diagnosability principle to support defence-in-depth in the nuclear industry: A Fukushima Daiichi accident case study

Surbhi Bansal[a] and Jon T. Selvik[a,b,*]

[a] University of Stavanger, P.O. Box 8600, N-4036 Stavanger, Norway
[b] NORCE Norwegian Research Centre, P.O. Box 8046, N-4068 Stavanger, Norway

*Corresponding author: jon.t.selvik@uis.no

ABSTRACT

'Defence in depth' (DID) is a fundamental safety principle applied in several industries, including nuclear. The key is to protect safety critical systems by employing multiple layers of protection, i.e. barriers. The principle states that one single barrier, regardless of how reliable, is insufficient to ensure acceptable safety performance. Obviously then, as the reliability of the layers are associated with the risk of hazardous events, a main safety management activity should be to monitor barrier conditions and performance. However, as experienced in the past, there could be situations where such monitoring is unsatisfactory, challenging the usefulness of the DID. One example, taken from the oil and gas industry, is the 2005 Texas City refinery explosion, where multiple layers of protection failed, resulting in an accident caused by operators with poor situational awareness. Motivated by this assumed weakness, a new principle called the 'Safety diagnosability principle' (SDP) has been suggested for use in the oil and gas industry, in combination with the DID principle. The SDP requires that, for DID to function as intended, any degradation of barriers must be diagnosable and reported. The link to DID makes it also relevant to other industries. In this article, we consider the principle for the nuclear industry. The objective of the article is to clarify the benefits, different ways of implementation, and the potential for using SDP in conjunction with DID in the nuclear industry. To assess the value added, we evaluate the principle against different criteria characterising usefulness. Overall, we find the principle attractive, as the detection and diagnosis of safety-critical events or failures are important for safety management. Having such information strengthens the DID. On the other side, it can also be claimed that acquiring such information is already an implicit part of DID. If so, the SDP adds limited value beyond compliance, i.e. making sure the information is acceptable. We conclude that particularly the relevancy, but also the achievability, related to the use of the SPD, do not point in favour of the principle. A discussion on the 2011 Fukushima Daiichi nuclear accident strengthens our conclusions. The case study indicates that the SDP would not have made the outcome very different. However, as a standalone principle, it might be of greater value. Having reliable information about barrier performance is clearly important to safety management.

Keywords: defence-in-depth, safety diagnosability principle, usefulness, nuclear industry, Fukushima Daiichi

1 Introduction

Defence in depth (DID) is a safety principle requiring multiple and independent layers of defence, i.e. barriers. Each subsequent layer plays a role in protecting the system, meaning that, always, more than one layer needs to fail for an accident to be possible. It is a principle implemented across several industries. History has also shown, however, that accidents occur, despite systems being designed according to this principle. Saleh et al. [1] examined the Texas City refinery accident and determined that, as a result of misleading information related to barrier conditions and performance, and low situational awareness, operators made the accident possible. Saleh et al. [1] argue that the low awareness originates from the system not being able to provide sufficient information about barrier conditions and the progression of hazardous events. This lack of understanding of what had failed and what was really going on resulted in operators making poor decisions, ultimately leading to the accident. It is acknowledged that, without the availability of updated and reliable barrier information, the value of DID can be questioned. To compensate for this assumed weakness, Saleh et al. [1] suggest pairing DID with a new principle called the 'Safety diagnosability principle' (SDP); see also [2]. The SDP is all about setting up capabilities that reliably detect and report safety-degrading events and barrier failures. It is basically a principle advocating information availability and safety-informed decision-making. For further description, see 2.2 and [1].

The SDP is motivated by the analysis of the 2005 Texas City refinery explosion. The application and conclusions, however, are of a more generic character and linked to the use of DID for various safety management purposes within the oil and gas industry. Saleh et al. [1] also invite other industries where DID is implemented, such as nuclear, to consider the value of implementing the SDP. This suggests that the nuclear industry could face similar challenges regarding the diagnosability of safety barriers. A main objective of this article is to assess why the SDP should also be implemented in the nuclear industry. The key is to assess the usefulness of the principle, which indicates whether it adds value to safety management.

The international standard ISO 12749-5 [3] notes that an objective of DID is to "maintain the effectiveness of the barriers". Clearly, for DID to be effective, either implicitly or explicitly, decision-makers should be informed about safety-critical failures and critical operational aspects related to barrier performance. Otherwise, DID will remain a passive principle, heavily relying on robust barriers. Given that DID encompasses diagnosability requirements, it is possible to manage barriers in a more flexible way, and it will be possible to take actions when and if system reliability is not acceptable. The question is, then, how to achieve such information, as some barriers, for example, could be passive, in the sense that they might have 'hidden failures'. Despite extensive monitoring programmes, some conditions might not be diagnosable before an actual demand. Within maintenance engineering, there is a concept called 'maintenance induced failures' that refers to the possibility that performing, for example, functional testing can cause failures and reduce the reliability. From a system performance perspective, then, collecting reliability information with frequent intervals could be unfavourable for safety, although, if the diagnosability is implicitly already covered, one might question whether there is any need for a second principle on this.

To be clear, we will not give our opinion on the SDP for use in oil and gas and will assume the argumentation and conclusions reached in Saleh et al. [1] to be sound; it is outside our scope to say otherwise. However, it is not obvious that such a principle is needed in the nuclear industry, as it has different sources of hazard (risks), use of technology, operational procedures, etc. [4]. That is where we direct our focus in this article.

As a starting point, we need to define some criteria for what is meant by 'useful' or 'value adding', as a basis for the assessment. For this, we will adopt a set of criteria from Rosencrantz et al. and Sørskår et al. [5, 6], used in different contexts to assess the usefulness of other safety principles, i.e. Vision Zero and ALARP (As Low as Reasonably Practicable). These criteria allow us to investigate whether SDP contributes value beyond DID and allows us to capture the relevant pros and cons of the implementation. For specificity, we build the argumentation around the 2011 Fukushima Daiichi nuclear accident. This is one of the most recent events and, with the maximum level 7 on the International Nuclear Event Scale, the most severe nuclear accident since the Chernobyl accident of 1986. In brief, a 9.0-magnitude earthquake off the Japanese coast caused a tsunami that hit the Fukushima nuclear power plant, causing major destruction and the release of radiation to the atmosphere. The plant was designed to withstand waves up to 6 metres and was thus unable to stand against the 14-metre-high tsunami wave [7], causing flooding and station-wide blackout at the Fukushima nuclear power plant. In the days following the tsunami, the plant experienced a series of explosions. Several barriers failed. The failure of monitoring and diagnostic instruments impeded the correct diagnosis of the plant and safety system status throughout. We will use this case study to indicate the effect that a hypothetical prior implementation of the SDP could have had for barrier management in this scenario.

The article is structured into six sections. Section 2 outlines the two safety principles in focus: DID and the SDP. This section also clarifies the rationale for using this principle in the oil and gas industry. Section 3 presents and clarifies the criteria adopted for assessing the usefulness of the SDP. Then, in Section 4, we give an overview of what happened at the Fukushima Daiichi accident, the failed safety barriers, and the causal factors. Among these particularly, factors related to the presumed failed diagnosability are identified. In Section 5, we discuss the extent to which improved diagnosability could have prevented the accident or reduced its consequences. Here, the role of failed monitoring systems (e.g. core temperature sensors, water level monitors) is compared with failed mitigatory barriers (e.g. evacuation plans, backup power and water supply) in accelerating the accident. We end the accident discussion by analysing whether restoration of diagnosing capability could have improved the outcome. Finally, Section 6 presents some conclusions and recommendations, based on the identified pros and cons related to use of the SDP in combination with DID in the nuclear industry.

2 Background

2.1 Defence in depth

As described above, DID is the principle of protecting safety or some asset by using multiple layers of successive barriers. The role of the barriers can be visualised with reference to a traditional bow-tie diagram, displaying both preventive and mitigating barriers. It depicts the pathway from causes, through some critical event, to the possible consequences. And it is particularly useful in identifying pathways not following a linear route. The DID complements such a presentation by adding requirements to the barriers displayed or communicated by the bow-tie diagram.

A key when considering DID is that the barriers are independent, and that each layer offers significant protection. It is pointed out for nuclear applications in the fundamental safety principles outlined by the International Atomic Energy Agency (IAEA) [8], "The independent effectiveness of the different levels of defence is a necessary element of defence in depth", meaning that a set of independent barriers must be penetrated for "the asset to be acquired" [9]. It is possible to define DID in different ways, and it has seen some widely discussed

developments (see e.g. [9, 10]), as might be expected for a principle used for decades in various industries, but the core understanding remains more or less the same.

There are two definitions given in ISO standards, both addressing nuclear applications. ISO 1709 [11] defines DID as "hierarchical deployment of different levels of diverse equipment and procedures (known as barriers) to prevent the escalation of faults to a hazardous condition", which is quite similar to the one given in ISO 12749-5 [3]: "hierarchical deployment of different levels of diverse equipment and procedures to prevent the escalation of anticipated operational occurrences or events". Both standards have adopted and modified the definition given in the IAEA safety glossary [12], where the wording is slightly longer: "A hierarchical deployment of different levels of diverse equipment and procedures to prevent the escalation of anticipated operational occurrences and to maintain the effectiveness of physical barriers placed between a radiation source or radioactive material and workers, members of the public or the environment, in operational states and, for some barriers, in accident conditions."

Typical descriptions of DID comprise terms such as 'successive compensatory measures', 'several layers of protection', 'hierarchical deployment of equipment/procedures', 'depth of penetration', etc. All these terms are associated with the idea of investing in multiple layers aimed at protecting the asset of importance. These are not necessarily safety assets; the asset can be the safety of the workers, society, environment, software, or other hardware (physical) assets. DID is also used for security applications; see [13] for security-related DID definitions. In principle, regardless of application area, the barriers should be effective in managing a system's response to any relevant hazard (human, mechanical and naturally caused events/failures). If one barrier fails to fulfil its intended function, the ongoing hazardous event sequence (e.g. rising reactor core temperature) should be handled in an effective way. The likelihood of severe accidents with serious consequences should be rendered extremely small, with accident prevention being the first priority [14]. For this, safety barriers (such as human, technical or organisational) are employed at every stage (before, during and after) in the event-to-accident escalation path. Barriers at different locations cater for accident prevention, ensuring barrier integrity (or block further escalation) and consequence mitigation [14-16]. However, the principle should be viewed beyond just the barriers, also capturing aspects of control for proper safety management [17], as also stated in the Fukushima Daiichi accident lessons learned [18-19]. The above understanding is summarised in the following three pillars, important for an effective DID strategy [2]:

1. Multiple lines of defence should be placed along potential accident sequences
2. Safety should not rely on a single defensive element (hence the 'depth' qualifier)
3. The successive barriers should be diverse in nature and include technical, operational, and organisational safety barriers (i.e., not only the physical defences).

The three strategy pillars together serve three fundamental safety functions, relevant to the nuclear industry [15]:

- Reactivity control
- Heat removal from the reactor and fuel store
- Confinement of radioactive material

The nuclear industry follows a five-level barrier system, to ensure the above safety functions. This is so that, should one level fail, the subsequent level comes into play [20]. Table 1 gives an overview of these five levels of defence in depth defined by the International Nuclear Safety Advisory Group (INSAG); for notes on the definition of 'defence in depth', we refer to [21].

Table 1 Overview of levels in defence in depth [21]

Level 1	Level 2	Level 3	Level 4	Level 5
Prevention of abnormal operation and failures	Control of abnormal operation and detection of failures	Control of accidents within the design basis	Control of severe plant conditions, including prevention of accident	Mitigation of radiological consequences of significant releases of radioactive material

At the first level, the focus is on typical activities and failures that could have a safety impact. Level 1 refers to main barriers failing, for example, activating redundant equipment to satisfy a given safety function, or instrumentation giving an alarm when safety-related performance is outside acceptable levels. At the second level, one could have failure of barriers linked to abnormal operational deviations. These are events that do not occur as frequently and might require barriers that have a more passive role in normal operations. The key is to detect and control the situation, so that it does not escalate. At level 3, if a hazardous event occurs, there should be barriers to shut this down in an effective way, to avoid consequences and return to safe operation. Then, for level 4, there should be barriers preventing or inhibiting the consequence development and escalation. Level 5 refers to mitigating barriers related to emergency response, as the final step before the consequences are realised. These levels are discussed in more detail in [21]. The levels can be illustrated by reference to a traditional bow-tie diagram, where levels 1 and 2 are on the left side of the diagram, dealing with causes, the third level being placed around the centre (hazardous event), and levels 4 and 5 being placed on the right side, dealing with mitigating measures and consequences. It is also common to group the levels into three safety layers: hardware, software and management control [19]. Such a combination of barriers, if implemented appropriately, is deemed robust against single or combined failures, unexpected failures and 'beyond design' situations. The key is to ensure independence amongst the barriers. One way to achieve this is by following criteria of diversity, physical separation, and functional isolation [15]. The idea is that independent barriers should not share common causes of failure. It is important that one failed barrier does not increase the probability of other barriers failing. Rather, it should minimise the escalation of deviations during normal operations, particularly to avoid so-called 'cliff-edge effects', i.e. an abruptly large variation in plant condition in response to a small variation in an input [22].

Over the years, the nuclear industry has continuously reviewed the DID content, to ensure it holds as an effective safety principle. This builds on a substantial collective knowledge base that the industry has acquired over the years, including the building, operating and maintaining a variety of nuclear plants, combined with lessons learned from several serious accidents and incidents [22]. The idea of DID has also evolved within different frameworks (such as design-DID, process-DID, and scenario-DID) of nuclear safety; refer to [22] for details. To some extent, this collected experience of lessons learned, observations and use cases contribute to a shared and improved understanding of DID and its value, visible in the regulatory standards of today. Overall, defence in depth is a key concept for better assurance of nuclear safety, by compensating for uncertainties and incompleteness in knowledge [23].

2.2 Safety diagnosability principle

According to Saleh et al. [1], the breakdown of barriers and effects, leading to the 2005 Texas City refinery accident, demonstrates an inherent weakness of DID. It shows that, by adopting this safety principle, one could have multiple independent barriers but still not be well protected. Salah et al. point to the lack of diagnosability, hindering the detection of hazardous states during operation, as a main failure mechanism. Diagnosability refers to the ability to determine whether

the system can detect a fault after its occurrence [24]. Poor diagnosability can also be seen as a side effect of redundancy of safety barriers, since it makes the system opaque to the people managing it [25]. For this particular accident, poor system diagnosability left 'blind spots' during operations, concealing the presence of an approaching hazard. This hazard materialised when the conditions in the system exceeded acceptable levels, without the operators being aware of it. The SDP is an initiative to reduce the likelihood of this happening, by requiring an ability to diagnose the hazard build-up concealed by such blind spots.

Saleh et al. [1] outline the SPD as follows: "This principle requires that all safety-degrading events or states that defence in depth is meant to protect against be observable/diagnosable. This principle requires that various features be put in place to observe and monitor for breaches of any safety barrier, and reliably provide this feedback to the operators". See also [26].

The core of the SDP is to reduce uncertainty related to barrier performance, meaning that any barrier should be observable, which in a way gives more control with respect to the issue of uncertainty. The principle requires actions if the conditions are not monitored or observable, given that the information achieved is credible or accurate. It requires reliable information to be available to reflect the barriers' conditions and performance at the relevant time. Facilitating such information allows for actions to make barriers diagnosable or to simply remove them, to avoid a false sense of safety.

A main motivation for this principle is to close the gap between the assumed and actual hazard levels, by increasing awareness of barrier conditions and performance. Its importance for accident prevention lies in the value of the information it supplies and the actions and interventions it spurs [2]. With reference to the Texas City accident, it has been demonstrated that non-compliance with the SDP can degenerate DID into an ineffective defence-blind safety strategy [26]. Violation of the SDP introduces an element of non-transparency regarding barrier effectiveness. Hence, it might lead to a sense of safety by falsely assuming the presence of functional barriers, which can translate into underestimation of hazardous event probabilities. We may end up facing implications of overconfidence in the safety barriers. Factors such as below-expectation barrier performance and a low response time window should obviously be captured by management, to prevent major accidents.

The SDP's usefulness is linked particularly to the left side of the bow tie and the implemented preventive barriers or measures. The availability of these build on the ability to detect and diagnose system conditions. In many situations, this will be necessary for them to perform the required function when needed. For example, there are preventive barriers, dormant in normal operations, such as redundant systems, which might require fault detection as a stimulus to activate them. For manually operated barriers, the sooner the hazardous situation is detected, the quicker barriers can be activated. Further, DID incorporates a need to diagnose safety conditions at different levels (see Table 1). Diagnosability is important to make the operator or decision-maker aware of what is really going on, so that the higher-level barriers are given sufficient attention. Based on this, it could be that the SDP places more weight on preventive compared with mitigatory measures. With robust preventive measures, there is small probability of any mitigating measures being activated in the first place. Based on the analysis of the Texas City refinery accident in [1], it appears that the greater focus is on preventing hazardous events rather than on mitigative measures minimising the consequences. However, that might not be intended. The principle should, nevertheless, not be seen as a way of prioritising between preventing (proactive) and mitigating (reactive) measures.

As it is relevant to basically all industries using DID, Saleh et al. [1] also invite the nuclear industry to consider implementing the SPD. The idea is to use this principle to complement DID, but it might also be considered as a standalone principle to strengthen barrier management. However, the SDP has not yet been recommended as a standalone principle, i.e. for situations where DID does not apply. In this article, our focus is on using the principles in combination, meaning that the usefulness or added value of the SDP comes from ensuring informed use of DID. Implicitly, it means that safety decision-making could be improved and could lead to different outcomes, compared with situations with no reference to the SDP.

The need for the SDP is motivated by past events and experience using DID in the oil and gas industry. This is an industry where barrier management overall is given a high level of attention, and where it is recognised as important to observe barrier conditions and performance, and update barrier reliability estimates, to demonstrate that performance satisfies the required safety integrity levels. Especially, there is much focus on barriers in systems with major accident potential. Despite this, for example as regards well design, safety-critical equipment could be installed downhole with limited or complex monitoring options. The oil and gas industry monitors several hazards due to the complex nature of operations that require constant vigilance. There is a wide spread of production activities taking place at several distinct locations, and implementing the latest technology to increase profitability is a common practice [4]. Hydrocarbons need to be moved across units (for example, from offshore platform to gas extraction unit to refinery), and their control is usually more decentralised compared with operations in the nuclear industry, where there is perhaps also less variability in the type of operations, while the potential worst-case consequences of accidents are considered less likely and more severe. Nuclear power plant operators typically have a greater time window to respond during disturbed conditions [4]. There are differences, obviously, but there is nothing in the operational differences to suggest that the SDP should not be transferrable from oil and gas to nuclear.

3 Usefulness assessment criteria for SDP

In the nuclear industry, DID has a role guiding managerial decisions about the sufficiency of levels of protection against the radiation risk. The idea is that the SDP complements the DID, by ensuring a higher focus on quality information feedback related to barrier performance. As a main safety objective is to have functioning barriers at any time, such information is seen as important for barrier management, meaning that, clearly, there are positive aspects. But we should also consider arguments for not implementing the SDP, which will contribute to a more nuanced evaluation of the principle, covering both pros and cons. For example, depending on the system considered, it might be challenging to achieve diagnosability in practice; see e.g. [27].

To assess the overall value of the SDP as a key principle for nuclear applications, we need an appropriate instrument: one that allows us to systematically evaluate its usefulness. What we look for is a set of criteria that can be used to assess whether the quality and value of the information provided by implementing the SDP are sufficiently in favour of the principle, in other words: how the principle influences safety management quality.

For a suitable set of criteria, we refer to Sørskår et al. [6], who use a set of criteria adopted from Edvardsson and Hanson [28] to assess the appropriateness of combining two other key safety or risk management principles, i.e. the ALARP and the Vision Zero principles. For the assessment, four rationality criteria (i.e. precision, evaluability, approachability, and motivating) are used to evaluate relevant aspects. The criteria allow for a consistent and transparent evaluation, while covering the main aspects of risk and safety management.

The four criteria suggested in [6] capture basically the same aspects as the criteria given by the SMART acronym: specific (S), measurable (M), achievable (A), relevant (R) and timely (T), enlisting them as an alternative set of criteria for appropriate quality [29-30]. Although the SMART criteria in [30] are not demonstrated specifically for safety principles, we interpret the two as interchangeable. Table 2 shows the correspondence among their criteria (refer to [6, 29-30] for more details).

Table 2 Similarity between rationality and SMART criteria

Rationality criteria	**SMART criteria**
Precise	A precise principle is one that is 'directionally, completely and temporally' precise. This corresponds to the 'specific' and 'timely' SMART qualities.
Evaluable	Performance towards the objective stated by the principle should be evaluable. This corresponds to the 'measurability' of progress towards attainment of an objective.
Approachable	Approachability refers to the quality of being 'achievable' or at least approachable to a reasonable degree.
Motivating	Motivating criterion refers to the ability to induce a suitable kind of action by agents. This inherently relates to the 'relevancy' criterion that decides the importance of the objective stated by the principle for business/safety purposes.

As can be seen from Table 2, the two alternatives prescribe similar criteria. This means that there are no practical implications of using one set over the other. One should arrive at the same conclusions, irrespective of which set was adopted for the assessment. The SMART framework is clearly the one most cited among the two and considered the most recognised. It is intuitive and quite simple to use in practice, and we will adopt it for our assessment of the SDP in this paper.

The five SMART criteria are further clarified below:

- *Specific:* The objective of the principle should be precisely and clearly defined. The implementing agents must have a clear understanding, to be able to use it consistently.
- *Measurable*: It should be possible to rationally measure the progress towards or achievement of the objective. Whether the objective is met, where we currently stand, and if we are going in the right direction, should be evaluable.
- *Achievable*: This refers to the degree to which the principle/objective is practically achievable. It concerns factors such as cost, knowledge and practical limitations affecting the certainty of achievement.
- *Relevant*: It should contribute to the organisation in a meaningful way, i.e. add value. The significance will be affected by conflicts or overlap with other business objectives and goals. A relevant principle will also motivate the agents to work for it persistently.
- *Timely*: The principle should have a time horizon in which the objective should be achieved.

The SDP should satisfy all these criteria to prove its informational value to DID and to demonstrate added value for overall safety management. This will serve as an input to evaluate its suitability, in combination with DID, for the nuclear industry.

4 Presentation of case for the SDP assessment

For more specificity, we will refer to an actual accident case scenario as a basis for the discussion. Several of the aspects related to the SMART criteria make little sense without such a reference, particularly achievability and relevancy. Without a more practical context for the discussion, it is difficult to conclude on its actual usefulness. Thus, before moving into an assessment of the SDP using the SMART criteria, we introduce a case based on the 2011 Fukushima Daiichi nuclear accident.

The Fukushima Daiichi plant and process design were guided by the DID principle. However, as history shows, DID's implementation could not prevent the accident from materialising. Below, we investigate whether the SDP would have made a significant difference to the accident outcome. We will use the findings from what happened in the discussion (in Section 5), along with arguments that can be given on an overall basis for the nuclear industry regarding the SDP. The discussion on its value-adding potential linked to this accident depends to some extent on the findings, but we might not necessarily be able to draw generalised conclusions based on this one accident scenario alone. However, should we conclude that the principle lacks usefulness for this scenario, there will be strong reasons to question the rationale for giving it a key role in the safety management of other nuclear power plants.

4.1 Fukushima Daiichi nuclear power plant overview

Fukushima Daiichi nuclear power plant (NPP) is located on the eastern coast of Japan. Figure 1 depicts the plant's layout. It has a total of six units (1-6). Units 1-4 are located on the left and the rest are on the right. Each unit has a reactor building (RB), a turbine building (TB), an emergency diesel generator (EDG) and relevant switchgear. The units share a common spent fuel building to store a large amount of fuel assemblies. The pumps located in front are used for pumping sea water and circulating water in the units. The administration building and the emergency response centre are in a seismically isolated building, located behind the units at an elevation. A back-wash valve pit used for filtering water is located in front of unit 3. The site has a seawall, to protect against tsunami waves of a height of up to 5.5m. It opens directly onto the ocean.

The setup of a typical unit in the power plant is shown in figure 2. The unit has two sides: a reactor building and a turbine building. The two sides together run a closed-loop steam cycle. The cycle begins with a nuclear fission reaction inside the reactor pressure vessel (RPV). The RPV is housed in the primary containment vessel (PCV) on the reactor side. The radioactive fuel in the RPV absorbs neutrons, triggering a chain reaction that releases energy. The process reactivity is controlled by control rods and immersing the fuel in water. The PCV is connected to the suppression chambers that store water to manage the reactor pressure. The nuclear reaction generates energy in the form of heat. The RPV has incoming water through a feedwater line. The generated heat vaporises this water, and it travels through the main steam line towards the TB. Here, the steam drives the turbine, so that a generator can produce electricity. After driving the turbine blades, the steam is condensed into water by a condenser. The condenser uses pumped ocean water as its cooling medium. The water is recirculated to the reactor side via the feedwater line, and the cycle keeps repeating. Clearly, ensuring a consistent water supply is important, as it plays multiple roles as a working fluid, coolant and moderator of reactivity.

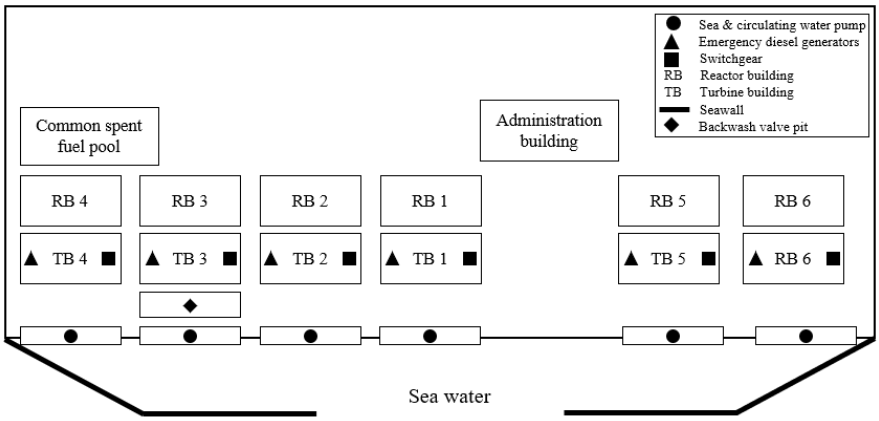

Figure 6 Fukushima Daiichi nuclear power plant layout

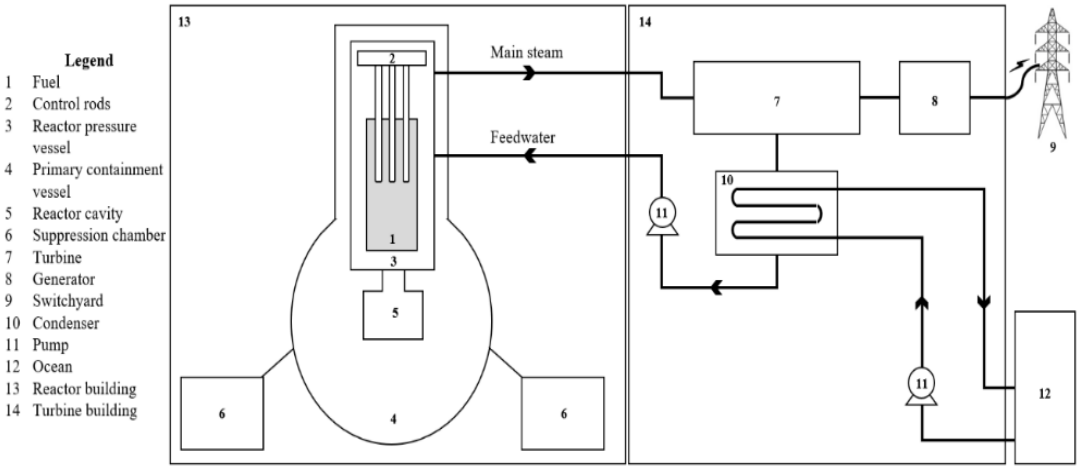

Figure 2 Setup of the NPP unit

4.2 Overview of safety barriers

The Fukushima Daiichi plant employed the defence-in-depth principle as its fundamental safety principle. It had three main barrier levels, as against the five levels prescribed in the IAEA standards. The plant should operate safely during normal circumstances, as well as under

emergency conditions. For this, several barriers for core cooling and radioactivity containment were ensured. Table 3 lists the safety barriers and their corresponding functions below:

Table 3 Safety barriers at Fukushima Daiichi units

1. Safety barrier against uncontrolled reactivity
1.1 Control rods – Scram system to shut down reactor
2. Safety barriers against reactor heating during operation
2.1 Condenser – Cools the feedwater that keeps fuel rods covered
2.2 Fuel pool cooling – Spent fuel (in the storage) kept submerged in water
3. Safety barriers against containment breach
3.1 Fuel protection – Zirconium cladding to protect fuel against corrosion
3.2 Primary containment vessel – Houses the RPV with nuclear fuel (primary containment barrier)
3.3 Reactor building – This concrete building serves as the secondary containment barrier between PCV and external environment
4. Safety barriers against loss of coolant event
4.1 Reactor core cooling – Sprays cooling water on top of the reactor, high-pressure injection system
4.2 PCV cooling – Sprays cooling water inside the PCV
4.3 Coolant cooling – Isolation condenser, Residual heat removal system, Suppression chamber
5. Safety barriers for other hazards
5.1 Hydrogen release – Hydrogen detection and removal system in the RPV
5.2 Fire hazard – Fire protection system (also a backup system for core cooling under accidents)

4.3 Accident sequence

Explosions at the Fukushima NPP spanned several days, following a complex sequence of events. The plant supervisors, operators and government authorities were unable to gather information about these events in time. We now look at the accident sequence that led to the explosions.

4.3.1 Initiating event sequence

An earthquake of 9.0 magnitude took place on 11 March 2011, off the Pacific coast of the north-eastern Japanese mainland [31]. The epicentre was 24km deep into the Pacific Ocean and 180km from the Fukushima Daiichi NPP [32]. On the incident day, units 1-3 were operational, while units 4-6 were in different stages of planned maintenance.

Unit 4: fuel offloaded to spent fuel pool and emitting a large amount of decay heat
Units 5 and 6: fuel assembly inside the reactor core but emitting low decay heat

The two-minute-long earthquake damaged the power transmission and distribution systems across the region. Fukushima NPP experienced a power outage. The power interruption triggered the automatic emergency response system and stopped the nuclear reaction in units 1-3. Their nuclear cores kept emitting decay heat in their surroundings, raising the temperature and pressure. For a safe halting of operations, a cold shutdown had to be achieved. Cold shutdown is the stage at which, after a few hours of reactor shutdown, actively cooling with recirculated water drops the temperature below 100°C, such that active cooling is no longer needed, and the reactor becomes passively safe [33].

The earthquake triggered a loss-of-offsite-power (LOOP) event in the plant. This refers to the loss of AC power at the plant. LOOP automatically initiated the onsite EDGs to supply the necessary AC power to the units (1-3). Consequently, the units could begin using the isolation

condensers to cool their reactor cores. Their temperature and pressure started lowering immediately. The earthquake also triggered the tsunami waves. Shortly after restoring the emergency power, the plant was flooded by tsunami waves of 16m height. The 5.5m seawall was entirely ineffective in preventing site inundation. The flood water entered the reactor, turbine and service buildings. Equipment necessary for ensuring the cooling function, such as pumps, EDGs, motors, power connections, switchgear, etc., were either damaged or immersed in water. The NPP had now also lost its emergency AC power source. This caused a station blackout, a specific event where the plant units experience a loss of AC power for more than five minutes [12]. The offsite emergency response centre and Japanese ministry declared a nuclear emergency.

4.3.2 Consequence sequence

Units 1, 3 and 4 shared a common sequence of events leading to explosions in their respective reactor buildings. These explosions spread out over several days following the tsunami. Given the similarities among their accidental path, we limit our analysis to unit 1, which was the first unit to experience an explosion.

The earthquake had caused the LOOP event. This triggered several emergency response systems: (1) The loss of AC power automatically started the emergency diesel generators. (2) The ventilation system stopped working, and the temperature and pressure inside the containment vessel started rising. The operators diagnosed this and started the cooling system manually. (3) After being shut down, the reactor became isolated from the turbine building's condenser cooling system; its rising pressure automatically started the isolation condenser (IC) system. The IC started removing the residual heat from the PCV. After some time, the IC was manually stopped, as it was decreasing reactor pressure and coolant temperature too rapidly. The NPP's safety barriers were operational, diagnosable, and the situation was now under control.

However, the earthquake was shortly followed by several tsunami waves. The tsunami flooded the basement of the reactor building. The emergency generators, DC panels and battery units located there were inundated. Unit 1 lost both the onsite AC and DC backup power. AC power was crucial to run the safety barrier equipment; the DC power supply was vital for plant safety, as it was needed for instrumentation and control and supplied AC power from inverters to a small number of essential components [32]. The tsunami had the following consequences:

(1) *Loss of backup AC power:* resulted in lost emergency core cooling barriers.
(2) *Loss of backup DC power:* Operators lost instrumentation, alarms and sensors that monitored the reactor water level, reactor pressure, cooling barriers' status, temperature and water level in the spent fuel pool, and status of the IC system.

The reactor lost all the cooling systems and the power necessary to energise and monitor them. Without the cooling function, the containment started to be pressurised by the evaporating water. As the water level dropped, the core would soon become uncovered. The heated core, if unchecked, could melt down and risk radioactive release.

Dissipating the decay heat became a priority in unit 1. The decay heat could accelerate water evaporation and reduce the water level in the core. If this evaporation remained unchecked, the nuclear core would be uncovered, overheated, and might end up in a core meltdown. Loss of AC/DC power due to a blackout triggered a downward spiral of events. The operators could not ensure the core cooling function, as it ran on electricity. They faced a twofold challenge. Firstly, the critical pumps and valves to achieve cold shutdown could not be operated, due to a loss of AC power. Secondly, there was uncertainty about the reactor status, as the unavailability of DC

power rendered the instrumentation useless. They decided to initiate their efforts to first arrange power to run the equipment.

They started formulating strategies for barriers that could stop the potential nuclear fuel degradation. For a short duration, the reactor water monitor activated and displayed a decreasing water level in the RPV. So, the team decided to cool the core by injecting water. They started arranging alternative equipment (such as the fire protection system, fire engines and freshwater tanks) for this, given that the existing cooling barriers had been rendered powerless. Additionally, there were repeated attempts to start the IC. The IC system condensed the incoming reactor steam pipeline by submerging it in a cold-water tank. As mentioned above, this system had been shut down just before the tsunami arrived. However, loss of AC power post-tsunami meant that its availability was unknown. The operators tried to restart it, believing that the valves inside the containment that routed steam to the IC were open. This assumption turned out to be wrong, when the IC failed to start. The timing and sequence of power loss had unknowingly closed the valves.

Fearing a degradation of the core, the operators had to manually read the reactor pressure, by visiting the reactor building. They confirmed that core pressure was increasing. By this time, the alternative water injection arrangement was complete, but it could not be initiated. High core pressure conditions rendered the alternative low-pressure water injection impossible. In the meantime, temporary batteries were used to restore DC power and energise the indicators. The readings on the water level monitor indicated that the reactor core was submerged. However, investigation reports suggest that the level indicators were unreliable [13].

After some time, two operators detected radiation outside unit 1, using their personal dosimeters. This was a sign that the core had started degrading, possibly due to low water level. As the radiation started spreading to the main control room, the failure of containment barriers also became a likely scenario. By the end of day one of the accident, the drywell pressure (inside the reactor) was found to be exceeding its maximum design pressure. This high pressure was a warning of an exceedingly critical situation in the unit. The site superintendent decided to vent the PCV to reduce this abnormal pressure level. This was also necessary to resume water injection. They communicated this to the Japanese government, who allowed the venting after residents in a 3-km radius were evacuated. Even after evacuation was complete, the ventilation kept on being delayed.

On 12 March, the following day, the operators managed to start water injection at 0400h, using a fire truck, which fetched water repeatedly from a freshwater tank. In the following hours, the operators noticed a drop in the containment vessel pressure, without any established ventilation paths. This observation, coupled with a significant increase in radiation dose rate, suggested that the primary containment was failing. In response to this, the government extended the evacuation zone to 10 km.

After a few hours, the workers were able to establish a continuous water injection line between the freshwater tank and the reactor. Although the team had clearance for manually venting the PCV, the ventilation had still not begun. Either the operators were forced to abandon the reactor building as a safeguard against radiation exposure and recurring tsunami threat or they faced challenges in opening the valves manually. After a few hours, they finally managed to open the PCV vent line valves. The pressure venting was done successfully, as a reduction in PCV pressure was observed. By 1530h, AC power restoration, water provisions and core cooling supplies had

been re-established in the unit. However, before they could be used, there was an explosion in the unit 1 reactor building topside. The explosion did not, however, affect the PCV. The source of the explosion is attributed to a hydrogen-air reaction. A reaction between zirconium (nuclear fuel cladding) and water under high temperature had released hydrogen gas, which had, unbeknownst to anyone, escaped to the reactor building via some unobserved path. There, it mixed with the air, causing a violent explosion. Being exothermic in nature, the hydrogen gas reaction produced heat that further accelerated fuel heating [8]. This released more radiation due to core melting, in addition to the radioactive gases released by the explosion. The explosion's pressure damaged the power cables and injection lines laid down for units 2 and 3. In the following days, unit 3 had a hydrogen explosion on the top floor of its reactor building. This was followed by another explosion in unit 4, wherein hydrogen had leaked through a vent from unit 3. Unit 2 did not experience an explosion, despite a damaged reactor core and pressure build-up. The investigators believe the opening of the top floor blow-out panels, due to the explosion in unit 1 nearby, and the lower hydrogen gas generation, to be the possible reasons [8]. The ceiling holes were also potential venting outlets for hydrogen gas accumulating inside the structure.

For further details, we refer to e.g. [19, 34].

5 Discussion - Assessment of usefulness

5.1 Basis for the discussion

In this section, we will use the above presented case to discuss arguments for and against complementing DID with the SDP principle. We will use the SMART criteria (see Section 3) as a basis for the discussion. The discussion will draw on the experiences from the Fukushima Daiichi accident. This will provide insights into the potential role of the SDP in nuclear accident situations.

5.2 Specificity discussion

This criterion can be assessed on a general basis for nuclear applications and is not specific to the scenario above.

According to definition, the SDP requires that all safety degrading events or states that DID is meant to protect against be observable or diagnosable. In other words, the principle requires the implementation of observing or monitoring features that look out for safety barrier breaches and reliably provide feedback to the operator. The SDP's precision lies in the clarity of its objective and direction to the implementing agent, by requiring actions if this is not fulfilled.

The principle allows for two ways of interpreting the objective: moderately and strictly, of which the moderate objective is substantially less demanding and requires that barrier degrading events are diagnosed and reported through feedback. For consistent implementation, monitoring features should be set up. The features should reliably provide information whenever DID-relevant events cause a safety barrier breach.

The stricter version of the objective leans towards a more extreme safety perspective. It requires the system to monitor the complete state of barriers. This implies that all the status parameters of a safety barrier need to be observable, not just the information about its breach event. Then, the combined scope of monitorable events and states increases exponentially. The rationale is that the barriers with even marginal deviations from the normal operating conditions may lead to a

potential barrier breach. The operator should have the maximum amount of information to predict a barrier failure considerably in advance. This will ensure the availability of a longer response window to the operator. The choice of moderate versus stricter SDP objective will depend on considerations such as risk appetite, cost-benefit evaluation, budget constraints, technology challenges, etc. This requires a managerial review and judgement and has been left to the management, as the principle cannot guide on this aspect.

Overall, the principle is seen as sufficiently specific, with a flexibly defined objective. It also provides a definite direction for the actions to achieve the objective. We argue that the principle is sufficiently specific.

5.3 Measurability discussion

Measurability is mostly a matter of which information it is technically possible to collect regarding barrier performance. Although it somewhat depends on the type of barrier, we will be able to draw inferences here based on general barrier understanding. Basically, what we want to know is whether there are obstacles hindering us in monitoring or collecting information on barrier conditions.

The level of barrier diagnosability should be measurable through some metric. To achieve this, we require information such as how many barriers are currently monitored and, amongst these, the number of states or critical events, or the development of degrading processes. But collecting such data can obviously be challenging. The size of the state space would increase exponentially with the system's complexity [35], especially for the stricter SDP objective. The analyst evaluating this metric might have difficulties in comparing the captured state space versus the real state space. Further, all these events/states need to be simulated to count the diagnosable fraction, which is quite challenging. This raises uncertainty about the background knowledge supporting this metric. It can be claimed that any measurement or evaluation made without the knowledge of this uncertainty would be meaningless. Instead, feedback or knowledge of past results can help in measuring and improving performance towards the objective [30]. Trend indicators can measure this progress. For example, for a nuclear reactor with a history of hydrogen gas leaks, an increasing trend of undiagnosed or delayed detections indicates poor diagnosability. The management implementing the SDP can then use this indicator to take actions that improve the diagnosability level in the future (e.g. installing gas detectors at the barriers and hidden escape paths). Such trend indicators also require careful judgement, especially when compiling and evaluating trends for normal operative periods or zero-missed detections.

Trend indicators could be useful in quantifying and assessing the system's ability to observe specific failures and events. Besides, the monitoring ability can be claimed to be simply a matter of cost and not really an issue with respect to the measurability. Overall, this ensures that the SDP's objective is measurable, and we conclude that the measurability criterion is satisfied.

5.4 Achievability discussion

The achievability criterion is highly scenario-specific. In a way, this criterion addresses the core of the principle: whether it is practically possible to obtain the barrier information with high confidence. It is a matter of removing uncertainty related to the barrier performance, while also considering the available resources and other business objectives.

Safety barriers experiencing failures are particularly important for this discussion. Motivated by the case presentation in Section 4, we focus on the performance of the following three barriers:

- The reactor core cooling barrier
- The containment integrity and hydrogen removal system
- The human-organisational barrier

A failure of these barriers was significant for the accident. For each of these barrier failures, we first consider the barrier monitoring capability already present (without following the SDP) and why it failed. Then, we will consider the potential benefits of the SDP: whether its diagnosis information had the potential, retrospectively, to avert the accident.

5.4.1 Loss of reactor core cooling barrier

Right from the start of the accident, the plant lost its normal and emergency core cooling barrier systems. The plant units were equipped with several sensors and instruments to monitor their status. Water level and temperature monitors were used to observe the barrier effectiveness against the accumulation of process decay heat. Additionally, valve status (open or closed) and activation indicators provided information on the barrier cooling's availability or failure. The units ensured diagnosability to a large degree, without mandating the SDP in the first place. This came from the diagnosis and monitoring requirements of DID. Following the tsunami-induced power blackout and site inundation, most of the units lost their safety barriers beyond defence level 2 (see table 1 for description of levels).

In retrospect, let us consider that the SDP was applied, such that all the monitoring features were functional. Often, normally reliable instrumentation becomes untrustworthy under extreme operating conditions of high pressure, temperature, radiation, etc. Then the reliability of the diagnosis received during accidental situations becomes uncertain. This also happened in the Fukushima accident. The erratic monitoring instrument readings misled the operators. Unit 1's water level indicator was key to monitoring and confirming the core cooling barrier's status. The instrument's unreliability became known only after the operators discovered that the actual reactor conditions and the displayed readings were incompatible. This uncertainty caused a loss of response time and induced stress among the operators. They made poor decisions that later required additional resources to retract. The operators had to physically verify the reliability of the indicators and lost precious time. They eventually shifted priorities towards re-establishing the integrity of safety barriers and arranging external help. The likelihood of unreliable diagnosis, which deteriorates further as the operating conditions become adverse, undermines the usefulness of the SDP.

5.4.2 Loss of containment integrity and hydrogen removal system

The hydrogen gas leaked from the PCV following unknown paths in unit 1. As per DID's monitoring requirements, the units were equipped with hydrogen detection instruments. These monitored the hydrogen level in the PCV that was filled with inert nitrogen gas as a barrier against explosion. But the plant's DID barriers were not designed to prevent hydrogen gas migration from the PCV to the reactor building. This was due to the assumption that hydrogen could not leak out of the PCV, which was the only standing barrier preventing hydrogen gas from leaking outside. However, eventually, the combination of core damage, high containment pressure and temperature compromised the containment, allowing hydrogen to escape from the PCV [13]. It is estimated that gaskets, flanges, cableways etc., weakened by high temperature, were possible escape routes that breached the PCV's leak seal and integrity [8]. As a result of this seriously flawed assumption, hydrogen gas build-up in the unit 1 reactor building remained hidden, as there were no monitors to detect it.

The hydrogen level monitors inside the PCV were unavailable due to power outage. The RPV could have accumulated 10,000 m^3 hydrogen in just half a day, due to the high decay heat soon after the reactor tripped [36]. The management blindly relied on the inert atmosphere and ventilation to prevent hydrogen accumulation and leakage. The operators focussed their efforts on core cooling and pressure venting rather than safely disposing of the hydrogen gas. Even the emergency response procedures did not emphasise hydrogen monitoring outside the PCV, despite it being a possibility. We know that the SDP requires that DID-relevant events should be diagnosable. Then, even if the SDP were implemented retrospectively, the units would not have had features installed to observe the PCV-barrier breach. The hydrogen gas breach was not anticipated in the DID barrier design. We can infer that there is a possibility that certain safety-degrading events/states are not within DID's scope. The SDP should have a broader scope, addressing such unaccounted-for hazardous events and unjustified assumptions. Then it could add safety-relevant information that is truly complementary to DID.

Even if the hydrogen detectors were functional, it is possible that hydrogen gas remained undetected. The PCV has a large complex surface area with several leaking paths. Unknown to anyone, gases may accumulate in hidden pockets and pipes for a long time. There is uncertainty about the diagnosis, as it would depend on the location of diagnosing instruments, their range and operating limits. Additionally, while the global containment pressure may remain below a certain safety level, a higher local concentration sensitive to hydrogen distribution may damage specific containment components, internal walls, and safety equipment [37]. This also affects the reliability and timely availability of the diagnosis. An improper design or poor positioning of diagnosing features can affect the extent to which the SDP can be successfully implemented.

After unit 1's explosion, the operators feared hydrogen explosions in other units. Even in the absence of diagnosability, they logically concluded that hydrogen containment barriers had failed in unit 3, which later turned out to be correct. The likelihood of a high hydrogen level concentration causing an explosion was predicted to be high. The operators were helpless and could not act on this information. The plant personnel did not have access to control equipment, and hydrogen gas ventilation was delayed. The presence of radiation, lack of light source and risk of hydrogen ignition prevented ventilation. Operators were waiting for the arrival of special equipment for cutting holes in the roof and knocking out the panels. Before it arrived, unit 3's building top had exploded. In such a situation, even if the hydrogen state and its barrier failure had been diagnosed due to SDP compliance, it would not have prevented the explosion from happening, due to ill preparedness. Instead, the timely availability of mitigatory measures, to stop the event escalating, would have had a positive effect.

Unit 4 had an unexpected hydrogen explosion, even though it was not operational to produce hydrogen gas. It received hydrogen gas from a reverse flow from unit 3, via the piping arrangement connected to a common vent stack. One design feature which may have prevented or mitigated the migration of hydrogen is backflow dampers, which were not included in the unit 4 venting system design [13, 38]. This is among those scenarios where a robust barrier design, rather than its failure diagnosis, needs to be emphasised. This does not undermine the need to monitor critical barrier states, but we need to compare the SDP's usefulness with mitigatory measures' effectiveness against such hidden hazardous event escalations.

5.4.3 Failure of human-organisational barrier

In Fukushima's case, the failure of the human-organisational barrier and the safety culture played a critical role in the failure of DID. The management of Tokyo Electric Power Company (TEPCO), the nuclear power plant's operating company, did not adopt a strict accident management strategy which could have prevented the simultaneous lack of power availability in all units [13]. Their managers also lacked experience and did not consider the importance of updated risk knowledge. Before the accident, a study had already revealed the likelihood of experiencing a tsunami beyond the Fukushima's handling ability. The organisation ignored the implications of such a study, even though the plant was under-designed. TECPO never addressed the possibility of a prolonged, total loss of power, which led to unpreparedness [38]. The poor safety culture is also visible in the continued use of outdated reactor design, improper placement of emergency generators, compact plant design to reduce land cost, other relaxed safety features, etc.

The Japanese government and regulatory barriers had also weakened. The regulators lacked the power to enforce new requirements emanating from operating experience in other parts of the world. The government had no provisions to manage an extended and widespread loss of power, since they assumed that the power transmission lines would go online quickly. These barriers' failures are difficult to detect and DID does not address them. The failure of these invisible non-technical barriers has more devastating consequences for accident escalation. The SDP lacks guidance on how to monitor the organisational barrier failures; see [26]. It does not add any value in diagnosing these barriers' failure.

5.4.4 Achievement of diagnosability

The SDP's objective is that the implementing agents should develop a system that diagnoses all the safety barrier breaches and delivers this information reliably to the operator. To assess the achievability of this principle, we need to address the uncertainties associated with diagnosability. These uncertainties may arise due to physical limitations, systemic risks, invalid design assumptions, and poor background knowledge. Thy can severely limit the ability to achieve the objective. In other words, targeted actions may have a less than desired effect on the progress towards the objective. For the SDP to satisfy these criteria, we need to evaluate whether diagnosability is actually achievable.

One of the important aspects for achieving the SDP's objective is the reliability of diagnosis feedback. Reliability is associated with multiple aspects such as timeliness, durability, accuracy, precision, etc. In Fukushima's case, negative externality and organisational factors led to a prolonged power interruption. This power blackout was a common cause failure event for the safety barriers and their monitoring instruments. Even though their instruments were reliable, accuracy-wise, they became unavailable and ineffective during hazardous conditions. Likewise, in risky and complex systems, the diagnosing features can simultaneously fail, along with the safety barriers, due to a common failure event (such as a tsunami, in the case of the Fukushima accident). Then, compliance with the SDP may not improve the situational awareness, as it claims. Safety diagnosability, even in the presence of reliable monitoring features, can, in some situations, be difficult to achieve.

In Fukushima's case, we saw that the failure of the reactor cooling barrier could not be confirmed, due to the erratic nature of the safety monitoring instrumentation. It has been commonly observed that instrumentations, while accurate under normal operating conditions, become unreliable under extreme physical conditions. This is due to being exposed to temperature, pressure or

radiation levels that are beyond their safe operating range. It becomes stressful to verify with high confidence whether they are performing their desired functions, when the accident is already quickly escalating.

Some safety barriers may not be completely diagnosable, due to practical limitations. Fukushima's hydrogen leak from the containment vessel into the unit 1 reactor building or the hidden hydrogen leakage from unit 3 to unit 4 are examples of this. Certain operational deviations may remain hidden, despite considerable investment in monitoring features. This can be attributed to factors such as the type of barrier design, its location, nature of hazardous substance, system complexity, and monitoring instrument location.

Overall, there are several uncertainties associated with achieving the SDP's informational benefits. These arguments suggest that the SDP only partially satisfies the achievability criteria.

5.5 Relevancy discussion

Based on the findings from the achievability discussion, there are also reasons to question the relevancy. The SDP's relevance is determined by the value of information its objective provides. Acquiring the information on safety barriers' breach is clearly valuable on a standalone basis, But, when paired with DID, its relevance lies in improving the informed use of DID, which already requires barrier diagnosis. Then, we need to determine whether the SDP-motivated barrier diagnosis is more reliable, of higher quality or holds more real-time value to the operator managing a potential accidental event. If it improves the outcome more than when it is not implemented, its pairing with DID can also be justified economically.

The SDP's maximum informational value or relevance should be observed under accidental conditions, i.e. when the demand arises. Throughout the Fukushima accident sequence, the operators struggled to obtain information on safety barrier status to make accurate diagnoses. As already indicated, even if the SDP had been implemented, it would likely not have made a significant difference in uncovering the information, partly due to a limited scope. This undermines its ability to convey relevant information to improve DID's effectiveness.

The Fukushima accident is considered a man-made disaster, due to the failure of safety culture, management, regulators and government. If the SDP provided guidance on monitoring the weakening of these barriers, such a diagnosis would be material to improving DID's implementation and overall emergency preparedness. Then, it is possible that the accident's outcome could have been different and added business value. However, this is not the case, as the SDP does not address the diagnosability of such non-technical barriers (i.e. human and organisational barriers).

There can be outlier accidental scenarios, when safety diagnosability may not be relevant in bringing the hazardous plant state under control. For example, the Fukushima unit 3 operators could not have made use of the barriers' failure diagnosis, without the capability to act on this information. For a nuclear plant to be prepared for such situations, they need to regularly validate their design assumptions and invest in mitigatory/control measures. In addition, the questionable reliability of diagnoses received during emergency scenarios adds very little value, beyond placing attention to the quality of the information and whether one is compliant to the SDP. Under high-stress and hazardous situations, operators can lose the motivation to follow the SDP. As the SDP takes an extreme safety perspective without consideration for the actual economic benefits for the business, even management may lose enthusiasm for it.

Overall, SDP may be partially relevant, if it requires organisations to invest in its compliance without considering its true costs, benefits and associated uncertainties.

5.6 Timeliness discussion

As mentioned in 2.2, barriers of distinct levels and types are monitorable in different time frames. While the timely availability of diagnosis is undoubtedly critical, the SDP's overall objective is to maintain a superior barrier diagnosability, by making improvements period over period. This makes achieving the SDP's diagnosability an ongoing objective. Quantifying its time horizon is neither realistic nor logical. Therefore, this criterion is not applicable to the SDP and does not provide information about its usefulness.

6 Conclusions

The rationale behind the SDP is that a violation of its requirements increases the probability of an accident conditioned on an initiating event. SDP compliance means that, if situational awareness is degraded during system operation, it can be adjusted appropriately if, or when, the barriers are breached. This prevents the shrinking of the operator response window required to intervene effectively. This motivation is sound but builds on the premise that the information is obtainable.

A main argument for making the SDP information attractive is the insufficiency of the DID principle, but this is perhaps more a question of how DID is managed in this industry. With proper management, one could claim that the SDP would add limited value, as the relevant safety information, corresponding to what would have been provided by the SDP, is already available. It is an argument challenging the benefits of adopting two principles instead of just using DID.

A fundamental part of the DID principle is that, for the barriers to be reliable, management should recognise the importance of monitoring tools to diagnose the barrier and plant status. In particular, the defence layer at level 2 requires that operating experience is sent as feedback and that diagnostic tools record and announce information about faults in the control room. This is implemented by setting up instrumentation and control capabilities over the necessary ranges and through the use of digital technology of proven reliability [39]. This presents an element of redundancy, since diagnosability and feedback fall under DID.

Table 4 summarises the result of the SDP rationality assessment. From section 5, it is concluded that SDP satisfies 'S' and 'M' and partially satisfies the 'A' and 'R', while 'T' is seen as inapplicable to this principle. The principle is clearly specific and measurable. Our discussion on its usefulness to the Fukushima nuclear accident case helped us derive general insights that strengthened the conclusions for 'A' and 'R'. These are important criteria that show that the SDP fails to completely satisfy these practical aspects. These are severe criticisms that can challenge the principle's usefulness, when employed to complement DID in the nuclear context, and it is a finding that can be generalised. This is because a specific and measurable safety principle has only limited usefulness if it is not completely achievable or lacks relevance to the business' safety. The Fukushima case study also shows that restoring the diagnosing capability, as per the SDP, would not have significantly improved the outcome.

Table 4 SDP usefulness assessment result

Rationality criterion	Criterion satisfied
Specificity	Yes
Measurability	Yes
Achievability	Partly
Relevance	Partly
Timely	Not applicable

On a standalone level, however, the situation might be different. It has not been our focus to assess this, and we recommend that future work should consider and conclude on the standalone benefits. We acknowledge that the SDP might show usefulness in combination with DID for some nuclear applications. Our conclusions, based only on this one accident, should not be generalised to cover all nuclear applications. Nevertheless, the SDP gaps pointed to are likely to apply to a wide range of applications, where the principle cannot be fulfilled, and might create a false sense of safety. Hence, we do not, on a general basis, recommend the implementation of the SDP for the nuclear industry.

References

[1] J.H. Saleh, R.A. Haga, F.M. Favarò, and E. Bakolas. "Texas City refinery accident: Case study in breakdown of defense-in-depth and violation of the safety–diagnosability principle in design." Engineering Failure Analysis 36 (2014): 121-133.

[2] J.H. Saleh, K.B. Marais, and F.M. Favaró. "System safety principles: A multidisciplinary engineering perspective." Journal of Loss Prevention in the Process Industries 29 (2014): 283-294.

[3] ISO 12749-5. "Nuclear energy, nuclear technologies, and radiological protection – Vocabulary - Part 5: Nuclear reactors". 2018.

[4] R.L. Boring. "Adapting human reliability analysis from nuclear power to oil and gas applications.", No. INL/CON-15-35411. Idaho National Lab. (INL), Idaho Falls, ID (United States), 2015.

[5] H. Rosencrantz, K. Edvardsson, and S.O. Hansson. "Vision Zero–is it irrational?" Transportation Research Part A: Policy and Practice 41(6) (2007): 559-567.

[6] L.I.K. Sørskår, J.T. Selvik, and E.B. Abrahamsen. "On the use of the vision zero principle and the ALARP principle for production loss in the oil and gas industry." Reliability Engineering & System Safety 191 (2019): 106541.

[7] M. Fackler. "Report finds Japan underestimated tsunami danger". The New York Times. 1 June 2011. Available from: https://www.nytimes.com/2011/06/02/world/asia/02japan.html

[8] International Atomic Energy Agency (IAEA). "Fundamental Safety Principles". Vienna: IAEA. 2006. Available from: file:///C:/Users/jts/Desktop/IAEA%20fundamental%20safety%20principles%20Pub1273_web.pdf

[9] L. Chierici, G. Fiorini, S. La Rovere, and P. Vestrucci. "The evolution of defense in depth approach: A cross sectorial analysis". Open Journal of Safety Science and Technology 06 (2016): 35-54.

[10] United States Nuclear Regulatory Commission (US-NRC). "Historical Review and observations of Defence in Depth." Brookhaven National Laboratory, Upton NY (2016).

[11] ISO 1709. "Nuclear energy - Fissile materials - Principles of criticality safety in storing, handling and processing". 2018

[12] International Atomic Energy Agency (IAEA). "IAEA Safety Glossary - Terminology Used in Nuclear Safety and Radiation Protection". Vienna: IAEA. 2018. Available from: file:///C:/Users/jts/Desktop/IAEA%20safety%20glossary%202018%20PUB1830_web.pdf

[13] International Atomic Energy Agency (IAEA). Nuclear Security Series Glossary. Version 1.3. Vienna: IAEA. 2015. Available from: https://www.iaea.org/sites/default/files/18/08/nuclear-security-series-glossary-v1-3.pdf
[14] International Nuclear Safety Advisory Group (INSAG). "Basic safety principles for Nuclear Power Plants". Safety Series NO. 75-INSAG-3, Rev. 1 INSAG-12, INSAG, Vienna, 1999.
[15] Western Europe Nuclear Regulatory Associations (WENRA), Reactor Harmonisation Working Group (RHWG). "Report- Safety of new NPP designs." WENRA, March 2013.
[16] United States Nuclear Regulatory Commission (USNRC). "NUREG-2150 A Proposed Risk Management Regulatory Framework". US Nuclear Regulatory Commission, Washington, DC. 2012.
[17] C.L. Smith. "Understanding concepts in the defence in depth strategy." In: Proceedings of the IEEE 37th Annual 2003 International Carnahan Conference on Security Technology 2003, pp. 8-16. IEEE, 2003.
[18] J. Ahn, C. Carson, M. Jensen, K. Juraku, S. Nagasaki, and S. Tanaka. "Reflections on the Fukushima Daiichi Nuclear Accident: Toward Social-Scientific Literacy and Engineering Resilience". Springer Nature, 2015.
[19] Tokyo Electric Power Company. "Fukushima nuclear accident analysis report." 2012.
[20] OECD/NEA. "Advanced Nuclear Reactor Safety Issues and Research Needs: Workshop Proceedings", Paris, France, 18-20 February 2002, Nuclear Safety, OECD Publishing, Paris.
[21] International Atomic Energy Agency (IAEA). "Safety of Nuclear Power Plants: Design". IAEA Safety Standards Series No. SSR-2/1, IAEA, Vienna. 2012.
[22] K.N. Fleming, and F.A. Silady. "A risk informed defense-in-depth framework for existing and advanced reactors." Reliability Engineering & System Safety 78(3) (2002): 205-225.
[23] O. Akira. "Where was the weakness in application of defense-in-depth concept and why?" In Reflections on the Fukushima Daiichi Nuclear Accident, (2015): pp. 131-164. Springer.
[24] B. Li, M. Khlif-Bouassida, and A. Toguyéni. "Reduction rules for diagnosability analysis of complex systems modeled by labeled Petri nets." IEEE Transactions on Automation Science and Engineering 17(2) (2019): 1061-1069.
[25] J. Reason. "Managing the risks of organizational accidents". Vermont: Ashgate; 1997.
[26] E. Bakolas, and J.H. Saleh. "Augmenting defense-in-depth with the concepts of observability and diagnosability from Control Theory and Discrete Event Systems." Reliability Engineering & System Safety 96(1) (2011): 184-193.
[27] A. Paoli, and S. Lafortune. "Safe diagnosability for fault-tolerant supervision of discrete-event systems." Automatica 41(8) (2005): 1335-1347.
[28] K. Edvardsson, and S.O. Hansson. "When is a goal rational?" Social Choice and Welfare 24(2) (2005): 343-361.
[29] G.T. Doran. "There's a SMART way to write management's goals and objectives." Management Review 70(11) (1981): 35-36.
[30] J.T. Selvik, S. Bansal, and E.B. Abrahamsen. "On the use of criteria based on the SMART acronym to assess quality of performance indicators for safety management in process industries". Submitted for possible publication in an international journal (2020).
[31] Japan Meteorological Agency (JMA). "Information on the 2011 off the Pacific Coast of Tohoku Earthquake". 2015. Available from: http://www.jma.go.jp/jma/en/2011_Earthquake/Information_on_2011_Earthquake.html
[32] International Atomic Energy Agency (IAEA). "The Fukushima Daiichi Accident, Technical Volume 1/5. Description and Context of the Accident". 2015.
[33] G. Brumfiel. "Fukushima reaches cold shutdown." Nature News (2011).
[34] K. Kurokawa. "Fukushima nuclear accident independent investigation commission by the National Diet of Japan." Nippon Genshiryoku Gakkai-Shi 55(3) (2013): 146-151.
[35] M.P. Cabasino, A. Giua, and C. Seatzu. "Diagnosability of discrete-event systems using labeled Petri nets." IEEE Transactions on Automation Science and Engineering 11(1) (2013): 144-153.
[36] G. Saji. "Root cause study on hydrogen generation and explosion through radiation-induced electrolysis in the Fukushima Daiichi accident." Nuclear Engineering and Design 307 (2016): 64-76.

[37] U. Bielert, W. Breitung, A. Kotchourko, P. Royl, W. Scholtyssek, A. Veser, A. Beccantini et al. "Multi-dimensional simulation of hydrogen distribution and turbulent combustion in severe accidents." Nuclear Engineering and Design 209(1-3) (2001): 165-172.
[38] International Atomic Energy Agency (IAEA), and the World Meteorological Organization. Flood Hazard for Nuclear Power Plants on Coastal and River Sites, IAEA Safety Standards Series No. NS-G-3.5, IAEA, Vienna (2003). (This publication is superseded by SSG-18 (2011)).
[39] International Nuclear Safety Advisory Group (INSAG). 1996. “Defence in Depth in Nuclear Safety”. INSAG Series No. 10. Vienna: IAEA. 1996. Available from: https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1013e_web.pdf