

Validation in Risk and Hazard Analysis in Safety-Critical Industries: Understanding
Current Practices to Develop a Framework for STPA Validation

by

Reyhaneh Sadeghi

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
May 2023

Dedication Page

*This thesis is dedicated to the strong and brave women of Iran
For #Woman_Life_Freedom*

TABLE OF CONTENTS

LIST OF TABLES	v
LIST OF FIGURES	vi
ABSTRACT	vii
LIST OF ABBREVIATIONS USED	viii
ACKNOWLEDGEMENTS	ix
CHAPTER 1: Introduction	1
1.1 Background and Motivation.....	1
1.2 State of the Art	3
1.2.1 State of the Art in Risk and System Safety Validation	3
1.2.2 State of the Art in STPA and STPA Validation.....	6
1.3 Objectives and Structure of the Thesis.....	11
1.4 Definition	13
CHAPTER 2: Research Methods.....	17
CHAPTER 3: Results	26
3.1 State of the Practice in Validation of Risk and Safety Approaches	26
3.1.1 An Empirical Study on the Validation of Academic Model-based Safety Analysis in Socio-Technical Systems	26
3.1.2 An Empirical Study on the Validation Practices of Hazard Analysis Techniques in Safety-Critical Industries	29
3.2 Towards a Formal Validation Approach for STPA.....	32
3.2.1 Theory-Based Framework	33
3.2.2 Empirical Confirmation of the Proposed Framework.....	42
CHAPTER 4: Discussion.....	46
4.1 Further Improvements in the General Concept of Risk and Hazard Analysis Validation	46

4.2 Further Improvements Related to the Proposed STPA Validation Framework..	48
CHAPTER 5: Conclusion	52
REFERENCES	54
APPENDIX.....	66

LIST OF TABLES

Table 1. The details of the research method for RQ 1	19
Table 2. The details of the research method for RQ 2	20
Table 3. Demographics of the interviewees	24
Table 4. Definition of the validation approaches identified in PI.....	26
Table 5. The list of proposed tests with their brief definitions, adopted from PIII.	37

LIST OF FIGURES

Figure 1. The overall structure of the thesis.....	13
Figure 2. Distinction between validation and efficacy	15
Figure 3. Definition of Validation in Risk and Hazard Analysis.....	15
Figure 4. The process of selecting papers adopted from PIII	22
Figure 5. The process of constructing the STPA validation framework based on PRISMA flow diagram, adopted from PIII.....	23
Figure 6. Further empirical research on the proposed STPA validation framework	24
Figure 7. The distribution of articles in terms of the adopted validation approach, for cases where validation is performed, adopted from PI	27
Figure 8. Distribution of papers in terms of the terminology used for validation, adopted from PI	29
Figure 9. The assigned tests to each element of STPA, adopted from PIII	39
Figure 10. The example of the proposed tests for Step 3 of STPA	40
Figure 11. Using the STPA validation framework in parallel with STPA implementation, adopted from PIII	41
Figure 12. Performing validation using the STPA validation framework after STPA implementation, adopted from PIII.....	42
Figure 13. Interviewees' experiences and opinions on each test, adopted from PIV	45
Figure 14. Summary of the future research directions.....	51

ABSTRACT

In recent years, scholars have increasingly delved into the theoretical foundations and issues surrounding risk and safety science. However, validation has not received much explicit attention although it has been highlighted as an important focus theme.

To contribute to closing this gap, this thesis first explores the current state of practice in risk and hazard analysis validation in both academic works and in the context of safety-critical industries. Two empirical research studies are performed to understand current validation practices, understand the extent to which validation takes place, to identify frequently used approaches, and to uncover challenges and directions for improvements. The findings suggest that validation is not a common practice among researchers, and a lack of clear guidance on how to perform validation makes it a challenging task for practitioners.

Then, the thesis proposes a formal validation framework for the Systems-Theoretic Process Analysis technique (STPA). This technique is selected as it has been identified as one of the few techniques capable of capturing the tenets of a systems view on accident causation and because it has gained increasing popularity in recent years. The framework aims to support a systematic assessment of the analysis's comprehensiveness, accuracy, and credibility using 15 proposed tests. The framework is based on theoretical validation concepts in related fields of study. It is recognized that the proposed framework should be further tested to confirm its practical usefulness.

This leads to the final issue addressed in this thesis. An evaluation of the proposed framework is accomplished through an interview-based study with STPA experts, who provide feedback on the individual tests comprising the framework and on its theoretical underpinnings. The experts appreciate the framework in that it provides clear guidance on how to validate each step of an STPA analysis systematically and find some additional theory-based tests interesting for consideration in practice.

Additionally, the thesis provides a comprehensive discussion of future research directions for validation of risk and hazard analysis techniques, such as developing a modular framework with associated guidance, tailored to a specific practical context, and integrating the proposed framework into the overall process of risk and safety analysis.

LIST OF ABBREVIATIONS USED

CAS	Credibility Assessment Scale
CAST	Causal Analysis Based on Systems Theory
DOI	Digital Object Identifier
FMEA	Failure Mode and Effect Analysis
NASA	National Aeronautics and Space Administration
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QRA	Quantitative Risk Assessment
SME	Subject Matter Expert
SRA	Society for Risk Analysis
STAMP	Systems-Theoretic Accident Model and Processes
STECA	Systems-Theoretic Early Concept Analysis
STPA	System Theoretic Process Analysis
UCA	Unsafe Control Action
V&V	Validation and Verification

ACKNOWLEDGEMENTS

First and foremost, I must thank my supervisor, Professor Floris Goerlandt, who trusted my abilities, accepted me to this program, and provided me with this great opportunity. His tremendous support and guidance were instrumental in helping me navigate through this process successfully. I am extremely grateful for the time he devoted to reviewing my drafts and providing invaluable feedback. Without his mentorship, none of this would have been possible.

I would also like to extend my thanks to Professor Ronald Pelot and Professor Paul Amyotte, my committee members, for their encouragement and guidance throughout my research journey. Their expertise and constructive criticism have been indispensable in shaping the outcome of my work. I would also like to express my gratitude to Dr. Ioannis Dokas for accepting to be the external examiner of this thesis despite his many other commitments and activities. Additionally, I am thankful to the interviewed experts who contributed to this thesis by generously sharing their time, knowledge, and experiences.

Special thanks go to my husband, Kia, for being so patient and for all his pep talks that kept me motivated. Thank you for always giving me a reason to smile, and for making the tough days of doing a Ph.D. during the pandemic bearable.

Last, but certainly not least, I would like to express my gratitude to my parents and sisters, Marjan and Melika, for their unconditional love and support throughout my studies. Their encouragement and unwavering faith in my abilities have been a constant source of strength, and I cannot thank them enough for all that they have done for me.

CHAPTER 1: Introduction

1.1 Background and Motivation

In recent years, scholars have increasingly focused on foundational issues in risk and safety science. Examples of such issues that have been studied and discussed include the concept and definitions of risk (Aven & Zio, 2014), safety (Alpeev, 2019; Aven, 2014; Hale, 2014; Hansson, 2012; Hollnagel, 2014), and plausibility (Glette-Iversen et al., 2022), prediction (Goerlandt & Reniers, 2018), accident theories (Saleh et al., 2010), credibility in risk and safety context (Busby & Hughes, 2006), and evaluation of system safety (Rae et al., 2010). One important area of study, which has not received much attention, is validation (Goerlandt et al., 2017b; Hale, 2014).

Although a lack of focus on validation in risk and safety science has been raised by some researchers (Aven, 2017; Goerlandt et al., 2017b; Rosqvist, 2010), there has been insufficient empirical research devoted to such validation in both academic and industrial settings. Risk analysis validation has been the focus of a call for further research (Goerlandt et al., 2017a), where it is highlighted that more extensive research is required in this field.

This lack of focus on validation has two major implications: (1) lack of knowledge on the quality and credibility of a given analysis performed (Goerlandt et al., 2017b; Rae et al., 2010); and (2) lack of evidence that an analysis effectively contributes to enhancing the system safety (Rae et al., 2012; Rae & Alexander, 2017). The former concerns how well an analysis is performed while the latter pertains to whether the analysis's effect on safety is actually positive (Hale, 2014; Rae et al., 2010).

This thesis focuses on the first implication of the lack of focus on validation. That is, how to ensure that an analysis is performed sufficiently well so that it generates comprehensive and accurate results about hazards, risks, or safety (depending on the analysis focus). A lack of evidence concerning the analysis's comprehensiveness and accuracy could further lead to a lack of credibility in the analysis.

In a manifesto by Rae et al. (2020), the authors highlight that a lack of empirical research and reality-based practices have resulted in the stagnation of safety science. According to

them, reality-based safety science is based on a “virtuous cycle of studying current practice to advance theory and applying theory to advance current practice.” Thus, it is crucial to gain a comprehensive understanding of the existing practices rooted in the real-world and then relying on theoretical ideas to prescribe changes to improve practice.

Aven (2014) also argues for the need for different types of safety and risk research. He classified different types of research in safety science into six categories, types A to F. For instance, type C is the “evaluation of/considerations related to specific methods and models.” Aven argues that such a work is essential to advance the scientific understanding of concepts, theories, methods, and approaches. He highlights that a scientific field can thrive only when academic communities commit to such conceptual work while stressing that (as it is done in this thesis) such work should be suitably accompanied by subsequent phenomenological/empirical research (type D).

Furthermore, the scarcity of available empirical information on the validation of risk and hazard analysis work in safety-critical industries is a key motivation to study the state of practice among practitioners. As previously noted, there exists a significant gap between academic safety science research and industry experience (Le Coze, 2019), which has led many practitioners to rely solely on the latter and neglect scientific evidence (Provan et al., 2019). Improved knowledge about this can also contribute to diminishing the gap between academic safety science research and the actual work of safety practitioners (Reiman & Viitanen, 2019).

Given the emphasis on the importance of empirical research and the need to narrow the gap between research and industry, combining empirical and theoretical work is valuable to contribute to closing the gap in risk and safety validation. This is done in this thesis, first, through an investigation of the state of the practice in validation of risk and safety approaches among both researchers and system safety practitioners. Such an empirical study can help to understand the extent of the issue of the lack of focus on validation and further lead to the understanding of the issues, challenges, and potential improvements in risk and safety analysis validation.

Eckerd et al. (2019) suggest that despite surface-level differences in validation practices among different academic communities, common underlying principles, and concerns

create an opportunity for knowledge exchange. Thus, based on the insights gained through an empirical study as well as the theoretical ideas in closely related fields of study, solutions can be provided to the identified challenges to improve the current state of the practice in the validation of risk and safety approaches. Consistent with the ideas in the manifesto by Rae et al. (2020), empirical work can be done to test the reasonableness of the developed ideas rooted in the theoretical concepts.

1.2 State of the Art

1.2.1 State of the Art in Risk and System Safety Validation

Validation has garnered some interest among scholars in the fields of risk and safety science, but there is relatively little work dedicated explicitly to the validation of safety and risk analyses and its underlying techniques. This lack of direct focus on validation is compounded by a lack of clear terminology, resulting in a situation where different authors work on issues related to validation while using the same terms for somewhat different concepts and purposes.

As explained in Section 1.1, validation concerns quality, i.e. how to ensure a good analysis is developed, and credibility, i.e. how to ensure that the stakeholders can trust the results of an analysis. This can further result in the analysis efficacy, i.e. how to ensure that the analysis can lead to a safer system (for the definitions of these concepts refer to Section 1.4). In this section, it is aimed to explain the previous research studies related to risk and safety analysis validation; with an emphasis on how their approaches to validation are different from the approach taken in this thesis.

Lathrop and Ezell (2017) present an insightful systems perspective on how to establish validity in risk analysis. Their approach involves a flowchart that links all the elements from inputs through risk analysis, risk reporting, and transparency. They discuss how reporting and transparency support the decision-making process of risk management, as well as third-party and stakeholder reviews, formal or informal, that determine the trust and acceptance necessary for the implementation of risk management actions. Within the flowchart, they identify sixteen critical elements and specify a validation test for each element. Validation, therefore, requires applying these sixteen tests to the risk analysis.

These authors' ultimate goal with validation is to ensure that analysis effectively supports the decision-making process of risk management, taking into account the role of trust and stakeholder review and acceptance. This is closely related to the credibility aspects of validation as understood in this thesis. They emphasize that completeness is also crucial and that failing to explicitly assess the completeness and report shortcomings and uncertainties can significantly affect the effectiveness of a risk analysis. The authors have identified three sections for completeness testing: completeness of scenario initiation, completeness of scenario unfolding, and explicit assessment of completeness. As such, their approach covers both the completeness and credibility aspects of validation.

Rouhiainen (1992) highlights that the main objective of a safety analysis is to support decision-making and that the applicability of a safety analysis depends on its quality. They consider four main areas in order to cover the quality aspect of an analysis: the analysis's ability to identify hazards and contributing factors, the accuracy of the estimated risks of an analysis, the effectiveness of the introduced remedial measures, and the cost-effectiveness of an analysis. To this end, he proposes a method for assessing the quality of safety analysis, in which the analysis process is evaluated, and its deficiencies are identified through a checklist. In light of this, his work focuses on one aspect of how validation is considered in this thesis, namely the quality dimension.

Rae et al. (2010) highlighted the lack of focus on validation in safety science by exploring the use of evaluation in system safety research. They noted that although safety engineering research generates numerous plausible and potentially beneficial ideas, very few of these ideas are academically tested for efficacy. They analyze a set of research articles using a classification scheme based on the knowledge (guidance or observation) and evaluation components (whether evaluation happened) of the papers. Their analysis revealed a significant disparity between a small group of observational research papers with robust evaluation and a large group of papers providing guidance which has not been evaluated. According to them, a technique may enhance safety, but without evidence of its efficacy, claims about its effectiveness should, from a scientific perspective, be met with skepticism. Thus, this study is concerned with the effectiveness aspect (refer to Section 1.4 for the definition of effectiveness and how it is different from validation).

Goerlandt et al. (2017b) presented a review focusing on the validity and validation of safety-related quantitative risk analysis. The authors classified theoretical contributions regarding the validity of Quantitative Risk Analysis (QRA) into three categories:

1. **Conceptual**, which pertains to the condition where the operationalisation of a concept measures what it intends to measure.
2. **Foundational**, which relates to different perspectives on validation in relation to the scientific foundations adopted in a risk analysis.
3. **Pragmatic**, which concerns the condition where a method satisfies the intended requirements in terms of the results achieved.

In this research five categories are proposed for the pragmatic validity following research by Suokas (1985). These categories are reality check (comparing the results of a model or a part of the model with real-world data), peer review (examining the model by technical expert's opinion according to a set of criteria), quality assurance (examining the process behind the analysis), and complete and partial benchmark exercise (comparing the result of a model with a parallel analysis either partially (a part of the model scope) or completely (full model scope)). While Goerlandt et al. (2017b) primarily focused on the quality and credibility aspects, they highlighted the need for further research to determine the effectiveness of the proposed validation methods. These authors also highlight the importance of building approaches to validation on explicitly defined foundations for risk analysis, as different views on, for example, what risk is as a concept, will have important implications for how risk analysis can be validated.

Aven and Heide (2009) investigated to what extent risk analysis fulfills the scientific quality requirements of validity, for a set of commonly used risk perspectives in quantitative risk analysis contexts. In their research, validation is defined as “the degree to which the risk analysis describes the specific concepts that one is attempting to describe.” They used four sub-criteria for the validity of risk analysis and discussed to what extent these are met in light of different perspectives of risk. These criteria are:

V1: The degree to which the produced risk numbers are accurate compared to the underlying true risk.

V2: The degree to which the assigned subjective probabilities adequately describe the assessor's uncertainties of the unknown quantities considered.

V3: The degree to which the epistemic uncertainty assessments are complete.

V4: The degree to which the analysis addresses the right quantities.

Their findings revealed that while quantitative risk analysis satisfies some of the basic scientific requirements, the validity requirements are generally not met. It should be noted that based on the defined criteria (V1 to V4), their focus is mainly on the accuracy and comprehensiveness aspects and it does not include the credibility aspect.

Despite the focus in the above-mentioned research studies on validation in a risk and safety analysis context, no research has proposed a specific method for establishing pragmatic validity using a formalized approach. Furthermore, the term "validation" is used to address different issues in these studies, or they may only focus on some elements of validation, rather than the approach presented in this thesis. This gap in the literature highlights the need for a formalized validation technique that can evaluate the completeness, accuracy, and credibility of an analysis.

1.2.2 State of the Art in STPA and STPA Validation

Through the performed empirical studies on risk and hazard analysis validation in publications PI and PII, it is found that there is a lack of a formalized validation framework for specific methods. Thus, as explained in Section 1.3, the main objective of this thesis is to develop a formalized validation framework for a specific technique. Given the broad range of analysis techniques (Ericson, 2015), a specific technique needs to be selected for which a validation framework can be developed.

As also highlighted by Gass (1983), one validation framework would not work for all analysis techniques. This is because each technique has a unique implementation process and different underlying foundations. Accordingly, the validation process should be tailored to fit the specific implementation process and foundations of the technique being used. For instance, developing a safety control structure is a step specific to the STPA technique, which is not performed in other hazard analysis techniques. Thus, specific consideration should be given to the safety control structure in the validation process.

In this thesis, the scope is limited to the STPA technique. A practical reason for choosing STPA is that it has gained increasing popularity for hazard analysis with application in different industries, with significant interest in the technique especially in aviation, maritime, automotive, and healthcare industries (Patriarca et al., 2022). Significantly, a theoretical analysis of the adequacy of various risk assessment methods in light of the tenets of accident causation in socio-technical systems according to Rasmussen's (1997) systems risk framework, also highlights STPA as one of the few currently available techniques which align with a systems view on accident causation (Dallat et al., 2019).

Nevertheless, some limitations of STPA have been identified which need to be addressed to facilitate a wider application in industry or to be widely recommended by regulatory authorities. Lack of formalism (Dakwat & Villani, 2018), dependence on available information and those who perform it (Harkleroad et al., 2013), its time-consuming nature (Patriarca et al., 2022), and use of abstraction for managing the complexity of a system (Baybutt, 2021) are some of the limitations of STPA, as currently applied. These limitations make the validity of STPA a debatable issue.

This section is further divided into Sections 1.2.2.1 and 1.2.2.2. The former briefly explains what STPA is and how it is different from traditional techniques, and the latter explains the state of the art of research in STPA validation.

1.2.2.1 STPA Theoretical Foundations and Implementation Steps

Traditional hazard analysis techniques or event-based models consider accidents as a result of a chain of events, which almost always involve some type of component failure (Leveson, 2004a). Leveson states that in such models, as linear causality relationships are emphasized, it is difficult to include non-linear relationships between components of a system, such as feedback. Accidents can also happen from dysfunctional interactions among system components, and it is not just limited to component failures (Dallat et al., 2019; Leveson, 2012; Rasmussen, 1997). In other words, according to this view on accident causation, the absence of appropriate control to impose required constraints on component interactions can lead to accidents (Leveson, 2004a). Thus, detecting causal factors that are not related to linear interactions between system components are challenging using traditional techniques, which can result in an inadequate identification of hazards.

Systems-Theoretic Accident Model and Processes (STAMP) is an accident causality model based on control system theory (Leveson, 2012). STAMP has three main components. First, as opposed to the traditional methods, STAMP considers safety a control problem, meaning that safety is achieved by imposing constraints on the system's behavior, rather than only preventing failures (Leveson, 2012). Hence, not only component failures but also accidents resulting from component interactions are considered in the analysis (Leveson, 2004a). Second, STAMP considers systems as hierarchical structures where each level controls the activity of the level beneath it (Leveson, 2004b). To determine the required control action, a process model is constructed and used (Fleming et al., 2013), which is the third component of STAMP. A process model is defined as a representation of the state of the controlled processes, which are kept updated through feedback control loops (Leveson, 2017).

The two most widely applied STAMP-based tools are System Theoretic Process Analysis (STPA) and Causal Analysis based on Systems Theory (CAST). STPA is a proactive analysis method to identify hazards in the design of a new or already operational system while CAST is a retroactive method used to understand the systemic factors involved in accidents during accident investigation processes.

STPA was conceived by Nancy Leveson in the early 2000s (Dakwat & Villani, 2018). The STPA method for hazard analysis focuses on analyzing the dynamic behavior of systems and is intended to provide advantages over traditional hazard analysis methods (Leveson, 2012). In this method, the potential causes of accidents are proactively analyzed so that they can be removed or controlled. This proactive analysis can happen in any system lifecycle from concept development to operation (Leveson & Thomas, 2018). In STPA, component failures are still included; however, the cause of accidents is extended to include component interactions, as well. Thus, accident prevention requires identifying and eliminating or mitigating unsafe interactions among the system components (Leveson, 2004a). According to Leveson & Thomas (2018), STPA is applied in the following four steps:

1. Defining the purpose of the analysis. STPA starts with specifying the purpose of the analysis, which has three main steps: (i) identifying losses; (ii) identifying system-level

hazards; and (iii) identifying system-level constraints. These steps form the foundations of the analysis, in which basic elements of the analysis, such as assumptions and system boundaries, are also specified.

2. Building a safety control structure. STPA relies on a process model of the system, called the safety control structure, which consists of components of a system and their functional relationships with feedback control loops. Different components of a control structure are controllers, control actions, feedback, and controlled processes. A controller issues control actions on a controlled process based on a control algorithm or procedure, that represent the controller's decision-making process and its underlying process models, i.e., the beliefs serving as a basis for those decisions. The development of the control structure is a critical step in STPA since it is used as a guide for identifying and mitigating Unsafe Control Actions (UCAs).

3. Identifying Unsafe Control Actions (UCAs). This step of STPA consists of determining how the controlled system can get into a hazardous state which can further lead to accidents/losses. The control actions are reviewed to investigate how they can, in a particular context and worst-case environment, lead to a hazard. Controllers may issue UCAs by: (i) not providing the control action; (ii) providing the control action; (iii) providing a potentially safe control action but too early, too late, or in the wrong order; (iv) providing the control action that lasts too long or stops too soon. Once UCAs and their causal factors have been identified, they are translated into constraints on the behavior of each controller.

4. Identifying loss scenarios. In this step, the potential causes of the identified UCAs in the previous step are determined. For instance, scenarios are developed to explain how unsafe controller behavior and inadequate feedback and information can lead to UCAs. In addition to hazards that can occur through UCAs, hazards can also be caused by not executing or improperly executing a control action. Therefore, all these loss scenarios are investigated and elaborated.

When all these steps are carried out, and the scenarios are identified, they can be used to, for instance, create additional requirements, identify mitigations, and make recommendations for improvement (Leveson & Thomas, 2018).

1.2.2.2 State of the Art on Research in STPA Validation

In a recent review article on STAMP/STPA/CAST by Patriarca et al. (2022), STPA validation has been raised as an important issue that has been missing to a great extent from the reviewed papers. Articles have been published in which the results of STPA are compared with other hazard analysis methods through a case study to determine their comparative merit, i.e. benchmark exercise. For instance, Sulaman et al. (2019) qualitatively compared the results of Failure Mode and Effect Analysis (FMEA) and STPA methods using a case study research methodology to compare the effectiveness of the methods and investigate their differences. Their results show that FMEA and STPA deliver similar analysis results.

Another example of a benchmark exercise is research by Hulme et al. (2022), through which the validity of STPA has been empirically tested. They studied the criterion-referenced concurrent validity of three systems-based methods, one of which is STPA. To achieve this, a test-retest study design is employed, utilizing the knowledge and expertise of 30 professionals in human factors, ergonomics, and safety science. The findings of this research indicate a weak to moderate level of reliability and validity for these techniques. They suggest that employing systems-based risk assessment approaches in the future is beneficial, provided that methodological improvements are implemented to bolster their reliability and validity.

Some attempts have been made to propose a formal verification approach for STPA. For example, Dakwat and Villani (2018) propose the use of STPA combined with model checking, which is a formal verification technique, to overcome the potential dependence of the results on the experience of the analysis team. The proposed model checking provides a formal representation of the system of interest and the identified threats through STPA analysis. They concluded that the proposed approach enhances the rigor and formalism of STPA, making it more reliable and effective for hazard analysis.

In addition, reviews by independent experts have been proposed as a way to improve the results of an STPA analysis, such as confirmation of the assumptions (Thomas et al., 2012). Harkleroad et al. (2013) conduct a technical report that evaluates STPA suitability for risk-based modeling of complex NextGen concepts. The study employs the National

Aeronautics and Space Administration (NASA) Standard for Models and Simulations (M&S) to assess the outcome of STPA. This standard utilizes a Credibility Assessment Scale (CAS) that considers eight factors, including Validation and Verification. To validate their findings, stakeholder and subject-matter expert reviews are conducted on the completed STPA outputs. The authors also provided recommendations for improving STPA, one of which is to compare system behavior with and without enforcement of STPA-generated safety constraints. This approach would confirm whether the identified safety constraints can effectively prevent hazards and accidents.

Having discussed the previous work related to STPA validation, it should be highlighted that none of the above-mentioned studies suggests a systematic, formalized validation approach for a particular STPA analysis, which covers all elements of an STPA process. STPA validation should not be limited to the results of STPA but should be expanded to different elements in an analysis, such as execution steps and assumptions. Validation needs to provide evidence about how well an entire analysis has been performed, that is whether it is comprehensive and accurate, and to what extent the performed analysis is credible (for a clear definition of validation, the way it is intended in this thesis in the context of the STPA validation framework, please refer to Section 1.4). Thus, to the best of the authors' knowledge, there has not been any work specifically focusing on proposing a comprehensive formalized framework to systematically approach the validation of STPA in academia or industrial contexts.

1.3 Objectives and Structure of the Thesis

This thesis first empirically investigates the current state of validation practices in risk and hazard analysis in both academia and industry, finding a lack of formalized validation approaches for specific techniques. Based on the finding of empirical research studies and theoretical foundations in related fields of study, a theoretical validation framework for a specific hazard analysis technique is proposed, namely STPA (see Section 1.2.2). Thus, this thesis has two high-level research objectives, as follows:

Research Objective 1: Investigate the current state of practice in validation of risk and hazard analysis techniques in academia and industry.

Research Objective 2: Propose a validation framework for the System Theoretic Process Analysis (STPA) technique, based on foundational concepts in risk analysis and prior theoretical work on validation in closely related disciplines.

The first research objective leads to the formulation of Research Questions (RQs) 1 and 2, addressed in publications PI and PII, respectively. The second research objective results in three research questions (RQs 3 to 5), with RQ 3 addressed in publication PIII, and RQs 4 and 5 both addressed in publication PIV. The publications can be found in the Appendix section of the thesis. The research questions are listed below:

Research Question 1: What is the state of practice in validation of model-based safety analysis in socio-technical systems in academia? (PI)

Research Question 2: What is the state of practice in hazard analysis validation for system safety among system safety practitioners in safety-critical industries? (PII)

Research Question 3: What is a suitable validation framework for the System Theoretic Process Analysis (STPA) technique? (PIII)

Research Question 4: To what extent are the ideas and tests in the proposed STPA validation framework in line with what experts do in practice? (PIV)

Research Question 5: Is the proposed STPA validation framework reasonable from the STPA experts' point of view? (PIV)

Figure 1 illustrates the relationships between the objectives, research questions, and publications (PI to PIV). The first two RQs aim to evaluate the scope of the problem regarding the insufficient focus on validation, and explore ways to enhance the situation, consequently laying the groundwork for future advancements. Based on the challenges and potential improvements resulting from these RQs, mainly the lack of clear guidance on how to perform validation in practice, these two RQs lead to the proposal of a structured approach for STPA validation. It is aimed to develop a well-grounded theory-based framework, recognizing that in practical applications, this may need to be modified. This also leads to the fourth and fifth RQs through which the reasonableness of the proposed framework is assessed.

It should be highlighted that the scope of this research has been narrowed down moving from RQ 1 to RQ 5. RQ 1 investigates model-based safety analysis as a general concept, while RQ 2 centers on hazard analysis techniques within the realm of model-based safety analysis. The last three RQs (RQs 3 to 5) concentrate on STPA, a specific hazard analysis technique.

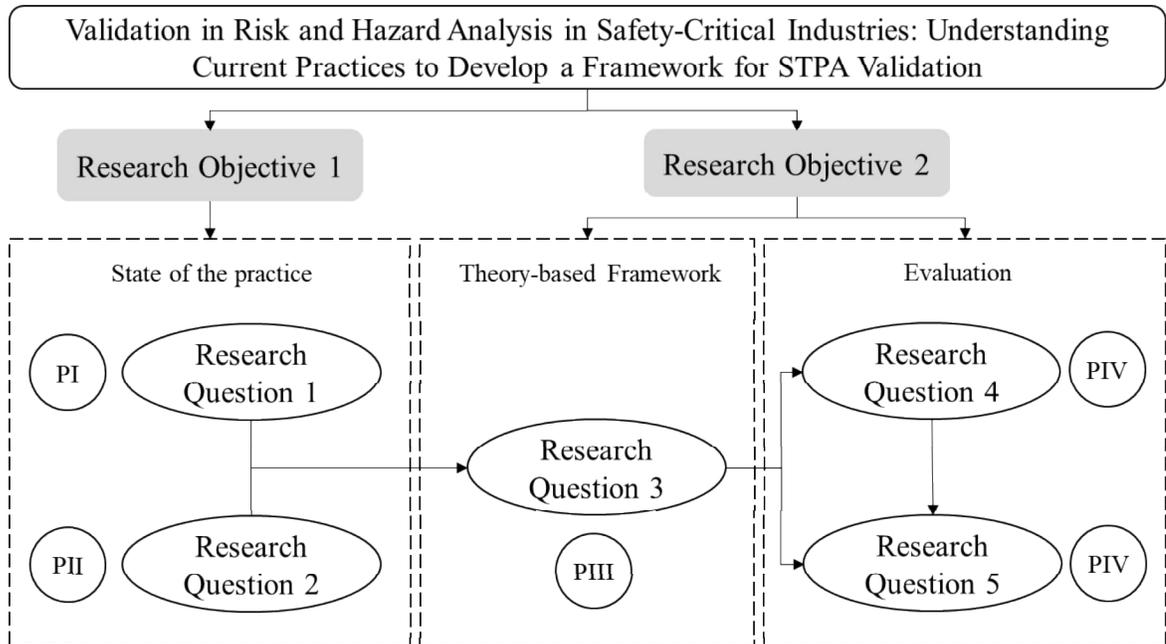


Figure 1. The overall structure of the thesis

1.4 Definition

One of the discussed issues in the field of validation and indeed a finding in PI and PII is the inconsistency in the terminologies used. There is no consensus on the words used for validation and there is a set of words that have been used interchangeably in the literature. Some of the words used in the literature are validation, evaluation, verification, comparison, effectiveness, usefulness, and trustworthiness. This issue is not limited to the validation in a safety context, and it is well known that this kind of terminological inconsistency is common in safety (Alpeev, 2019), risk (Aven & Zio, 2014), and validation research (Oberkampf & Trucano, 2008). This results in a somewhat chaotic situation (Aven, 2012), which has been and continues to be a problem (Goerlandt et al., 2017b).

According to Augusiak et al. (2014), one of the major obstacles to a sound understanding of what model validation is, how it works, and what it can deliver is unclear terminology. Thus, one of the challenges in the validation process is distinguishing the validation concept from its associated terms. It is imperative to provide a clear definition for the term validation, specifically in the context of risk and hazard analysis.

In this thesis, an open-minded approach is taken first, to comprehend how validation has been approached and defined by researchers and practitioners through Research Objective 1 (Section 1.3). Subsequently, in Research Objective 2, a precise definition of validation is presented in the context of an STPA analysis. This definition is grounded in safety science literature and draws upon the concept of safety of work vs. work of safety, as described by (Rae & Provan, 2019). This concept is explained later in this section.

For the sake of clarity throughout this dissertation, it is imperative to distinguish between two concepts: validation and effectiveness. As shown in Figure 2, the concept of validation comprises two major components: quality and credibility, each addressing a completely different issue. The first component, quality, concerns how to ensure that a good analysis is developed. The second component, credibility, concerns how to ensure that the stakeholders can trust the results of an analysis (Sargent, 2013).

The second concept, effectiveness concerns whether the analysis indeed leads to a better and safer system. As mentioned in Section 1.2.1, there is a lack of ample evidence that supports the effectiveness of the current risk and hazard analysis techniques (Goerlandt et al., 2017b; Rae et al., 2010). Having a valid analysis could be one element of providing evidence of effectiveness. This is shown in Figure 2 by adding a dashed arrow from validation to effectiveness.

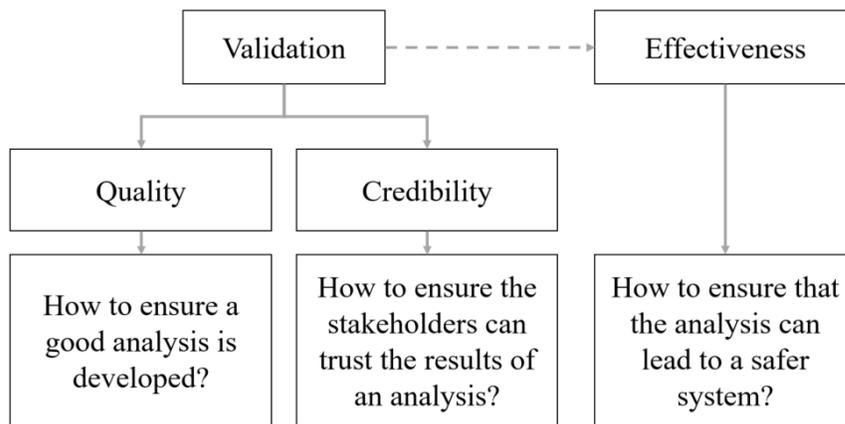


Figure 2. Distinction between validation and efficacy

The quality component of validation further falls into comprehensiveness and accuracy. The former concerns the adequacy of the scope, assumptions, implementation steps, and results of an analysis with respect to the stated purpose of the analysis (Goerlandt et al., 2017b). The latter focuses on assessing whether the analysis and its results are correct and free of errors (Sargent, 2013).

In light of the above definitions for the quality and credibility components, validation in the context of hazard analysis, in general, and STPA, in particular, is taken to mean the process of ensuring that a hazard analysis is accurate, comprehensive, and credible. These are called the functions of the proposed STPA validation framework (please refer to Section 3.2), which are illustrated in Figure 3. This definition is also provided in Appendix A of PII.



Figure 3. Definition of Validation in Risk and Hazard Analysis

These functions each can be linked to different types of safety work. Safety work, as defined by Rae & Provan (2019), encompasses activities carried out in the pursuit of safety and is divided into four aspects: social, demonstrated, administrative, and physical safety work. The comprehensiveness and accuracy functions are at the core of the proposed STPA validation framework. These primarily aim to support safety-related decisions (a type of administrative safety work) and ultimately lead to physical safety work (i.e., operational work which would not occur if not for safety concerns). The credibility function is primarily concerned with demonstrated safety (showing safety to stakeholders) as it deals with ensuring that stakeholders can trust the results of an analysis. Together, these functions can contribute towards the overall safety of work, which means the absence of harm arising from operational work (Rae & Provan, 2019).

CHAPTER 2: Research Methods

In order to address the overarching research objectives posed in Section 1.3, a combination of empirical and theoretical research methods is employed. Specifically, to investigate Research Objective 1, a range of empirical methods are used to gain insight into the current state of risk and safety analysis validation practices in the both academic and industrial contexts. The goal of this research is to explore the extent to which validation is performed by system safety researchers and practitioners, identify commonly used validation approaches, understand organizational challenges to perform validation, and potential improvements. The use of an empirical research design aligns with Rae et al.'s (2020) view that a comprehensive understanding of existing practices in the real world is crucial before venturing into theoretical concepts to suggest changes or new approaches.

To gain a deeper understanding of the state of the practice in academia and industry, Objective 1 is further elaborated into two RQs, RQs 1 and 2 (Section 1.3). These RQs seek to take advantage of the strengths of different empirical research methods to provide comprehensive insights. To obtain data-based insights for the academic context (RQ 1), document analysis is employed. The analysis of previously published research studies can help in synthesizing research findings to identify areas that require more research (Snyder, 2019). This is a critical research component before creating theoretical frameworks (Webster & Watson, 2002).

Also, for understanding validation practices in hazard and risk analysis work in safety-critical industries, the research design adopts interviews as the methodological basis, because these can provide a more content-rich contextual understanding of work practices (RQ 2). The interview research methodology is intended to generate knowledge grounded in human experience (Sandelowski, 2004). Thus, it generates in-depth knowledge of the actual validation practices among system safety practitioners, the reasons behind the choices made, and challenges and avenues for future work. This choice is also based on the impracticality of obtaining safety documents from industry practitioners, which can be challenging due to various factors, such as confidentiality and undocumented processes.

To determine the answer to RQ 1, a representative sample of papers spanning a decade (2010-2019) is selected from the *Safety Science* journal. This journal is selected because it

is one of the leading journals in safety research, with a comparatively long publication history (Li & Hale, 2016). With a top-ranking position and a strong reputation among scholars, it is widely acknowledged for its significant academic impact (Reniers & Anthone, 2012). Furthermore, as a multidisciplinary journal, model-based safety analyses represent an important cluster in its publication records (Li & Hale, 2016).

The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) method is used to identify, screen, determine eligibility, and include studies for analysis (Moher et al., 2009). The details of the performed steps to find the sample articles are summarized in Table 1 and can be found in Section 2 of PI. Once the sample articles are selected, the close reading method (Brummett, 2019) is used to extract data, including the title of the paper, name of the author/authors, Digital Object Identifier (DOI), safety concept (as intended in the scope of this study, the concepts closely related to safety, which are risk, reliability, and resilience are also included in the search, see Table 1), year of publication, country of origin, stage of the system life cycle, industrial application domain, model type/approach, validation approach, and terminology used for validation.

To facilitate the analysis, the extracted data is classified into relevant categories. To achieve this, the categories are initially formed based on the existing relevant categorizations available in the literature, and then further refined through the identification of emerging themes in the studied sample.

The potential correlation between validation work and other above-mentioned variables is examined. With this, it is aimed to investigate whether validation has been more focused on in relation to these variables. For instance, it is tested whether articles proposing a model for a specific industry or originating from specific countries are more likely to address validation. This is rooted based on the understanding that safety analyses are often executed as part of regulatory requirements, the specifics of which may differ significantly between countries and industries (Rae & Provan, 2019). Hence, these questions aim to provide some insight into whether such contextual factors lead to significant differences in the degree of validation of model-based safety analyses originating from different countries or industry sectors.

The year of publication is an ordinal categorical variable, while others are nominal categorical variables. A new nominal categorical variable is added to the dataset, labeled ‘validation’, which shows whether a model proposed in an article is validated or not, so it has two states. To investigate the statistical correlation between validation and the nominal variables, Fisher’s exact test is used. This is an alternative to Pearson’s chi-square test of independence when the sample size is small (Agresti, 2007; McCrum-Gardner, 2008). The relationship between validation and the year of publication is tested using a Kruskal-Wallis test (Hecke, 2012). This test is the non-parametric equivalent of one-way ANOVA, and it is best for cases when we have one ordinal and one nominal variable.

Table 1. The details of the research method for RQ 1

Procedure performed	Results
The query which is run on WoS	TS = ((“Safety” OR “Risk” OR “Reliability” OR “Resilience”) AND (“Model” OR “Method” OR “Approach” OR “Framework”)) Combined with IS = (0925–7535)
Number of documents after initial search results	1477 research studies
Number of documents after screening and eligibility checking	247 research studies
Criteria for determining the sample size	A confidence level of 95% and a confidence interval of 5%
Studies included in the analysis after sampling	151 research studies selected from 247 research studies using a proportional stratified sampling strategy

To investigate the answer to RQ 2, a semi-structured interview method is used through which qualitative data is gathered (Magaldi & Berler, 2020). To select interviewees, a combination of two non-probability sampling techniques is used, namely purposive and snowball sampling. In purposive sampling, participants are selected based on specific qualities they possess (Etikan et al., 2015). Snowball sampling, on the other hand, involves asking participants to recommend others who may also meet the selection criteria, creating a referral-based approach (Bhattacharjee, 2012). In this study, sampling begins with purposive sampling and then continues with snowballing to identify additional participants who met the selection criteria.

Various methods are used to recruit participants for this study. An initial list of prospective interviewees is prepared, searching the term "system safety" on the people tab of LinkedIn to whom requests for participation are sent. Four relevant industry groups on LinkedIn are also identified in which a research poster is posted inviting members to participate in an interview. In addition, benefiting from the snowballing approach, the initial participants are asked to recommend others who meet the selection criteria and have related work experience. Personal recommendations by the researcher's network are also utilized for snowballing.

In order to ensure the adequacy of the number of interviewees in qualitative research, data collection should continue until a saturation point is reached, where no new information is being added to the data (Corbin & Strauss, 2008). Thus, it is important to clarify the steps taken to reach saturation (Bowen, 2008). In this research, after conducting and transcribing the first five interviews, preliminary categories are identified using NVivo software. As subsequent interviews are conducted, new themes emerged, and categories are amended accordingly. After conducting a total of 15 interviews, no new data or themes are identified, and the same answers repeatedly surfaced. Therefore, no new categories are added. However, to confirm the saturation of ideas, an additional 5 interviews are conducted.

Table 2 provides a summary of the research methodology employed to address RQ 2 and the details can be found in Section 2 of PII.

Table 2. The details of the research method for RQ 2

Procedure performed	Results
Research Ethics Board approval	From Dalhousie Research Ethics Board (REB) under approval number 2021-5761.
Participant recruitment approach	A combination of purposive and snowball sampling
Interview questionnaire	Structured into three parts and included 24 questions (Refer to Appendix B, Paper II)
Number of interviewees	20
Number of interviews with each participant	1
Duration of interviews with each participant	60 to 90 minutes
Data analysis software	NVivo

To answer RQ 3, the gained insights through answering RQs 1 and 2 as well as the theoretical validation concepts and ideas in closely related fields of study are relied upon to develop a validation framework for STPA. This approach is consistent with reality-based safety science, which involves studying current practice to advance theory and using theory to improve practice (Rae et al., 2020). The selected scientific fields for developing the validation framework are risk science, social science, systems dynamics, simulation modeling, and operations research.

The investigation of theoretical ideas and validation tests from the above-mentioned fields involves the following steps. First, articles addressing aspects of risk analysis validation are selected based on the papers included in the *Safety Science* Special Issue "Risk analysis validation and trust in risk management" (Goerlandt et al., 2017a), and through a process of backward snowballing, which involves checking the references in those papers for relevant articles. Second, articles in social science validation are chosen as found useful in a prior work on a validation framework for expert-based models by Pitchforth & Mengersen (2013). Finally, articles related to operations research, system dynamics, and simulation modeling disciplines are gathered through a keyword search ("validation", "validity") in key journals that address these model types, followed by further backward snowballing of the identified articles.

A list of all the identified papers is compiled. The papers are then evaluated to determine if they presented a comprehensive set of validation tests. If they do not, they are excluded from the list. However, if they meet the criteria, their citation scores are checked using Scopus and recorded as of March 2022. From this refined list, two highly cited articles in each discipline are carefully selected for further analysis. These articles are particularly noteworthy because they often build on and integrate previous work in their respective fields and represent the key concepts underlying validation. Therefore, they serve as a solid foundation for the development of the STPA validation framework. The process of selecting papers is detailed in Figure 4, adopted from paper III.

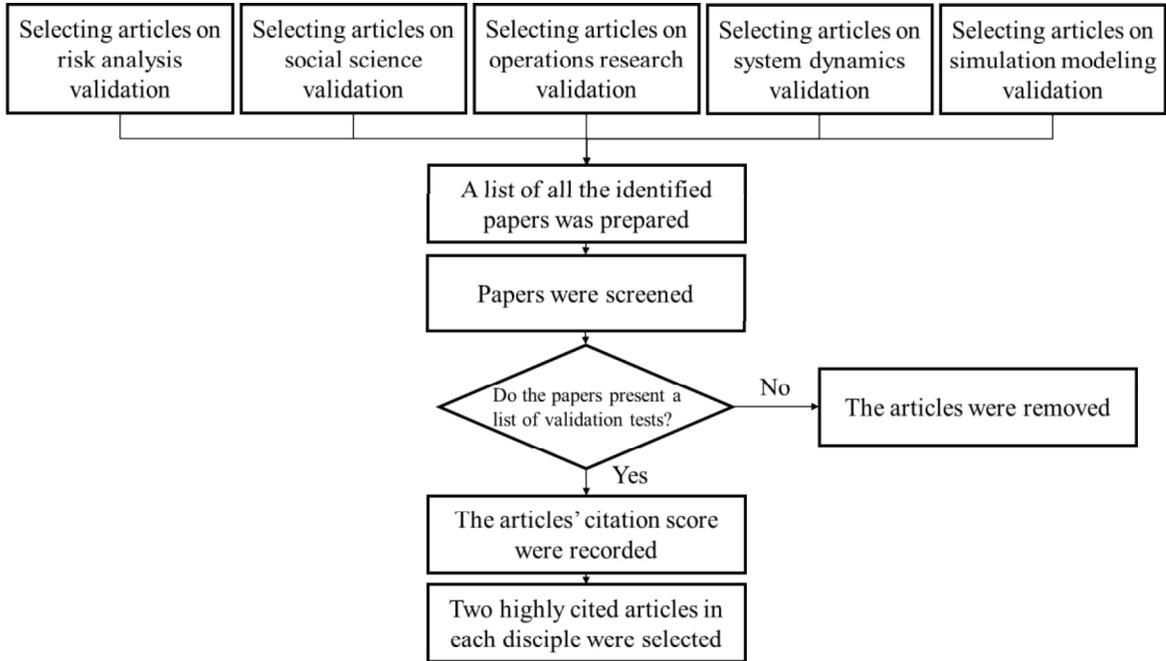


Figure 4. The process of selecting papers adopted from PIII

The PRISMA statement (Moher et al., 2009) is used to identify, screen, and compile a list of validation tests proposed in the selected papers (Figure 5) for developing an STPA validation framework. This results in a list of 150 tests. The list is screened to eliminate duplicates and unrelated tests, and similar tests are categorized into a set of categories. This process results in a final list of tests that can be used as a basis for building an STPA validation framework. This process is illustrated in Figure 5, which is adopted from PIII.

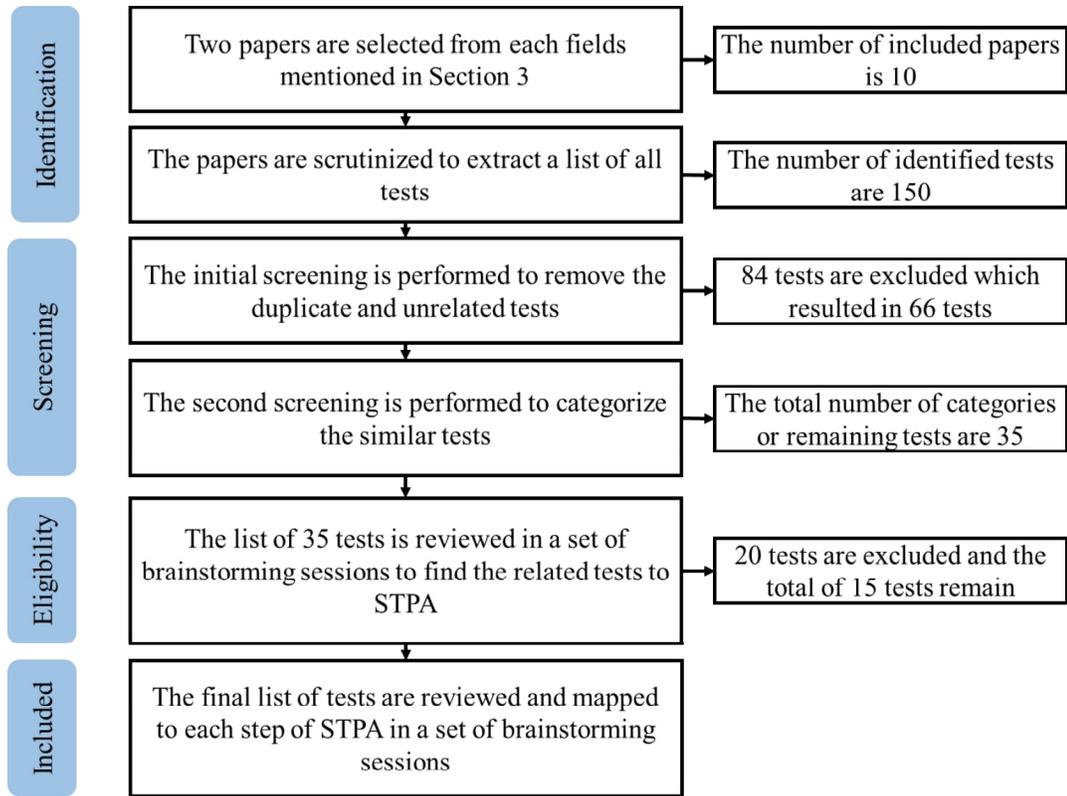


Figure 5. The process of constructing the STPA validation framework based on PRISMA flow diagram, adopted from PIII

Finally, to address RQs 4 and 5 (Section 1.3), semi-structured interviews (Magaldi & Berler, 2020) are conducted with STPA experts to gain a deep understanding of their current STPA validation approaches as well as their judgments on this framework. This approach aligns with the recommendations put forth in Rae et al.'s (2020) manifesto, which suggests that empirical research can be used to validate, criticize, and further develop theoretical ideas. Thus, this thesis not only proposes a framework rooted in theoretical validation ideas (Paper III) but also tests the framework through empirical research by seeking the STPA experts' judgments, see Figure 6. To accomplish this, the current STPA validation practices employed by experts are investigated to identify similarities and differences between the framework and current practices (PIV). Then, feedback from experts is solicited regarding the concepts and tests within the framework to evaluate its reasonableness (PIV).

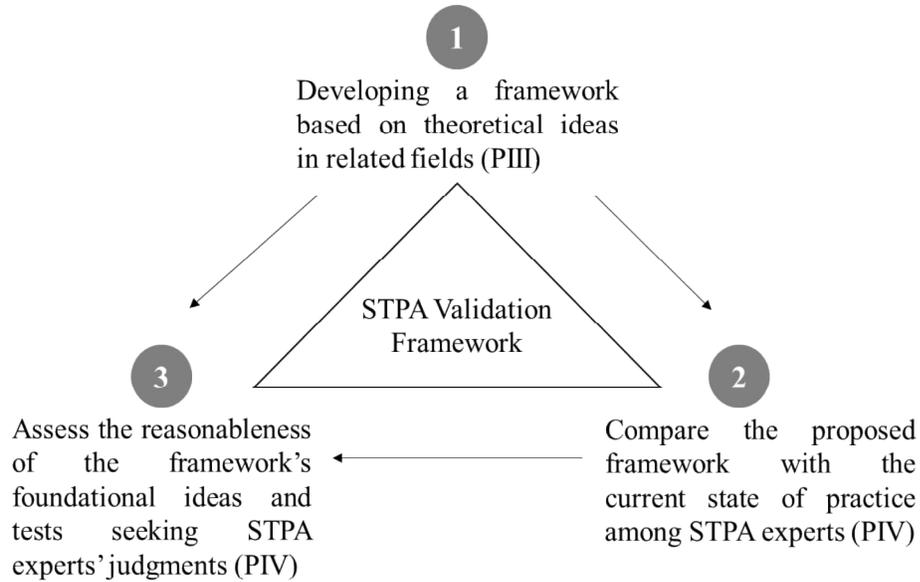


Figure 6. Further empirical research on the proposed STPA validation framework

To perform this interview study, the research ethics approval is received from Dalhousie Research Ethics Board (REB) under approval number 2021-5761. The participants in this study are recruited from various sources, including attendees of the 10th European STAMP workshop and conference, experts in STPA identified through the 2022 MIT STAMP workshop, and academic authors of key papers identified through a recent scoping review on STAMP by Patriarca et al. (2022). The participants are researchers or industry practitioners with experience using STPA-related tools, making them a suitable group for obtaining insights into STPA validation. The demographic information of the interviewees is summarized in Table 3.

Table 3. Demographics of the interviewees

Demographic information	Values and distribution (N, %)	
Field	Academia (7; 54%)	Industry (6; 46%)
Years of experience	[5,10] (8; 62%)	[10,15] (5; 38%)
Highest education level	Master (7; 54%)	PhD (6; 46%)

The research involves two in-depth interviews with 13 experts to gain insights into STPA validation by domain experts. The first interview consists of questions regarding the interviewee's background, as well as any form of validation approaches and practices they

adopt for an STPA analysis (Item 2 in Figure 6). The interviews are transcribed and analyzed using thematic analysis (Braun & Clarke, 2006). The identified themes are then mapped with the tests and proposed concepts in the theoretical validation framework. This reveals the similarities and dissimilarities between the validation framework developed based on theoretical foundations, and what STPA practitioners commonly do to develop (what they consider to be) valid analyses (RQ 4).

In the second interview session, the developed STPA validation framework is first presented to the interviewee. Then, the results of the mapping of the results from the first interview with the theoretical framework are presented to the interviewee (Item 2 provides input to Item 3 in Figure 6). Further, the interviewee is walked through each test which they did not mention in their first interview, to investigate whether such a test would nevertheless be reasonable to be used in practice from their point of view. Thus, the reasonableness of the whole framework and each test are assessed in the second interview (Item 3 in Figure 6 and RQ 5). The results of the second interview are also analyzed to identify themes and gain further insights into the obtained qualitative data.

To perform data analysis in this research, NVivo software (QSR International Pty Ltd. NVivo, 2020) is used. In terms of data saturation (O'Reilly & Parker, 2013), after ten interviews, no new themes are identified. However, to ensure that saturation occurs, three more interviews are conducted. A final thematic analysis is also performed once all interviews are performed to map the identified categories from all data to the proposed theory-based validation framework and to obtain overall insights for the research study.

CHAPTER 3: Results

This section is divided into two parts, each dedicated to one research objective, i.e. Research Objectives 1 and 2, as explained in Section 1.3. Section 3.1 explains the results of Research Objective 1. Section 3.2 elaborates on the results of Research Objective 2.

3.1 State of the Practice in Validation of Risk and Safety Approaches

3.1.1 An Empirical Study on the Validation of Academic Model-based Safety Analysis in Socio-Technical Systems

As highlighted in Section 1.3, to investigate Research Objective 1, two RQs are defined. This section presents the results of RQ 1, which involves investigating the state of practice in model-based safety analysis validation for socio-technical systems through an empirical study of relevant articles published in the *Safety Science* journal. The hazard/accident analysis method is considered one of the model types/approaches in this study. The results of this study have been published in the form of PI.

The data analysis shows that in 63% of the articles, no model validation is presented, while in only 37% of the articles, a validation of the proposed or applied models is performed. The validation approaches applied in articles where validation is performed fall into seven categories. The categories are initially formed based on those proposed by Goerlandt et al. (2017b), which are *reality check*, *peer review*, *quality assurance*, and *benchmark exercise*. Three more categories are further added to incorporate additional approaches identified in sample papers that do not fit within the initial categories, which are *validity tests*, *statistical validation*, and *illustration*. The definition of each category is provided in Table 3.

Table 4. Definition of the validation approaches identified in PI

Validation Approach	Definition
Reality check	Comparing the output of the model with real-world data, such as experimental or field data
Peer review	Examining the model by Subject Matter Experts' judgments according to a set of criteria
Quality assurance	Examining the process behind the analysis
Benchmark exercise	Comparing the result of a model with a parallel analysis

Validation Approach	Definition
Validity tests	Applying a set of tests to a model for analyzing its validity using mainly quantitative techniques, such as sensitivity analysis
Statistical validation	Using the statistics-based quantitative methods, such as tests of means, analysis of variance or covariance, the goodness of fit tests, regression and correlation analysis, spectral analysis, and confidence intervals
Illustration	Applying a model to a specific illustrative case study to show how the model works as a means to support the reasonableness of the proposed model

Figure 7 presents the distribution of validation approaches used in the papers that report performing validation. *Illustration* is the most frequently used method, appearing in 23.9% of the papers in which validation is performed, while quality assurance is the least common, appearing in only 2.8% of them. It is important to note that no judgment is made about the quality of the validation methods used in these papers. Rather, any paper that claimed to have validated its model is listed as having performed validation. Therefore, the quality and rigor of the validation methods used in these papers cannot be assumed. For instance, while illustrations are a popular technique, they may not be sufficient for fully validating a model. Merely demonstrating that the model can be applied as intended through an illustrative case study does not necessarily indicate that the model is "good" or that its assumptions are valid.

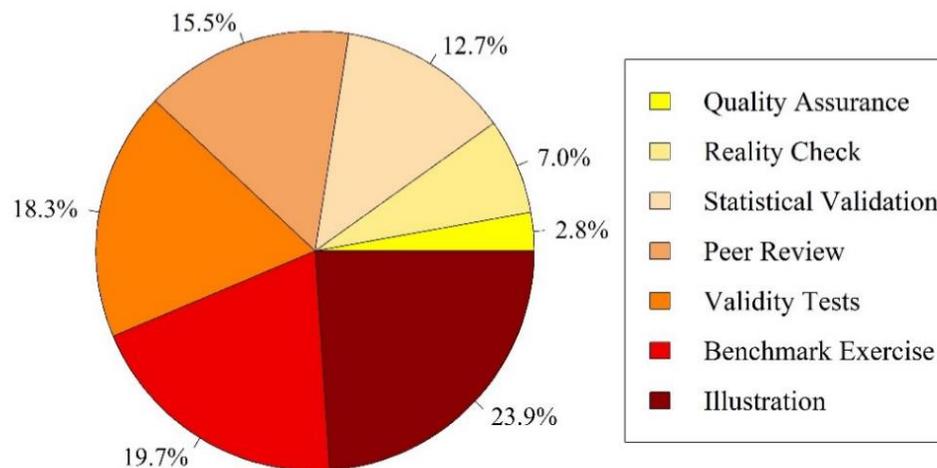


Figure 7. The distribution of articles in terms of the adopted validation approach, for cases where validation is performed, adopted from PI

It is worth noting that some of the papers, in which validation has been performed, utilize a combination of the above-mentioned validation approaches, such as a mix of *benchmark exercises*, *validity tests*, and *peer review* categories. While these mixed approaches may be beneficial, the results suggest that even in articles where some form of validation is performed, the processes are typically ad-hoc.

This study also examines the potential correlation between validation work and several other variables extracted from the selected papers, including the year of publication, safety concept, model type/approach, country of origin, industrial application domain, and stage of the system life cycle. The results indicate that there is no significant correlation between the frequency of validation and these factors, as demonstrated by the non-rejection of the null hypotheses based on p-values. In particular, the study found no evidence to suggest whether performing validation is associated with safety-related concepts, model type/approach, country of origin, industry, or stage of a system's life cycle. These findings imply that the consideration of validation is not necessarily influenced by these variables, suggesting that lack of focus on validation is prevalent across subdomains of safety science, across different communities working on different theoretical or methodological foundations, and in various industrial application domains.

Having analyzed all the articles in the selected sample, it is discovered that the language used to describe validation in model-based safety analysis is inconsistent. The terms *validation*, *evaluation*, *effectiveness*, *verification*, *comparison*, and *usefulness* are used interchangeably in the selected papers, with validation being the most frequently used term. Figure 8, adopted from PI, illustrates the distribution of papers according to the terminology used for validation.

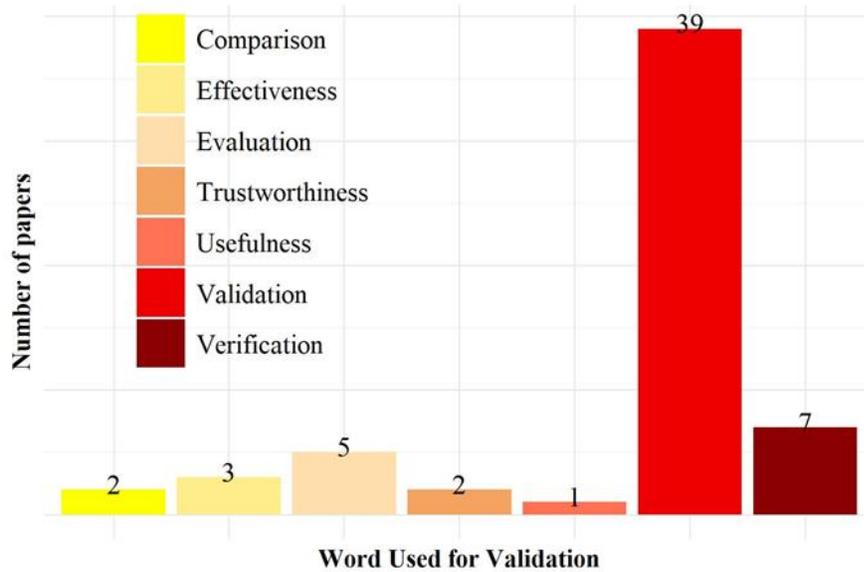


Figure 8. Distribution of papers in terms of the terminology used for validation, adopted from PI

3.1.2 An Empirical Study on the Validation Practices of Hazard Analysis Techniques in Safety-Critical Industries

According to the findings of RQ 1 (Section 3.1.1), validation approaches are employed in some academic settings. However, the analysis is limited to academic papers, and it is not possible to assess the quality of the techniques used or the challenges faced by practitioners in real-world safety practices based solely on this data. Therefore, the goal is to expand beyond academic literature and acquire a qualitative understanding of validation practices of hazard analysis techniques in safety-critical industries. Thus, this section aims to explore RQ 2 (Section 1.3) through which the state of practice in hazard analysis validation for system safety among system safety practitioners is investigated.

More specifically, the main objective of this study is to examine the various approaches and techniques used by professionals to validate hazard analyses, as well as the underlying motivations, driving factors, major obstacles, and limitations of such validations. Additionally, the study aims to identify areas for improvement and opportunities to enhance the validation of hazard analyses. The scope of this study is narrowed down to hazard analysis techniques, as opposed to the broader scope of model-based safety analysis techniques as focused on PI. The rationale behind this approach is that a more focused examination of hazard analysis techniques will provide a more comprehensive

understanding of the subject matter, whereas a wider scope would only offer a superficial overview of the research objectives. Also, due to the terminological inconsistencies in the field, it is considered that participants would have been confused about the focus of the research.

The findings of this research have been published in PII. The study findings indicate that all interviewees recognized the importance of validation as a key factor in achieving a comprehensive hazard analysis. Practitioners reported that validation provides a means to ensure that every aspect is adequately covered, identify gaps in the analysis, and detect errors or discrepancies in assumptions. Out of the twenty participants, only one individual reported not performing any form of validation. The overwhelming majority of participants confirmed that they consistently validate their hazard analyses. These results emphasize that participants recognize hazard analysis validation as a significant value driver while other involved parties, such as stakeholders, may not see the value readily, and convincing them to perform validation is a challenge.

The study reveals that reviews by independent experts (or peer review) are the most commonly used approach among almost all practitioners to validate their hazard analyses. Rather than relying solely on one person's expertise, combining the experiences of several experts leads to a more robust analysis. In addition, several practitioners reported using benchmark exercises as another form of validation. This exercise involves comparing the results of an analysis with those of a parallel analysis to cross-check and ensure that nothing has been overlooked. This approach instills confidence that all aspects have been thoroughly considered.

There are specific motivations that drive practitioners to validate their hazard analysis. According to several practitioners, the primary motivation often stems from internal organizational policies. They note that the decision on whether to perform validation or not can depend on the specific company they work for. Another reported driving factor is the level of novelty of the system. If a system is well-known and has been used extensively before, the level of validation needed is judged lower compared to when there is a new system or feature in an existing system, which requires a higher level of validation. Also,

sometimes validation is performed to ensure that the analysis is comprehensive and that no essential components have been overlooked.

The practitioners identified five primary challenges to successfully performing validation. Practitioners highlighted that they must convince stakeholders about the importance of validation as they may not understand why validation is necessary. Even when they grant access, they may not realize the level of experience required for proper validation. Lack of competency is also noted as a barrier to validation, with a worldwide shortage of competent personnel and insufficient time taken to train novice engineers. Schedule pressures and budget limitations are also mentioned as hindrances, with a trade-off required to decide how much time and money to spend on hazard analysis validation.

Even when validation is included as an explicit task in project schedules and budgets, and the necessary competency is provided to the team, practitioners still face significant challenges due to the lack of clear guidelines on how validation should be performed. Questions such as how to perform validation, what validation tests to use, when to stop the validation process, and who should be involved in the validation process are often unclear or not well-defined.

Two limitations associated with current validation methods are highlighted by practitioners. According to them, the most significant of these limitations is subjectivity. Practitioners have emphasized that the individuals responsible for conducting the validation are a critical component of the analysis process and that the comprehensiveness and accuracy of the analysis are heavily dependent on their skill, knowledge, and experience. If individuals without the necessary expertise are included in the analysis process, there is a risk that they may not ask the right questions or make accurate assumptions, which could result in important details being overlooked. Hence, the analysts consider that identifying the appropriate individuals to participate in the analysis process is a crucial factor in reducing subjectivity and achieving comprehensive and accurate results.

Another highlighted weakness is that validation of hazard analysis is an expensive and time-consuming process. This situation is exacerbated if stakeholders do not understand the value of validation and consider it to be a waste of time, effort, and money (as explained above as one of the challenges for performing validation). It is highlighted that there is

always that push from companies to save money and spend less time on a project. If the validation process fails to uncover anything of significance, some may consider the effort and resources invested to be a mere confirmation that the primary analysis is sound, without adding much value.

To tackle the highlighted limitation, several opportunities for improvement are proposed by the practitioners. A common opinion among the participants is that a formal way of how to validate hazard analysis is required. They suggest having standard processes proposed by regulatory authorities, and that validation frameworks proposed by academia could add significant value to improve the current situation. It is also essential to enhance the visibility of validation and to educate the relevant actors in an organization about the importance of validation. Finally, increasing awareness could also lead to top management support, which is mentioned by a few practitioners as an important opportunity for improvement. A leader's approach can instill commitment or indifference.

3.2 Towards a Formal Validation Approach for STPA

The results of RQs 1 and 2, which are outlined in Sections 3.1.1 and 3.1.2, respectively, suggest that although some form of validation exists both in academia and industry, the lack of clear guidance and a formal validation approach make the process of validation challenging. Specifically, the practitioners interviewed in PII acknowledge the significance of validation and make efforts to carry it out. However, the absence of clear guidance regarding the validation process, whether from regulatory authorities or researchers, is a significant obstacle that needs to be addressed in order for them to improve their hazard analysis validation practices. Consequently, additional research is required to determine a formalized process for the validation of a hazard analysis.

One approach to overcome this challenge is to rely on theoretical frameworks and best practices to inform and guide researchers and practitioners in their validation efforts. This is consistent with the idea of reality-based safety science (Rae et al., 2020), which is based on a “virtuous cycle of studying current practice to advance theory and applying theory to advance current practice.” That is, by leveraging theory, a formal validation framework is proposed. However, in order for the proposed framework to be used either in academic or

industry settings, it is essential to test the proposed framework and subject it to empirical testing and provide evidence of the framework's reasonableness (Rae et al., 2020).

RQ 3 is established to develop a theoretical validation framework that can be utilized by both researchers and practitioners. Given the broad range of hazard analysis techniques (Ericson, 2015), the scope is narrowed down by focusing on the STPA technique. As mentioned in Chapter 2, the framework is developed based on fundamental principles in risk and safety science and previous theoretical research on validation in related fields. The results of RQ 3 are discussed in Section 3.2.1 and the details can be found in PIII.

Further RQs 4 and 5 are defined to empirically test the proposed theoretical validation framework. RQ 4 aims to scrutinize the extent to which the proposed ideas and tests in the validation framework align with practices already adopted in current practice by STPA experts. RQ 5 evaluates the reasonableness of the proposed STPA validation framework seeking STPA experts' judgments. The results of these two RQs are outlined in Section 3.2.2, with further details found in PIV. This research can also contribute to diminishing the gap between academic safety science research and the actual work of safety practitioners, which is needed according to Reiman & Viitanen (2019).

3.2.1 Theory-Based Framework

To better reflect on the proposed validation framework (RQ 3), this section is divided into two sections. Section 3.2.1.1 explains conceptual foundations for and assumptions behind the proposed STPA validation framework. Moreover, Section 3.2.1.2 centers on the validation tests put forth within the framework.

3.2.1.1 Conceptual foundations for the proposed STPA validation framework

There are four main conceptual foundations for the proposed STPA validation framework: (i) it assesses an analysis' validity subjectively; (ii) the framework is theory-based; (iii) it considers validation work formatively; and (iv) it relies on independence between analysis implementation and validation teams. These ideas are elaborated on below.

In this thesis, validation is understood as a tool to formalize judgments about the comprehensiveness, accuracy, and credibility of a risk and hazard analysis. That is,

validation is decided subjectively (i). The proposed STPA validation framework aims to enhance the intersubjective agreement among analysts, users, and stakeholders by evaluating validity as a judgment made by an assessor. This relies on the idea of risk and hazard analysis being believed to be inherently subjective, which is supported by different views on the ontological status of risk, associated with realism and constructivism (Aven & Renn, 2009; Rosa, 1998; Solberg & Njå, 2012).

The concept of risk in realism is based upon the idea that a certain state of the world can objectively be defined as risk, whereas since these states are not predetermined, they are uncertain (Rosa, 1998). Rosa argues that the definition of risk moves from an objective state to a subjective state based on an assessor's ability to identify, measure, and understand that state. Therefore, social factors play a role in shaping risks. On the other hand, in the constructivist view, the core concept of risk is associated with the assessor's knowledge of the situation and the ability to imagine a possible future state of affairs, which are subjective (Solberg & Njå, 2012). Solberg and Njå (2012) suggest that risk is not considered in isolation, but rather it is connected to specific activities.

Therefore, a risk analysis is a report on the uncertainties expressed by analysts, which is inherently subjective but rooted in evidence, which can be strong and compelling or weak (Aven & Guikema, 2011). This is what Kaplan (1992) called “evidence-based” risk assessment and decision-making. Based upon this idea, in addition to the analyst’s experience and knowledge, a “consensus body of evidence” is required to make a decision. The proposed framework aligns with this notion. That is, validity is assessed subjectively but rooted in evidence.

One such a subjective aspect in this framework is the decision on the validation cessation. As suggested in the framework, validation cessation happens through a brainstorming session between the analysis implementation and validation teams, where they assess the level of agreement. If the agreement is high, the validation can be stopped. However, if the agreement is low, the analysis is revised iteratively until an acceptable level is reached, indicating that validation can be stopped. Ceasing validation is essentially a judgment, which cannot be simply reduced to some quantitative criteria. Practical limitations, such as budget and schedule constraints, mean that regardless of the quantitative criteria

established, validation is likely to be stopped. Therefore, incorporating flexibility in the process to allow for expert judgments would be advantageous.

As mentioned in Chapter 2, to develop this framework the theoretical validation concepts and ideas in closely related fields of study are relied upon, referring to the second conceptual foundation (ii). The rationale behind this is that the underlying validation ideas and concepts in different sciences and application domains are similar (Eker et al., 2019). This is because different sciences will commit to similar philosophies of science, i.e. realist vs constructivist ideas, as explained above. This similar foundation provides an opportunity for knowledge exchange between academic communities. Some researchers have engaged in such bridge-building work by drawing on approaches from different fields to develop validation frameworks. For instance, Pitchforth & Mengersen (2013) incorporates approaches from the fields of psychometrics and system dynamics to develop a high-level validation framework for Bayesian Networks. Schwanitz (2013) draws on operations research and simulation modeling disciplines to develop a validation framework for integrated assessment modeling of global climate change.

The selected fields of study for developing the STPA validation framework are risk science, social science, and three other narrower areas of scholarship, which are operations research, system dynamics, and simulation modeling disciplines. The literature on validation in risk science is relevant as the whole hazard analysis process can be framed with a risk management context. As defined by Society for Risk Analysis (SRA) (Aven et al., 2018), a hazard is “a risk source where the potential consequences relate to harm.” Hazards should be identified to specify their inherent and unique risks (Ericson, 2015). Hence it is imperative to review how validation is approached in risk science.

Literature on validation in social science can be also a useful base. This relates to the realist/constructivist debate in risk science, where state-of-the-art views on the risk concept and hazard analyses consider these to be socially shared constructs (as explained above). If hazard and risk analysis are best understood as an expression shared by a group of experts, it can be approached as a social phenomenon. Hence, social science concepts regarding validation become meaningful to consider.

As per Section 1.2.2.1, STPA explains accidents as a result of inadequate control over a system's behavior, in contrast to traditional hazard analysis techniques that focus on chain-of-event sequences. To develop an STPA validation framework, it is possible to draw on insights from the operations research, system dynamics, and simulation modeling disciplines because of STPA's reliance on modeling a system as a safety control structure. STPA involves creating a model of system components and their functional relationships and interactions through feedback control loops (Leveson & Thomas, 2018). This safety control structure is not a quantitative simulation or a mathematical model, but a conceptual model used to structure analysts' knowledge and understanding of the system. This model is then used to systematically inspect Unsafe Control Actions (UCAs) (Section 1.2.2.1).

In addition, because of relying on a control structure as a basis for hazard analysis, STPA can be categorized as a model-based safety analysis. Model-based safety analysis can be built upon qualitative methods (Boolean formalisms such as fault trees or event trees) or quantitative methods (Transition systems such as Markov chains and Petri nets) (Abdellatif & Holzapfel, 2020).

Additionally, STPA shares conceptual similarities with system dynamics models, which are particular types of simulation models, as these aim to model complex dynamic systems through various feedback loop structures (Keys, 1988). In the Ph.D. dissertation by Dulac (2007), it is explained that system dynamics and STAMP have similarities, which are exploited to propose a dynamic risk management approach by combining STAMP safety control structures with system dynamic modeling principles. Hence, the validation literature in these modeling disciplines is also considered a useful foundation for developing an STPA validation framework.

The developed validation framework is formative in nature (conceptual foundation iii) and serves to assist peers and stakeholders in systematically reasoning about the analysis and providing advice for improvement or further elaboration. According to Barlas (1996), model validation is a gradual "confidence-building" process, rather than a binary "accept/reject" division. Forrester (1980) asserts that there is no single test that can validate a model. Instead, confidence in a model accumulates gradually as the model passes more tests and new points of correspondence between the model and empirical reality are

identified. Therefore, the formative process serves as a thinking tool, rather than attempting to rate an analysis according to a specific quantitative standard, as Busby and Hughes (2006) suggest.

The proposed framework suggests that, ideally, two distinct teams should carry out the STPA implementation and validation processes (conceptual foundation iv). This is because those who conduct the analysis may be less likely to question their own judgments (Pitchforth & Mengersen, 2013), which can be linked to the psychological phenomenon referred to as the 'IKEA effect,' whereby people overestimate the value of their own creation (Norton et al., 2012). Therefore, to mitigate this potential bias, it is recommended to involve two separate groups: one responsible for analysis implementation and another responsible for analysis validation.

3.2.1.2 Specific Validation Tests

The proposed STPA validation framework consists of 15 validation tests, each targeting a specific element of an STPA analysis. The list of proposed tests with their brief definitions are provided in Table 4 and the details can be found in PIII. The STPA implementation process is scrutinized to identify key elements of an STPA analysis (Figure 9). Fourteen critical elements are identified, and each is matched with appropriate validation tests from Table 4.

Table 5. The list of proposed tests with their brief definitions, adopted from PIII.

Proposed validation tests	Definition of the validation tests
Nomological Validation	Establishing confidence that an analysis fits within a wider domain based on the literature
System Boundaries Validation	Evaluating how a change in the identified boundary would affect the identified losses, system-level hazards and constraints
Data Validation	Ensuring that the sources of data (e.g. design documents) are clearly specified and validated
SMEs Validation	Investigating the criteria and processes for SMEs selection, elicitation, and aggregation
Assumption Validation	Identifying, describing, and documenting the assumptions, and ensuring that they are understood and agreed upon

Proposed validation tests	Definition of the validation tests
Content and structure Validation	Investigating the elements as well as their functional relationships in the control structure
Concurrent Validation	Comparing the control structure of an already existing STPA analysis for an identical system
Convergent Validation	Investigating how similar the control structures are in the identified analyses of similar systems
Face Validation	Reviewing the analysis (e.g. loss scenarios) to judge whether they appear reasonable and accurate
Extreme Condition Validation	Evaluating the plausibility of the controller's constraints under extreme conditions
Behavior Validation	Comparing the system's behavior with and without enforcement of the identified constraints
Historical Validation	Reviewing the historical data (e.g. accident/incident data of the studied system or identical or similar systems) to ensure that the associated contributing factors are covered in the identified scenarios
Traceability Validation	Checking the traceability of all items, and making sure they are logical and documented
Documentation checking	Reviewing the finalized documents and ensuring that they are correct, clear, complete, and in formats understandable to stakeholders
Review of Presentation of results	Reviewing the presentation to ensure the sources of uncertainty and the limitation of the analysis are included

As a heuristic, it is intended to assign at least three tests to each of the four steps of STPA (Section 1.2.2.1), to provide reviewers of an STPA analysis with sufficient guidance to systematically evaluate the accuracy and comprehensiveness of each aspect and analysis step. As the validation of the final STPA results is more concerned with the credibility of an STPA analysis, two tests are assigned. Combining the different elements of STPA and the identified tests in a logical structure finally results in the proposed STPA validation framework. The resulting list of STPA elements with their corresponding tests is presented in Figure 9, adopted from Paper III.

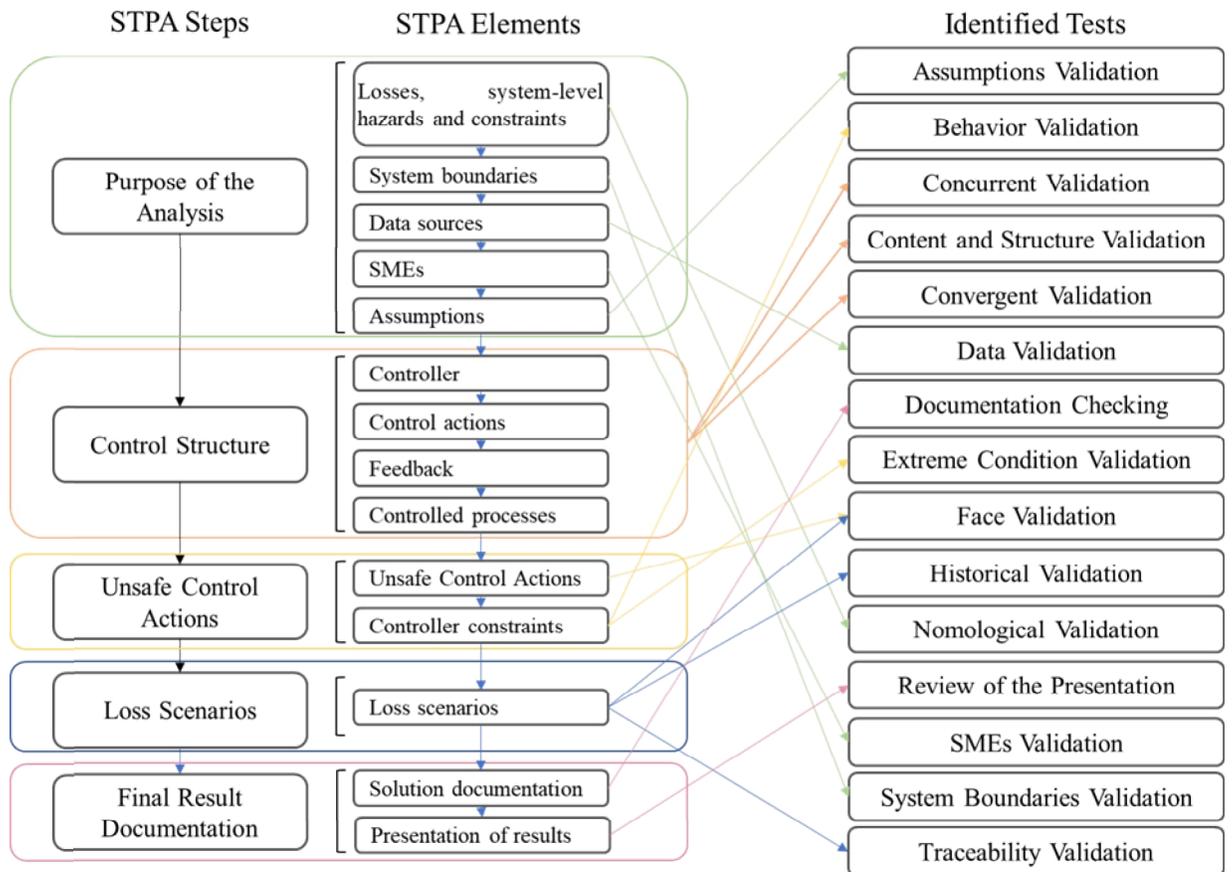


Figure 9. The assigned tests to each element of STPA, adopted from PIII

The proposed tests are further elaborated as high-level guide questions. The guide questions are designed in three different ways: (1) extracted from literature and amended to be used in the context of STPA, (2) developed based on important notes in the STPA handbook, or (3) developed based on the author's knowledge and experience in the field of hazard analysis, risk research, and validation, using the ideas captured in the validation tests extracted from the literature described in Section 2.

As an illustration, Figure 10 shows two validation tests proposed for the third step of implementing STPA, along with their respective guided questions. The two tests are *face validation* and *behavior validation*, each targeting a different element in the third step of STPA. Face validation seeks to validate the identified Unsafe Control Actions (UCAs), while behavior validation focuses on the identified controller constraints. Section 5 in PIII provides detailed explanations of each test and their associated guided questions.

STPA Step	STPA Elements	Proposed Tests	Guiding Questions
Step 3: Identifying unsafe control actions	Unsafe Control Actions	Face Validation	<ul style="list-style-type: none"> • Are the identified UCAs logical and accurate from the validation team's perspective? • Are there any other possible UCAs that have not been identified by the STPA implementation team? • Are the identified UCAs accurately translated into constraints on the behavior of each controller?
	Controller Constraints	Behavior Validation	<ul style="list-style-type: none"> • How does the enforcement of the identified constraints change the behavior of the system? • Are the changes in the system's behavior as expected by the STPA implementation team?

Figure 10. The example of the proposed tests for Step 3 of STPA

The proposed validation framework is primarily intended to be used in two types of processes: (1) in parallel with the STPA implementation (Figure 11), or (2) after a complete STPA is performed (Figure 12).

As can be seen in Figure 11, in a parallel process, the implementation and validation of STPA occur simultaneously, wherein each implementation step is followed by its corresponding validation step. This approach is inspired by the simulation field's widely adopted use of parallel processes (Landry et al., 1983; Law, 2014; Oral & Kettani, 1993). Conducting STPA validation parallel to its implementation enables the identification of potential errors during STPA execution as soon as a particular step is finished. This eliminates the risk of carrying errors or omissions over to subsequent analysis phases. This approach is consistent with system engineering's validation philosophy, where validation is conducted throughout the system's lifecycle to detect flaws as early as possible (Engel, 2010).

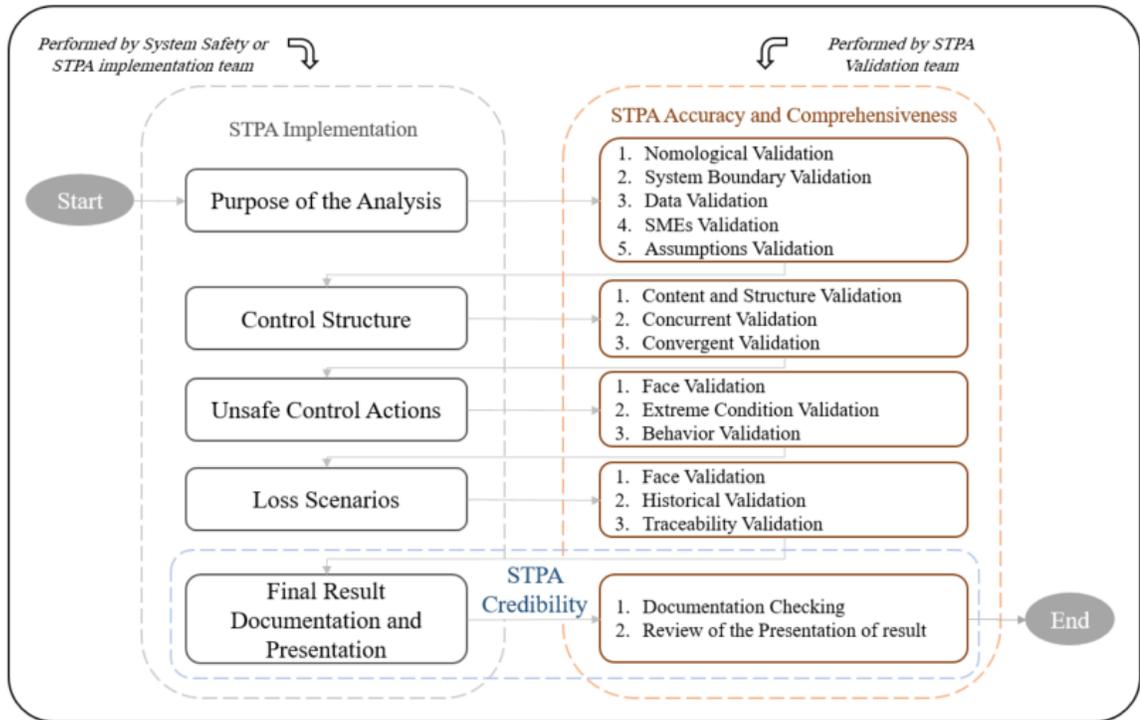


Figure 11. Using the STPA validation framework in parallel with STPA implementation, adopted from PIII

The proposed validation framework can also be used to validate an existing and completed STPA in a post hoc manner. This approach can be useful, for instance, when an external regulatory authority or certification body wants to validate a company's hazard analysis. Another example is when a company outsources the hazard analysis process to an external consultant. In such a case, this framework can be used by the company's internal resources to validate the results provided by the external consultant. Additionally, if a company cannot allocate more time to the validation steps due to scheduling constraints, the validation team can be involved in later analysis stages, and all the tests can be conducted once the STPA implementation is complete. In this case, as depicted in Figure 12, the analysis results are provided to the validation team to perform validation. Nonetheless, using the framework in a post-hoc manner has limitations, including the risk of not considering validation results and finding errors too late for the STPA implementation team to influence the analysis results.

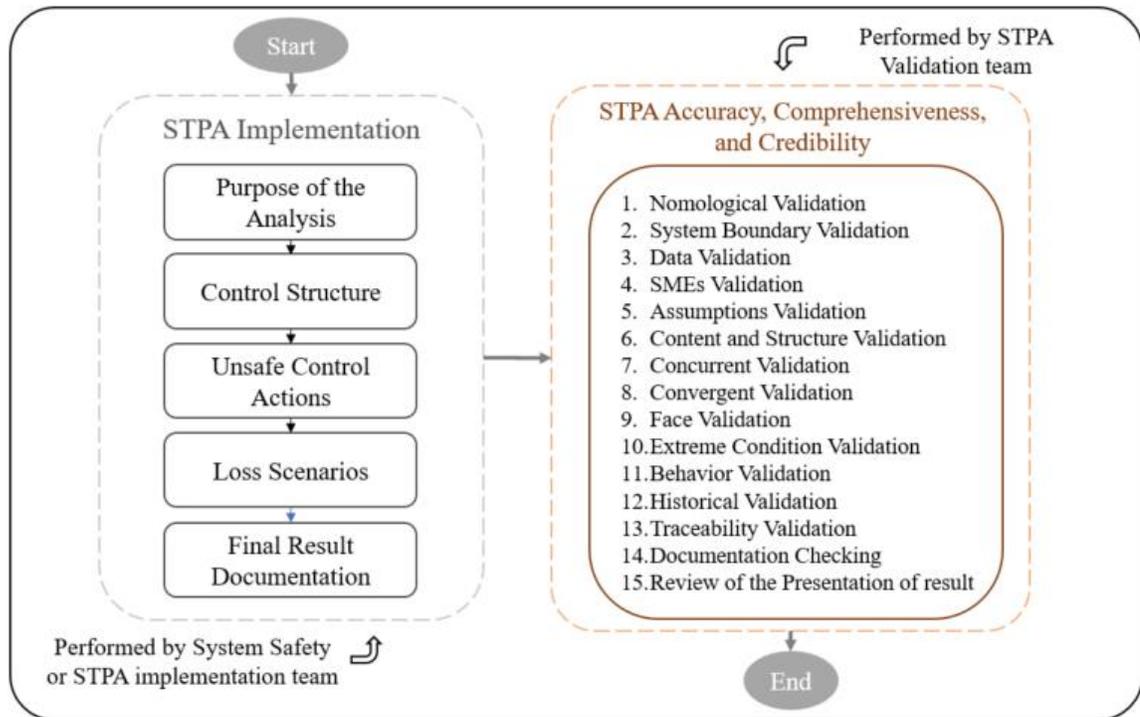


Figure 12. Performing validation using the STPA validation framework after STPA implementation, adopted from PIII

3.2.2 Empirical Confirmation of the Proposed Framework

The findings of RQ 4 (Section 1.3) indicate that all interviewees acknowledge the importance of validation in an STPA analysis context and endeavor to carry it out, although it may not always be feasible. Certain STPA experts, particularly those who work as consultants, emphasize that their decision to conduct validation hinges on the project they are engaged in and their role in it. For instance, if a project pertains to a regulated industry, there may be numerous formalities to adhere to, and compliance with a standard that mandates validation could be necessary. Additionally, depending on their role in a project, such as a facilitator or leader, validation may not be within their purview. Alternatively, if they are leading the analysis, the extent to which validation is performed is contingent on the resources available for the project.

The interviewees are queried about whether they adhere to a structured or formalized STPA validation process. The majority of STPA experts reported that validation is carried out on an ad-hoc basis and that they neither follow a formalized process nor are guided by a systematic list of validation tests to select from. As a result, they pointed out that the actual

effect of utilizing such ad-hoc validation practices on the results of their analysis remains an unresolved issue.

As explained in Chapter 2, the current STPA validation tests employed by STPA experts are investigated in the first interview session with participants, as part of the inquiry into RQ 4. This is performed to identify the validation tests in the STPA validation framework that are not already used in practice. This list is prepared after the first interview with each participant and then presented to them in the second interview session. This is performed to seek their judgments on the reasonableness of the tests in the framework that are not mentioned by experts, as part of the exploration of RQ 5.

The combined results of the two above-mentioned investigations regarding each validation test of the theory-based framework are categorized into four groups: (1) The interviewee already applies this test; (2) The interviewee has not used this test before but considers that it makes sense to use it in practice; (3) The interviewee has not used this test before but considers that it makes sense to use it in practice, with some caveats and limitations; and (4) The interviewee believes that this test does not make sense to be used in practice. Figure 13 summarizes the interviewees' views about each test, using the above categories.

As can be seen in Figure 13, all theory-based proposed validation tests have already been used in practice by at least one STPA expert. The most frequently used validation tests are *Face Validation* and *Content and Structure Validation tests*, which are already commonly applied by all 13 interviewees. As suggested by interviewees, the reason for the common application of *Face Validation* is that it mainly benefits from and relies on SMEs' knowledge and experience, which may not be readily available from other sources. However, this reliance on experts is also noted as a potential drawback, as the results may not be entirely reliable, especially if the selection of experts is not carefully considered.

The frequent use of *Content and Structure Validation* test is driven by the recognition of the safety control structure as a critical foundation of the analysis, as noted by interviewees. They all emphasized the importance of this test because the rest of the analysis rests upon the accuracy and comprehensiveness of the control structure. When experts are available, interviewees rely on them to review the control structure. Otherwise, they utilize available

technical documentation, such as flow diagrams, hardware and software diagrams, and documented procedures, to validate the control structure.

Only one expert highlighted four tests as not making sense to be used in practice which are *Nomological Validation*, *System Boundaries Validation*, *Concurrent Validation*, and *Convergent Validation*, and one interviewee highlighted the *Extreme Condition Validation* test as not making sense to be used in practice. The interviewee's reason for not considering *Nomological*, *Concurrent*, and *Convergent Validation* as reasonable tests is that these three tests rely on other STPA/hazard analyses, raising concerns about the validity of those studies.

In terms of the *System Boundaries Validation* test's unreasonableness, from this participant's point of view, different views on the system boundaries would only cause controversy and the conversation around it would continue for a long time without reaching an agreement. The reasoning for the *Extreme Condition Validation* test is that STPA already is a worst-case scenario analysis as is conceived so extreme conditions should have been considered already within the analysis. This interviewee also highlighted that the analysis team does not have influence over anything outside of the boundary and that therefore, the boundaries should be defined carefully and validated which would be done using the *Boundary Validation* test.

Most interviewees agreed that the framework provides a good foundation to formalize the STPA validation process. However, several experts found that they may face challenges in applying some tests in practical cases. A common comment is the lack of clear guidance on how exactly to perform each test and some experts suggested developing a formalized technique for each test. Several STPA experts consider the guide questions too generic and not exhaustive. That is, there is a lack of evidence to show that asking the proposed guide questions for each test would cover everything and there is no guarantee that they do not need to ask any other guide questions.

Some experts also believed that while it is helpful to have an explicit, documented STPA validation framework that can be used as a guideline, the framework may need to be amended for each use case. For example, the framework can be helpful for small organizations "as is", while mature organizations may add more tests to be added to this

framework. Some interviewees expressed concerns about probative blindness, a phenomenon that creates false assurance about safety (Rae & Alexander, 2017). One interviewee also questioned how the proposed framework could lead to better results and how it could be shown.

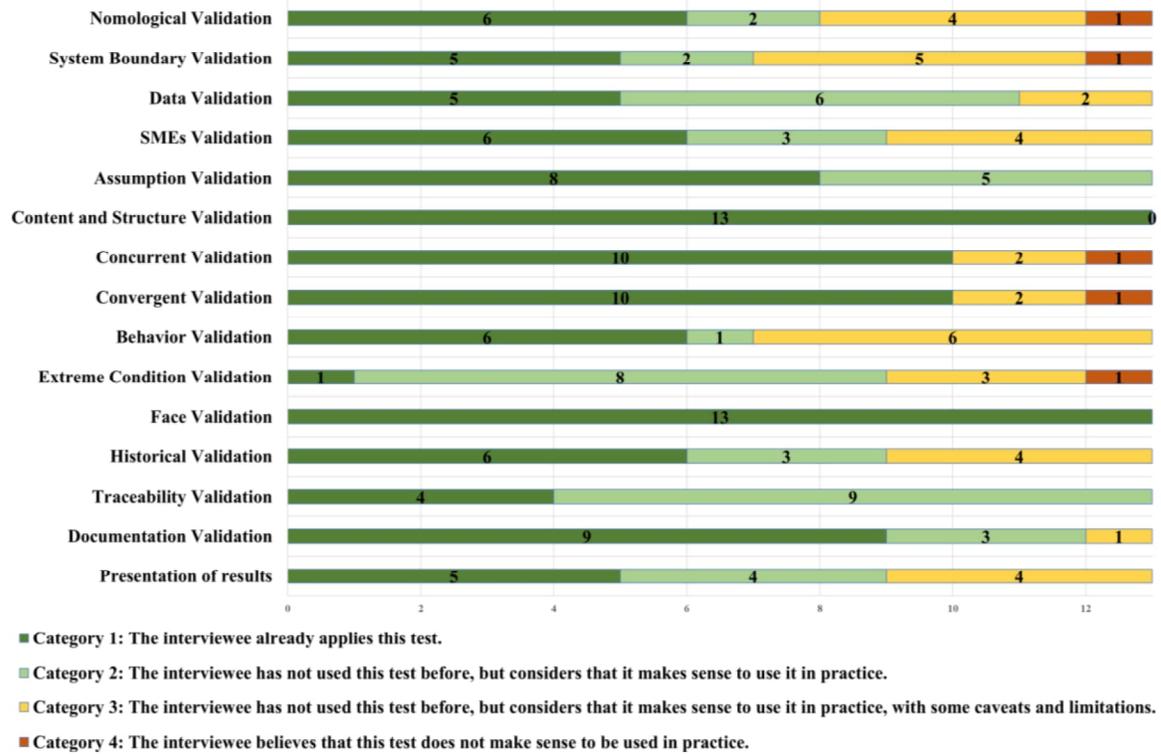


Figure 13. Interviewees' experiences and opinions on each test, adopted from PIV

CHAPTER 4: Discussion

As with any research study, this research is not without limitations. To address these and further advance research on risk and hazard analysis validation, future research directions are discussed. These can be broadly categorized into two areas: further improvements in the general area of risk and hazard analysis validation, and further improvements related to the proposed STPA validation framework. The former is explained in Section 4.1, and the latter is elaborated in Section 4.2. The summary of the future research directions is also illustrated in Figure 14.

4.1 Further Improvements in the General Concept of Risk and Hazard Analysis Validation

Improving the general concept of risk and hazard analysis validation hinges on an important question: can validation, particularly a formalized validation framework or process, truly improve safety outcomes? This question is closely linked to the concept of the effectiveness of risk and safety analysis, as outlined in Section 1.4. Goerlandt et al. (2017b) provide a discussion on the effectiveness of QRA validation, where they highlight a lack of sufficient focus on this topic. For instance, they report a lack of research on the effectiveness of benchmark exercises and reality checks.

Performing validation without evidence of its efficacy can lead to the occurrence of a phenomenon called probative blindness, i.e. the false assurance of the safety of an activity, where this assurance is not aligned with reality (Rae & Alexander, 2017). If validation practices are perceived as effective while they are not, it can lead to unjustified confidence that safety goals have been achieved (Rae & Alexander, 2017). As reported in PIV, some interviewees express concerns about the occurrence of this phenomenon if the proposed STPA validation framework are to be used. However, it is important to note that this issue is not exclusive to the proposed framework. Indeed, it is applicable even to the current practices used by researchers or industry practitioners, such as benchmark exercises as highlighted by Goerlandt et al. (2017b).

Testing the effectiveness of a safety-related technique, including a validation practice, is not a straightforward process, because its results, in terms of improved safety, cannot be

checked directly (Rae et al., 2014). To assess the efficacy of risk and hazard analysis validation, first, it is necessary to develop approaches and criteria that enable such testing. Once developed, they can then be applied to evaluate the efficacy of having validation practices and their effects on actual safety outcomes. Empirical investigation through comparative case study research can be one possible approach. This involves analyzing a case study (or multiple case studies) with and without the application of a specific validation practice and comparing the results using measurable criteria, such as the number of identified hazards.

Even if a formalized validation framework leads to a better hazard analysis, in terms of quality and credibility (Section 1.4), and a safer system, in terms of the effectiveness of an analysis (Section 1.4), its practicality in terms of resource allocation must also be considered. After all, usefulness without cost-effectiveness is limited, as a technique with poor return on investment may actually reduce the overall usefulness of a safety program (Rae et al., 2014). This raises important questions about the cost-effectiveness of validation as a type of safety work, especially for industry practitioners who often struggle to convince stakeholders of its value and necessity, as found in PII.

In order to address this issue, it may be useful to draw on knowledge of human performance in time-critical work, which suggests that errors are more likely to occur when available time is limited (Hall et al., 1982). This knowledge could inform an investigation into the optimal allocation of resources to validate risk and hazard analyses in relation to the level of improvement achieved. The achieved improvement should be identified in relation to both whether the analysis is improved and whether it has an effect on safety.

The results of PI and PII suggest that better training and awareness about safety and validation are required for both researchers and practitioners. This need has been raised already a long time ago by researchers as an improvement opportunity (e.g. Kletz, 2001), yet it remains an ongoing challenge in the field. Improving education is highlighted as an ongoing need, for example by Amyotte et al. (2019) regarding the importance of continuous improvement in process safety education. This implies that there is room to improve the general understanding of the need for safety, hazard analysis, and validation in the engineering profession.

Having raised this as an opportunity for improvement, there is still relatively little evidence to support the idea that awareness and education, whether in the broader field of safety or in the specific area of validation, would have an actual positive safety effect. Although there has been some scholarly work on safety education, e.g., (Mkpat et al., 2018), more work needs to be done in this area to investigate the actual effect of awareness and education on safety. One way to improve this would be the better dissemination of ideas, experiences, and research findings among practitioners as well as safety scientists. This could happen in a form of a conference presentation, publication, or as suggested by a practitioner (refer to PII), knowledge sharing via a database. These all suggest that regulatory bodies, organizations, and academic institutions can do much more to improve the current state of the practice.

4.2 Further Improvements Related to the Proposed STPA Validation Framework

As explained in Section 1.4, the proposed STPA validation framework aims to evaluate the quality and credibility of an STPA analysis, with quality further divided into accuracy and comprehensiveness. These three, which are called the functions of the proposed framework (Figure 3), each can be linked to different types of safety work (Rae & Provan, 2019). The comprehensiveness and accuracy primarily aim to support safety-related decisions (a type of administrative safety work), and ultimately lead to physical safety work (i.e., operational work which would not occur if not for safety concerns). The credibility function is primarily concerned with demonstrated safety (showing safety to stakeholders) as it deals with ensuring that stakeholders can trust the results of an analysis.

Together, these functions can work towards the overall safety of work, which means the absence of harm arising from operational work (Rae & Provan, 2019). However, it is important to acknowledge that this assumption requires testing. It is possible that different types of safety work do not necessarily lead to the safety of work. For example, the various types of safety work may exist for social reasons, but may not truly improve operational safety. Therefore, effectiveness, beyond validation, needs to be examined in order to see whether the various activities are contributing to improving safety.

In light of this, two future research directions are suggested to confirm or disconfirm the usefulness of the proposed framework. First, it is crucial to investigate whether the framework can achieve its envisioned goals by determining whether its application truly enhances the comprehensiveness, accuracy, and credibility of an analysis. One such research study can be using laboratory experiments. Similar studies have been performed to compare the results of STPA with other techniques, e.g. comparison of STPA and FMEA by Sulaman et al. (2019). Research can be done to analyze and compare the results of a case study with and without the application of the proposed STPA validation framework.

Second, should such evidence suggest that the application of the validation framework indeed increases comprehensiveness, accuracy, and credibility, the next step is to investigate whether this leads to improved safety of work. This is in line with the discussion in Section 4.1 regarding effectiveness that is whether it indeed enhances system safety or lowers system risk. As also highlighted in Section 4.1, testing the results of applying validation on the safety of a system is not easy because it cannot be checked directly (Rae et al., 2014). Therefore, a reasonable method for conducting such a study must be developed. One possible approach is to use Quantitative Risk Assessment (QRA). That is, QRA can be performed on the system first after applying STPA and then after applying validation on the results of STPA to comparing the results to see how the risk numbers are changed.

It is important to examine the effectiveness of the proposed framework and to understand the conditions (e.g. the required training) under which the framework can be most effective. This will help to provide a deeper understanding of the impact of the proposed framework on system safety, and its potential to mitigate risks associated with operational work. By addressing these research directions, insights can be gained that can inform further improvements to the framework. Additionally, it could further facilitate the application of this proposed framework in real-case scenarios.

One of the challenges highlighted by STPA experts interviewed in PIV is that the proposed validation framework may need to be customized for each use case. For example, the results of STPA can be used for various purposes, such as developing system architecture and creating safety requirements (Leveson & Thomas, 2018). The proposed validation

framework does not consider any particular purpose for STPA. Another example is that STPA can be used for different stages of a system's lifecycle, such as design and operational contexts, with the same generic implementation steps (Leveson & Thomas, 2018). Similarly, the proposed STPA validation framework is assumed to be applicable to various system lifecycle stages, with the same validation tests and guide questions.

With this framework, it is aimed to offer a high-level guide. Drawing on Gass (1983)'s argument, one validation framework is unlikely to work for all cases in various industrial contexts or for different applications of STPA. Ideally, each project team or company needs to tailor the framework to their specific needs. A promising future research direction would be to develop a modular framework with associated guidance that can be customized to different practical contexts. This way, the relevance and applicability of specific validation tests can be evaluated on a case-by-case basis to ensure that the framework is effective for the particular project context.

One of the limitations of the proposed validation framework highlighted in PIII, which is also raised by the interviewees in PIV, is that the proposed tests and the guide questions may not provide sufficiently clear guidance on how each test can be used in a real case study. While the framework aims to highlight key areas of focus, additional research may be necessary to clarify the practical use of each test. Thus, proposing a formalized technique for each test to tackle this challenge could be a fruitful future research direction. For instance, there are several ways to enhance the utilization of *convergent and concurrent validation* tests. One such approach is to create techniques for comparing systems based on complexity or establishing criteria to differentiate between similar or identical systems. To facilitate the use of the *Nomological Validation* test, it could be beneficial to propose methods for constructing a Nomological map to compare STPA with hazard analyses of identical or similar systems. It could be also helpful to develop techniques to assess the criticality and effects of assumptions.

Additionally, research may explore the integration of the proposed STPA validation framework into the overall safety assurance process. In this thesis, STPA validation is treated as a separate and independent process and is not integrated with other Validation and Verification (V&V) activities of the whole system engineering process. Future research

could explore the interplay between STPA validation and V&V activities as an integrated set of processes, instead of solely validating STPA independently.

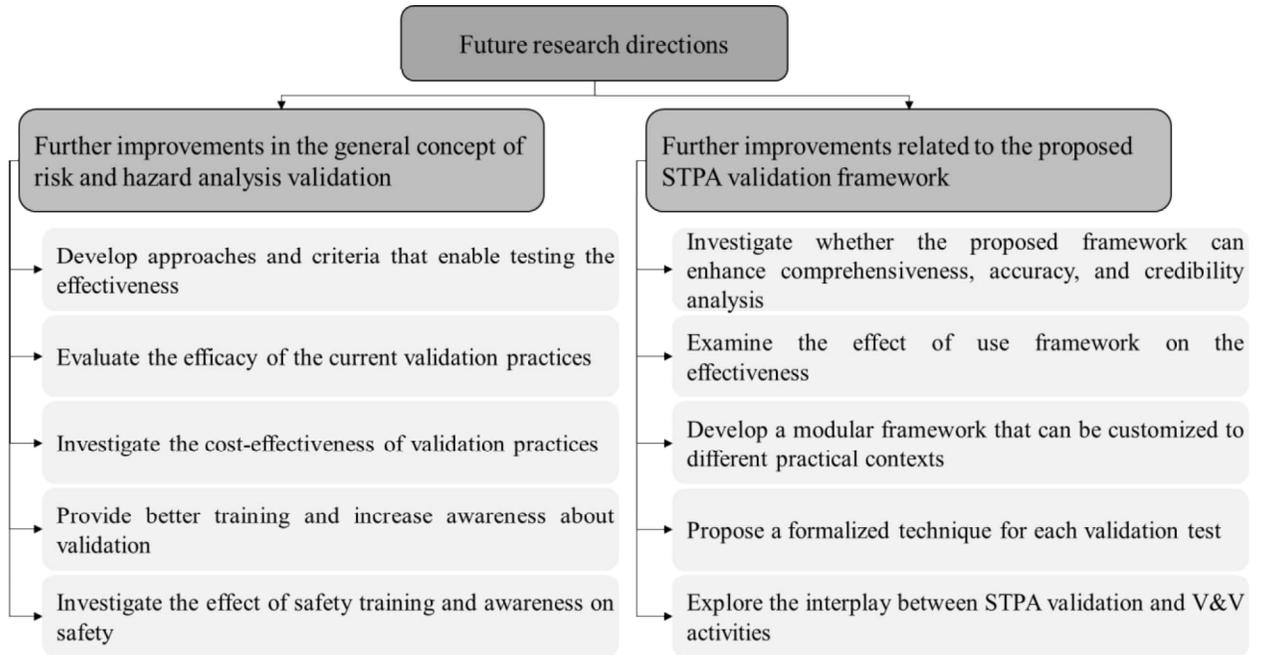


Figure 14. Summary of the future research directions

CHAPTER 5: Conclusion

The general aim of this thesis has been to assess the state of the practice in validation of risk and hazard analysis techniques in both academia and industry. This has been used as a basis for further proposing a solution to an existing challenge, which is a lack of clear guidance on how to perform validation, for a specific hazard identification technique.

The empirical analysis in PI, as a response to RQ 1, showed that validation has not been a focus in the model-based safety analysis, where hazard analysis is one of the identified model types. Seven validation approaches have been identified with illustration and benchmark exercise being the most frequently used validation techniques. In order to investigate RQ 2, an interview study is performed with system safety practitioners to investigate the state of the current validation practices in safety-critical industries. The results of PII show that although practitioners see value in validation and strive to validate their analysis, a lack of clear guidance on how to do validation has created a challenge for them. Thus, proposing a formalized validation framework is identified as one of the most important future research directions which could alleviate this issue.

To address this need, RQs 3 to 5 have been introduced, aiming to explore the development of a validation framework that can be utilized by both researchers and practitioners. Drawing on the insights gained from answering RQs 1 and 2, as well as theoretical validation concepts in related fields, a validation framework is proposed for a specific hazard analysis technique, namely STPA. The proposed STPA validation framework, presented in PIII, consists of 15 validation tests, each targeting a specific element of an STPA analysis. These tests are designed as high-level guide questions to further facilitate the use of the framework.

In response to RQs 4 and 5, the reasonableness of the validation framework is evaluated in PIV through an empirical approach. This aims to compare the current STPA validation practices with the ideas and tests in the framework as well as to test the reasonableness of the developed framework using STPA experts' judgments. The findings of this research indicate that all interviewed STPA experts acknowledge the importance of validation in an STPA analysis and endeavor to carry it out, although it may not always be feasible, due to practical challenges, such as the project's tight schedule. All theory-based proposed

validation tests have already been used in practice by at least one STPA expert. The most frequently used validation tests are found to be the *Face Validation* and *Content and Structure Validation* tests, which are already commonly applied by all interviewees. Most interviewees agreed that the framework provides a good foundation to formalize the STPA validation process, but they may face challenges in applying some tests due to a lack of clear guidance on how to perform each test in a real case study.

Overall, this thesis has highlighted the importance of validation in risk and hazard analysis. Following empirical work to investigate the extent of the issue of lack of focus in risk and hazard analysis validation, it has contributed to closing this gap by proposing a formalized structure for validation. Further empirical research focused on the reasonableness of the framework has revealed several strengths and weaknesses. While various limitations are highlighted and future research directions are outlined to confirm and advance the outcome of the work presented, the overall objectives of the thesis have been achieved.

REFERENCES

- Abdellatif, A. A., & Holzapfel, F. (2020). Model Based Safety Analysis (MBSA) Tool for Avionics Systems Evaluation. *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, 1–5. <https://doi.org/10.1109/DASC50938.2020.9256578>
- Agresti, A. (2007). *Introduction to Categorical Data Analysis* (2nd ed.). John Wiley & Sons.
- Alpeev, A. S. (2019). Safety Terminology: Deficiencies and Suggestions. *Atomic Energy (New York, N.Y.)*, *126*(5), 339–341. <https://doi.org/10.1007/s10512-019-00560-y>
- Amyotte, P., Khan, F., & Irvine, Y. (2019). Continuous Improvement in Process Safety Education. *Chemical Engineering Transactions*, *77*.
<https://doi.org/10.3303/CET1977069>
- Augusiak, J., Van den Brink, P. J., & Grimm, V. (2014). Merging validation and evaluation of ecological models to ‘evaludation’: A review of terminology and a practical approach. *Ecological Modelling*, *280*, 117–128.
<https://doi.org/10.1016/j.ecolmodel.2013.11.009>
- Aven, T. (2012). Foundational Issues in Risk Assessment and Risk Management. *Risk Analysis*, *32*(10), 1647–1656. <https://doi.org/10.1111/j.1539-6924.2012.01798.x>
- Aven, T. (2014). What is safety science? *Safety Science*, *67*, 15–20.
<https://doi.org/10.1016/j.ssci.2013.07.026>
- Aven, T. (2017). Risk Analysis Validation and Trust in Risk Management: A postscript. *Safety Science*, *99*, 255–256. <https://doi.org/10.1016/j.ssci.2017.08.009>

- Aven, T., Ben-Haim, Y., Andersen, H. B., Cox, T., Droguett, E. L., Greenberg, M., Guikema, S., Kröger, W., Renn, O., Thompson, K. M., & Zio, E. (2018). *Society for Risk Analysis Glossary*. 9.
- Aven, T., & Guikema, S. (2011). Whose uncertainty assessments (probability distributions) does a risk assessment report: The analysts' or the experts'? *Reliability Engineering & System Safety*, *96*(10), 1257–1262.
<https://doi.org/10.1016/j.ress.2011.05.001>
- Aven, T., & Heide, B. (2009). Reliability and validity of risk analysis. *Reliability Engineering & System Safety*, *94*(11), 1862–1868.
<https://doi.org/10.1016/j.ress.2009.06.003>
- Aven, T., & Renn, O. (2009). On risk defined as an event where the outcome is uncertain. *Journal of Risk Research*, *12*(1), 1–11.
<https://doi.org/10.1080/13669870802488883>
- Aven, T., & Zio, E. (2014). Foundational Issues in Risk Assessment and Risk Management. *Risk Analysis*, *34*(7), 1164–1172. <https://doi.org/10.1111/risa.12132>
- Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, *12*(3), 183–210. [https://doi.org/10.1002/\(SICI\)1099-1727\(199623\)12:3<183::AID-SDR103>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1099-1727(199623)12:3<183::AID-SDR103>3.0.CO;2-4)
- Baybutt, P. (2021). On the need for system-theoretic hazard analysis in the process industries. *Journal of Loss Prevention in the Process Industries*, *69*, 104356.
<https://doi.org/10.1016/j.jlp.2020.104356>

Bhattacharjee, A. (2012). *Social Science Research: Principles, Methods, and Practices*.

Digital Commons University of South Florida.

Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: A research note.

Qualitative Research : QR, 8(1), 137–152.

<https://doi.org/10.1177/1468794107085301>

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative*

Research in Psychology, 3(2), 77–101.

<https://doi.org/10.1191/1478088706qp063oa>

Brummett, B. (2019). *Techniques of close reading* (2nd edition.). SAGE Publications,

Inc.

Busby, J. s., & Hughes, E. j. (2006). Credibility in risk assessment: A normative

approach. *International Journal of Risk Assessment and Management*, 6(4–6),

508–527. <https://doi.org/10.1504/IJRAM.2006.009542>

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research techniques and*

procedures for developing grounded theory. (3e [ed.] / Juliet Corbin, Anselm

Strauss.). SAGE.

Dakwat, A. L., & Villani, E. (2018). System safety assessment based on STPA and model

checking. *Safety Science*, 109, 130–143. <https://doi.org/10.1016/j.ssci.2018.05.009>

Dallat, C., Salmon, P. M., & Goode, N. (2019). Risky systems versus risky people: To

what extent do risk assessment methods consider the systems approach to accident

causation? A review of the literature. *Safety Science*, 119, 266–279.

<https://doi.org/10.1016/j.ssci.2017.03.012>

- Dulac, N. (2007). *A Framework for Dynamic Safety and Risk Management Modeling in Complex Engineering Systems* [Massachusetts Institute of Technology].
<http://sunnyday.mit.edu/safer-world/dulac-dissertation.pdf>
- Eker, S., Rovenskaya, E., Langan, S., & Obersteiner, M. (2019). Model validation: A bibliometric analysis of the literature. *Environmental Modelling & Software*, 117, 43–54. <https://doi.org/10.1016/j.envsoft.2019.03.009>
- Engel, A. (2010). *Verification, Validation, and Testing of Engineered Systems*. Wiley.
<http://ezproxy.library.dal.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=e000xna&AN=329992&site=ehost-live>
- Ericson, C. (2015). *Hazard Analysis Techniques for System Safety* (2nd ed.). Wiley.
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2015). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), Article 1. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Fleming, C. H., Spencer, M., Thomas, J., Leveson, N., & Wilkinson, C. (2013). Safety assurance in NextGen and complex transportation systems. *Safety Science*, 55, 173–187. <https://doi.org/10.1016/j.ssci.2012.12.005>
- Forrester, J., & Senge, P. (1980). *Tests for building confidence in system dynamics models* (pp. 209–228).
- Gass, S. I. (1983). Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis. *Operations Research*, 31(4), 603–631.

- Glette-Iversen, I., Aven, T., & Flage, R. (2022). The concept of plausibility in a risk analysis context: Review and clarifications of defining ideas and interpretations. *Safety Science*, *147*, 105635. <https://doi.org/10.1016/j.ssci.2021.105635>
- Goerlandt, F., Khakzad, N., & Reniers, G. (2017a). Special Issue: Risk Analysis Validation and Trust in Risk management. *Safety Science*, *99*, 123–126. <https://doi.org/10.1016/j.ssci.2017.07.012>
- Goerlandt, F., Khakzad, N., & Reniers, G. (2017b). Validity and validation of safety-related quantitative risk analysis: A review. *Safety Science*, *99*, 127–139. <https://doi.org/10.1016/j.ssci.2016.08.023>
- Goerlandt, F., & Reniers, G. (2018). Prediction in a risk analysis context: Implications for selecting a risk perspective in practical applications. *Safety Science*, *101*, 344–351. <https://doi.org/10.1016/j.ssci.2017.09.007>
- Hale, A. (2014). Foundations of safety science: A postscript. *Safety Science*, *67*, 64–69. <https://doi.org/10.1016/j.ssci.2014.03.001>
- Hall, R. E., Fragola, J., & Wreathall, J. (1982). *Post-event human decision errors: Operator action tree/time reliability correlation* (p. 48).
- Hansson, S. O. (2012). Safety is an inherently inconsistent concept. *Safety Science*, *50*(7), 1522–1527. <https://doi.org/10.1016/j.ssci.2012.03.003>
- Harkleroad, E., Vela, A., & Kuchar, J. (2013). *Review of Systems-Theoretic Process Analysis (STPA) Method and Results to Support NextGen Concept Assessment and Validation* (ATC-427).

- Hecke, T. V. (2012). Power study of anova versus Kruskal-Wallis test. *Journal of Statistics and Management Systems*, 15(2–3), 241–247.
<https://doi.org/10.1080/09720510.2012.10701623>
- Hollnagel, E. (2014). Is safety a subject for science? *Safety Science*, 67, 21–24.
<https://doi.org/10.1016/j.ssci.2013.07.025>
- Hulme, A., Stanton, N. A., Walker, G. H., Waterson, P., & Salmon, P. M. (2022). Testing the reliability and validity of risk assessment methods in Human Factors and Ergonomics. *Ergonomics*, 65(3), 407–428.
<https://doi.org/10.1080/00140139.2021.1962969>
- Kaplan, S. (1992). ‘Expert information’ versus ‘expert opinions’. Another approach to the problem of eliciting/ combining/using expert knowledge in PRA. *Reliability Engineering & System Safety*, 35(1), 61–72. [https://doi.org/10.1016/0951-8320\(92\)90023-E](https://doi.org/10.1016/0951-8320(92)90023-E)
- Keys, P. (1988). System dynamics: A methodological perspective. *Transactions of the Institute of Measurement and Control*, 10(4), 218–224.
<https://doi.org/10.1177/014233128801000406>
- Kletz, T. A. (2001). *Learning from accidents* (3rd ed.). Gulf Professional Publishing.
- Landry, M., Malouin, J.-L., & Oral, M. (1983). Model validation in operations research. *European Journal of Operational Research*, 14(3), 207–220.
[https://doi.org/10.1016/0377-2217\(83\)90257-6](https://doi.org/10.1016/0377-2217(83)90257-6)

- Lathrop, J., & Ezell, B. (2017). A systems approach to risk analysis validation for risk management. *Safety Science*, 99, 187–195.
<https://doi.org/10.1016/j.ssci.2017.04.006>
- Law, A. (2014). *Simulation Modeling and Analysis* (5th edition). McGraw Hill.
- Le Coze, J.-C. (2019). *Safety Science Research Evolution, Challenges and New Directions*. CRC Press LLC.
- Leveson. (2012). *Engineering a Safer World: Systems Thinking Applied to Safety*. Cambridge, Mass : The MIT Press. https://web-s-ebSCOhost-com.ezproxy.library.dal.ca/ehost/ebookviewer/ebook/ZTAwMHhuYV9fNDIxODE4X19BTg2?sid=e9969089-f149-426b-bb6e-776f0eca0b81@redis&vid=0&format=EB&lpid=lp_1&rid=0
- Leveson. (2017). Rasmussen’s legacy: A paradigm change in engineering for safety. *Applied Ergonomics*, 59, 581–591. <https://doi.org/10.1016/j.apergo.2016.01.015>
- Leveson, N. (2004a). A new accident model for engineering safer systems. *Safety Science*, 42(4), 237–270. [https://doi.org/10.1016/S0925-7535\(03\)00047-X](https://doi.org/10.1016/S0925-7535(03)00047-X)
- Leveson, N. (2004b). A systems-theoretic approach to safety in software-intensive systems. *IEEE Transactions on Dependable and Secure Computing*, 1(1), 66–86.
<https://doi.org/10.1109/TDSC.2004.1>
- Leveson, N., & Thomas, J. (2018). *STPA Handbook*.
https://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf
- Li, J., & Hale, A. (2016). Output distributions and topic maps of safety related journals. *Safety Science*, 82, 236–244. <https://doi.org/10.1016/j.ssci.2015.09.004>

- Magaldi, D., & Berler, M. (2020). Semi-structured Interviews. In *Encyclopedia of Personality and Individual Differences* (pp. 4825–4830). Springer International Publishing. https://doi.org/10.1007/978-3-319-24612-3_857
- McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, *46*(1), 38–41.
<https://doi.org/10.1016/j.bjoms.2007.09.002>
- Mkpat, E., Reniers, G., & Cozzani, V. (2018). Process safety education: A literature review. *Journal of Loss Prevention in the Process Industries*, *54*, 18–27.
<https://doi.org/10.1016/j.jlp.2018.02.003>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, *339*, b2535. <https://doi.org/10.1136/bmj.b2535>
- Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, *22*(3), 453–460.
<https://doi.org/10.1016/j.jcps.2011.08.002>
- Oberkampf, W. L., & Trucano, T. G. (2008). Verification and validation benchmarks. *Nuclear Engineering and Design*, *238*(3), 716–743.
<https://doi.org/10.1016/j.nucengdes.2007.02.032>
- Oral, M., & Kettani, O. (1993). The facets of the modeling and validation process in operations research. *European Journal of Operational Research*, *66*(2), 216–234.
[https://doi.org/10.1016/0377-2217\(93\)90314-D](https://doi.org/10.1016/0377-2217(93)90314-D)

- O'Reilly, M., & Parker, N. (2013). 'Unsatisfactory Saturation': A critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research : QR*, 13(2), 190–197. <https://doi.org/10.1177/1468794112446106>
- Patriarca, R., Chatzimichailidou, M., Karanikas, N., & Di Gravio, G. (2022). The past and present of System-Theoretic Accident Model And Processes (STAMP) and its associated techniques: A scoping review. *Safety Science*, 146, 105566. <https://doi.org/10.1016/j.ssci.2021.105566>
- Pitchforth, J., & Mengersen, K. (2013). A proposed validation framework for expert elicited Bayesian Networks. *Expert Systems with Applications*, 40(1), 162–167.
- Provan, D. J., Rae, A. J., & Dekker, S. W. (2019). An ethnography of the safety professional's dilemma: Safety work or the safety of work? *Safety Science*, 117, 276–289. <https://doi.org/10.1016/j.ssci.2019.04.024>
- QSR International Pty Ltd. NVivo (released in March 2020). (2020). <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>
- Rae, A., & Alexander, R. D. (2017). Probative blindness and false assurance about safety. *Safety Science*, 92, 190–204. <https://doi.org/10.1016/j.ssci.2016.10.005>
- Rae, A., Alexander, R., & McDermid, J. (2014). Fixing the cracks in the crystal ball: A maturity model for quantitative risk assessment. *Reliability Engineering & System Safety*, 125, 67–81. <https://doi.org/10.1016/j.ress.2013.09.008>
- Rae, A., & Provan, D. (2019). Safety work versus the safety of work. *Safety Science*, 111, 119–127. <https://doi.org/10.1016/j.ssci.2018.07.001>

- Rae, A., Provan, D., Aboelssaad, H., & Alexander, R. (2020). A manifesto for Reality-based Safety Science. *Safety Science*, *126*, 104654.
<https://doi.org/10.1016/j.ssci.2020.104654>
- Rae, McDermid, & Alexander. (2012). *The Science and Superstition of Quantitative Risk Assessment*.
- Rae, Nicholson, & Alexander. (2010). The state of practice in system safety research evaluation. *5th IET International Conference on System Safety 2010*, 1–8.
<https://doi.org/10.1049/cp.2010.0838>
- Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science*, *27*(2), 183–213. [https://doi.org/10.1016/S0925-7535\(97\)00052-0](https://doi.org/10.1016/S0925-7535(97)00052-0)
- Reiman, T., & Viitanen, K. (2019). Towards Actionable Safety Science. In *Safety Science Research* (pp. 203–222). CRC Press. <https://doi.org/10.4324/9781351190237-13>
- Reniers, G., & Anthone, Y. (2012). A ranking of safety journals using different measurement methods. *Safety Science*, *50*(7), 1445–1451.
<https://doi.org/10.1016/j.ssci.2012.01.017>
- Rosa, E. A. (1998). Metatheoretical foundations for post-normal risk. *Journal of Risk Research*, *1*(1), 15–44. <https://doi.org/10.1080/136698798377303>
- Rosqvist, T. (2010). On the validation of risk analysis—A commentary. *Reliability Engineering & System Safety*, *95*(11), 1261–1265.
<https://doi.org/10.1016/j.ress.2010.06.002>
- Rouhiainen, V. (1992). QUASA: A method for assessing the quality of safety analysis. *Safety Science*, *15*(3), 155–172. [https://doi.org/10.1016/0925-7535\(92\)90002-H](https://doi.org/10.1016/0925-7535(92)90002-H)

- Saleh, J. H., Marais, K. B., Bakolas, E., & Cowlagi, R. V. (2010). Highlights from the literature on accident causation and system safety: Review of major ideas, recent contributions, and challenges. *Reliability Engineering & System Safety*, *95*(11), 1105–1116. <https://doi.org/10.1016/j.res.2010.07.004>
- Sandelowski, M. (2004). Using Qualitative Research. *Qualitative Health Research*, *14*(10), 1366–1386. <https://doi.org/10.1177/1049732304269672>
- Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation*, *7*(1), 12–24. <https://doi.org/10.1057/jos.2012.20>
- Schwanitz, V. J. (2013). Evaluating integrated assessment models of global climate change. *Environmental Modelling & Software*, *50*, 120–131. <https://doi.org/10.1016/j.envsoft.2013.09.005>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Solberg, Ø., & Njå, O. (2012). Reflections on the ontological status of risk. *Journal of Risk Research*, *15*(9), 1201–1215. <https://doi.org/10.1080/13669877.2012.713385>
- Sulaman, S. M., Beer, A., Felderer, M., & Höst, M. (2019). Comparison of the FMEA and STPA safety analysis methods—a case study. *Software Quality Journal*, *27*(1), 349–387. <https://doi.org/10.1007/s11219-017-9396-0>
- Suokas, J. (1985). *On the reliability and validity of safety analysis: Dissertation* [Dissertation]. VTT Technical Research Centre of Finland.

Thomas, J., Lemos, F. L., & Leveson, N. (2012). *Evaluating the Safety of Digital Instrumentation and Control Systems in Nuclear Power Plants* (Research Report NRC-HQ-11-6-04-0060; p. 66).

Webster, J., & Watson, R. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), xiii–xxiii.

APPENDIX

Appendix A. List of Publications

This thesis consists of an overview of the following publications, which are referred to in the thesis summary by their Roman numerals.

PI. The State of the Practice in Validation of Model-based Safety Analysis in Socio-technical Systems: An Empirical Study

PII. Validation of system safety hazard analysis in safety-critical industries: An interview study with industry practitioners

PIII. A proposed validation framework for the System Theoretic Process Analysis (STPA) technique

PIV. Evaluation of the proposed System Theoretic Process Analysis (STPA) validation framework by seeking STPA experts' judgments

Publication I

Sadeghi, & Goerlandt, F. (2021). The State of the Practice in Validation of Model-Based Safety Analysis in Socio-Technical Systems: An Empirical Study. *Safety (Basel)*, 7(4), 72–. <https://doi.org/10.3390/safety7040072>

Article

The State of the Practice in Validation of Model-Based Safety Analysis in Socio-Technical Systems: An Empirical Study

Reyhaneh Sadeghi *  and Floris Goerlandt

Department of Industrial Engineering, Dalhousie University, Halifax, NS B3J 1B6, Canada; floris.goerlandt@dal.ca
* Correspondence: Reyhaneh.sadeghi@dal.ca

Abstract: Even though validation is an important concept in safety research, there is comparatively little empirical research on validating specific safety assessment, assurance, and ensurance activities. Focusing on model-based safety analysis, scant work exists to define approaches to assess a model's adequacy for its intended use. Rooted in a wider concern for evidence-based safety practices, this paper intends to provide an understanding of the extent of this problem of lack of validation to establish a baseline for future developments. The state of the practice in validation of model-based safety analysis in socio-technical systems is analyzed through an empirical study of relevant published articles in the *Safety Science* journal spanning a decade (2010–2019). A representative sample is first selected using the PRISMA protocol. Subsequently, various questions concerning validation are answered to gain empirical insights into the extent, trends, and patterns of validation in this literature on model-based safety analysis. The results indicate that no temporal trends are detected in the ratio of articles in which models are validated compared to the total number of papers published. Furthermore, validation has no clear correlation with the specific model type, safety-related concept, different system life cycle stages, industries, or with the countries from which articles originate. Furthermore, a wide variety of terminology for validation is observed in the studied articles. The results suggest that the safety science field concerned with developing and applying models in safety analyses would benefit from an increased focus on validation. Several directions for future work are discussed.

Keywords: validation; model-based safety analysis; risk; resilience; reliability; socio-technical systems



Citation: Sadeghi, R.; Goerlandt, F. The State of the Practice in Validation of Model-Based Safety Analysis in Socio-Technical Systems: An Empirical Study. *Safety* **2021**, *7*, 72. <https://doi.org/10.3390/safety7040072>

Academic Editor: Raphael Grzebieta

Received: 14 February 2021
Accepted: 12 October 2021
Published: 18 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. The Need to Understand the State of the Practice in Validation of Model-Based Safety Analysis

Safety science is an interdisciplinary field of study that contains a broad range of theories, ideas, and scientific traditions [1]. While research in safety science produces many ideas and approaches for safety assessment, assurance, and ensurance, few of these are systematically tested according to academic procedures. Consequently, the scientific validity of many approaches and activities remains open for debate, which is one of the several factors contributing to difficulties in establishing evidence-based safety practices [2,3]. This, furthermore, can lead to uncertainty in industrial contexts in terms of choosing concepts and tools in what has been labeled by some practitioners a “nebulous safety cloud” [4].

In papers and commentaries addressing fundamental issues in risk and safety science, lack of focus on validation has repeatedly been raised as an important issue [5–7]. Indeed, validation has been discussed and investigated in some subfields of safety science. For instance, the effectiveness of the occupational health and safety management system is reviewed through a systematic literature review [8]. Another example is a literature review paper focusing on maturity models for assessing safety culture [9]. In this paper, the authors assert that “validity of the use of maturity models to assess safety culture tends to be the exception, rather than the rule”. Furthermore, some articles have been published in

which the results of different safety analysis methods are compared through a case study, such as a comparison of FMEA and STPA safety analysis methods [10] and a comparison of three waterway risk analysis methods [11]. This can also be seen in the work of Suokas and Kakko [12], Amendola et al. [13], and Laheij et al. [14] where comparative model-based safety analyses are presented in different industrial contexts.

Notwithstanding the existence of such comparative case studies, there is very little explicit focus on validation of model-based safety analysis, in the sense of providing evidence that models are useful as intended in the envisaged practical contexts. From an academic perspective, the extent of this problem is furthermore not clearly understood, i.e., there is, to the best of the authors' knowledge, a lack of systematic evidence of the extent to which model-based safety analyses presented in the academic literature have been validated. This is not merely an academic issue. This can also lead to uncertainties in industrial contexts, because it may result in models being implemented and used even though they provide unreliable, incomplete, or even misleading results [4]. From a practical safety perspective, the scientific validation of models should thus be a concern.

Validation has been discussed elaborately in different fields focusing on the development of modeling approaches, such as system dynamics [15], simulation [16], and environmental and decision sciences [17]. In contrast, although model-based safety analyses are widely applied in the academic field of safety science and practical safety work, there is scant literature on validation of model-based safety analysis [7,18]. The literature on model-based safety analysis has been mainly focused on proposing new models, adjusting or integrating existing ones, or employing an existing model to obtain insights into safety issues for particular problems in various industries, such as the chemical industry [19], the nuclear industry [20], the maritime industry [21], and the transportation industry, including railway [22] and road safety [23]. However, establishing the validity of such models is still a major challenge. Goerlandt et al. [6] argue that the reason for this challenge is two-fold. First, there are different perspectives on how to understand validation as a concept. Second, there is a lack of consensus of appropriate criteria and processes for how to assess validity, or sometimes even a lack of awareness that such criteria need to be specified.

In modeling contexts, validation is often seen as an important step to establish the credibility of a model [17], so that it can be used appropriately as a basis for practical decision making. Hence, model validation is an important topic in general, and arguably, even more so in a safety context since the results obtained from a safety analysis model can exert a considerable influence on safety improvements [24].

To the best of our knowledge, there is a lack of research aiming to provide empirical insights into the state of the practice in validation in the context of model-based safety analysis in socio-technical systems. Thus, the current work intends to address this gap as a step towards understanding the extent of this problem in the scientific community. In addition to providing a baseline understanding of the state of the practice, the aim is to raise further questions and to explore pathways for improving the current situation.

1.2. Scope of This Research

Before stating the research questions, the scope of this research needs to be clarified. The first issue concerns the meaning of model-based safety analysis in the context of this work. In general, models are a way to provide information in a simplified form [25]. Complex systems cannot be comprehensively understood without modeling [26]. Models make informal ideas formal and clear, based on which implications of the underlying assumptions can be systematically approached [27]. The purpose and use of models vary, ranging from prediction to social learning [28], and they may describe components, processes, organizations, events, dependencies, factors, or causation [25]. In our current study, we include different types of models, such as mathematical, statistical, and qualitative, which are also sometimes referred to as methods, approaches, and/or frameworks in the literature. Although these terms may have different meanings in different contexts, in the scope of this research, they are all taken to have the overall objective of dealing

with a safety challenge in a socio-technical system through a structured way of thinking involving the development of a model-based representation of safety-relevant aspects of a socio-technical system.

Second, in addition to safety, the closely related concepts of risk, reliability, and resilience are also included in the scope of this research, hence including a wide range of model-based safety analyses. These concepts represent different approaches to achieve safety, often based on diverging theoretical commitments to accident causation. Accordingly, these concepts are collectively referred to as “safety concept(s)” throughout this paper.

Third, for clarity of scope, we define the term socio-technical systems, as this is the context in which we frame our study of model-based safety analysis validation. According to Kroes et al. [29], modern complex systems comprise different elements: social institutions, human agents, and technical artifacts, which interact to deliver outcomes that cannot be achieved by humans or technology in isolation. Therefore, such systems, known as socio-technical systems, need to be investigated in terms of their interactions and interrelationships between the relevant human, technical, and social aspects.

Fourth, this research only focuses on studies addressing harm/accidents to people or systems (human and industrial safety). Thus, other types of risks, such as financial or environmental risks, are excluded from the scope.

Finally, we limited the scope of this research to one journal, *Safety Science*, which publishes work on model-based safety analysis in complex socio-technical systems. There are two main reasons for this scope limitation. First, it proved unfeasible to accurately delineate the wider literature of model-based safety analysis. Second, a poorly defined study population would lead to significant methodological flaws and unreliable results. The journal *Safety Science* was selected as it is one of the leading journals in safety research, with a comparatively long publication history [30]. It is among the highest-ranked journals in safety research, with a high reputation among academics [31], and hence is widely considered to be academically impactful. Furthermore, as a multidisciplinary journal, model-based safety analyses represent an important cluster in its publication records [30]. Based on this, further acknowledging that related empirical work on the state of practice of system safety evaluation [2] makes a similar scope limitation; the authors believe limiting the scope to *Safety Science* to be a defensible choice for the current purposes.

1.3. Research Questions

The main, overarching research question of this paper is “What is the state of practice in the academic literature regarding the validation of model-based safety analysis in socio-technical systems?” To more precisely answer this broad question based on empirical insights, the relevant literature is interpreted considering the following specific sub-questions:

RQ 1. In what percentage of relevant published articles did the authors attempt to validate their models?

RQ 2. Which validation approaches are used for model-based safety analysis in the articles, and what are the frequencies of the approaches?

RQ 3. Is there any trend in the ratio of the number of articles in which models are validated to the total number of papers in each year?

RQ 4. Are articles utilizing specific model types more likely to address validation?

RQ 5. Are articles focusing on a specific safety concept more likely to address validation?

RQ 6. Are articles focusing on a specific stage of a system life cycle more likely to address validation?

RQ 7. Are articles proposing a model for a specific industry more likely to address validation?

RQ 8. Are articles originating from specific countries more likely to address validation?

RQ 9. What terminology is used for validation, and what are the frequencies of the terms used?

RQ 1 is chosen to investigate the percentage of the papers in our sample in which the models were validated. It has been raised previously that validation has not been a topic of much explicit focus in safety research, but there is no empirical evidence available

regarding the extent of this issue in articles proposing or using models to analyze safety in socio-technical systems. Hence, this question aims to contribute to building evidence.

RQ 2 is selected to investigate the existing validation approaches in the model-based safety analysis in the literature. The aim of this is to shed some light as to what authors believe they should do to validate a model-based safety analysis. As there are different approaches available, with their comparative merits and limitations not conclusively agreed upon in the academic and professional communities, the relative frequency of different validation approaches is of interest.

RQ 3 is included to investigate whether validation has gained more attention over time in the studied period. As mentioned in Section 1.1, several articles and commentaries about fundamental issues in risk and safety science have raised the lack of focus on validation in safety research as an important issue. Hence, this question explores whether such commentaries have led to a gradual increase in models being validated by the authors.

RQ 4 is included to explore the hypothesis that some of the safety analysis model types could have been more frequently validated than others. The rationale behind this hypothesis is that, as mentioned in Section 1.1, validation has been more elaborately considered in the parent academic disciplines focusing on the theory and development of specific modeling approaches, such as simulation, which is one model type identified as being used for model-based safety analysis (see Section 2.2.4). The existence of rich validation literature on simulation models would suggest that such models may be more validated also in a safety analysis context. If this is the case, this may suggest a more mature application community, from which proponents and users of other safety analysis model types may learn.

RQ 5 concerns the possible relationship between validation and different relevant concepts to model-based safety analysis. As mentioned in Section 1.2, in addition to safety, the closely related concepts of risk, reliability, and resilience are also included in the scope of this research. As these concepts are the associated analysis methods to a large degree proposed and studied by different communities within safety science, this question investigates whether different conceptual focuses lead to different degrees of attention to validation of the associated models.

In RQ 6, the phase of a system's lifecycle is taken as another factor with a possible relation to the validation of model-based safety analysis. According to Amyotte et al. [32], inherently safe design, which focuses on considering safety requirements early in the design phase and eliminating hazards, is one of the principles that could prevent major accidents. While the subsequent phases of a system's life cycle are clearly important as well, the design phase is often seen as having a major role in the overall system safety performance, with emphasis on the design phase being necessary to avoid re-design and extra costs [33,34]. In addition, considering that validation may not be equally feasible to be performed in practice for analyses in different system lifecycles, this question investigates whether validation has been given more consideration in different stages of the system lifecycle, particularly in the design phase.

In the last two questions (RQ 7 and RQ 8), the assumptions concern the relationship between validation and the countries of origin of the publication, and the industrial sector in which the model is applied. These questions are rooted based on the understanding that safety analyses are often executed as part of regulatory requirements, the specifics of which may differ significantly between countries and industries. Hence, these questions aim to provide some insight into whether such contextual factors lead to significant differences in the degree of validation of model-based safety analyses originating from different countries or industry sectors.

The remainder of this article is organized as follows. In Section 2, the process of constructing the dataset is described, which includes identifying the relevant literature and the sampling strategy. This section also provides a descriptive overview of the resulting sample. Section 3 presents the analysis results, providing answers to the above-listed research questions. Subsequently, Section 4 summarizes the findings and connects the

specific findings of the research questions to make an overall assessment of the state of the practice in validation of model-based safety analysis. This section also identifies the limitations of the study and discusses future research directions. Section 5 concludes.

2. Materials and Methods

2.1. Identifying Relevant Literature and Sampling

The preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement is used to identify, screen, determine eligibility, and include studies for analysis from search results [35]. The flow diagram is shown in Figure 1.

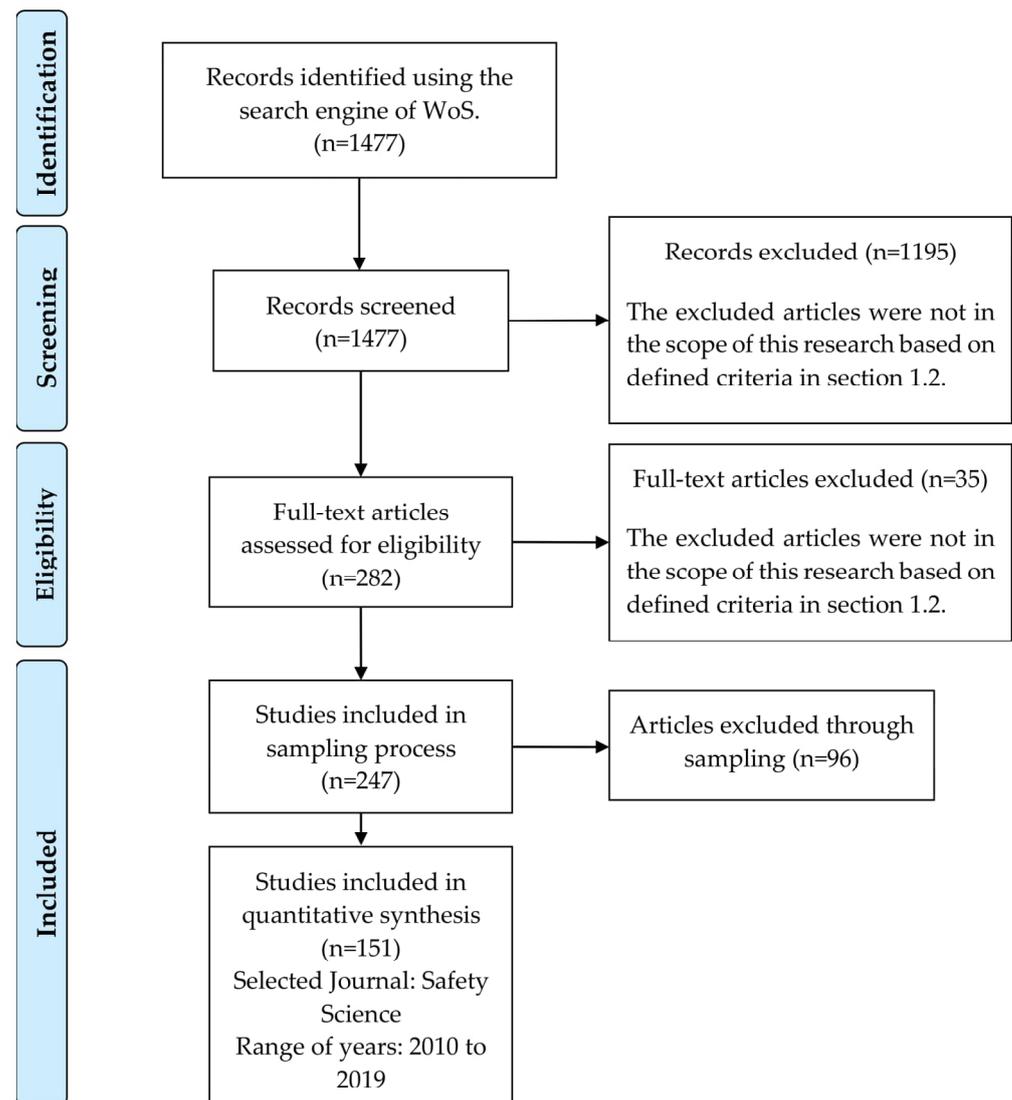


Figure 1. The process of constructing the dataset based on PRISMA Flow Diagram.

In literature analyses, it is critical to use an appropriate set of keywords to identify and include an adequate range of papers [36]. We used two sets of keywords. One of these is a term related to safety, risk, reliability, and/or resilience to identify safety-related papers. The second set of terms relates to our focus on the use of a model, method, approach, and/or framework. The search was executed using Web of Science (WoS) in July 2020, limited to articles published in *Safety Science*. WoS is a database that includes bibliographic information of articles of the world's most impactful and high-quality journals [37]. A further scope restriction is made to retain only published articles in the period 2010 to 2019,

to focus on the more recent developments, and to investigate the potential trends in this period. This research only includes articles written in English.

To identify records following process is performed:

- a. The below query is run on WoS:

$$TS = ((\text{"Safety"} \text{ OR } \text{"Risk"} \text{ OR } \text{"Reliability"} \text{ OR } \text{"Resilience"}) \text{ AND } (\text{"Model"} \text{ OR } \text{"Method"} \text{ OR } \text{"Approach"} \text{ OR } \text{"Framework"}))$$
- b. To limit the search to the *Safety Science* journal, its ISSN code is considered in a new query, which is IS = (0925–7535). Additionally, the 2010 to 2019 period is selected in WoS, further limiting the search.
- c. Finally, the result of the first query is combined with the second query, using the "AND" operator.

The title and abstract of the identified papers were thoroughly reviewed to provide an initial screening for applicability mentioned in Section 1.2. After removing articles in the initial screening phase, the dataset contained 282 documents for further analysis. The text of each of the 282 remaining articles was scrutinized using the close reading method [38] to ensure that they all are within the intended scope of this research. As a result, an additional 35 documents were dropped during the eligibility review, resulting in 247 retained papers. To limit the number of papers for further analysis, a sample size with a confidence level of 95% and a confidence interval of 5% was selected [39]. This culminated in 151 papers selected from the 247 papers to give a representative sample. Since the number of articles differed between years, acknowledging the upward trend in the number of articles published from 2010 to 2019 [40], a proportional stratified sampling strategy is used. In this approach, the number of papers selected for each year is based on the proportion to their size in the population [39]. Based on the calculated number of samples, papers are randomly drawn within each category. In Table 1, the number of selected papers from each year is shown.

Table 1. Number of articles selected for each year.

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Number of Articles	8	7	12	10	11	16	13	16	28	30

2.2. Data Retrieval Process and the Overall Trends in the Data

The close reading method [41] for extracting data from selected papers is employed. For this, papers are thoroughly read, focusing on statements related to the research questions listed in Section 1.3. If the required data are explicitly stated in the text, they are recorded. If not, text clues are identified. The extracted data from each paper consist of the title of the paper, name of the author/authors, digital object identifier (DOI), safety concept, year of publication, country of origin, stage of the system life cycle, industrial application domain, model type/approach, validation approach, and terminology used for validation. These are all referred to as 'variables' throughout this paper. The first three variables, which are the title of the paper, the name of the author/authors, and DOI, are recorded easily based on the bibliometric information. The other variables each have their own specific categorizations, informed by the relevant literature and emerging from the studied sample.

To define the categories, related categories available in the literature are identified and considered as a first version. Then, the data extracted from the articles in the sample are analyzed, with repeating themes found and coded for each variable. Combining the categorization from the literature with the identified themes in the dataset, the final categories are determined. In the following subsections, each variable and its associated categories, along with the reasons for selecting these categories, are provided. Furthermore, a visual overview of the information about the variables are provided to give high-level

insights in the contents of the investigated sample. The variables and the associated categories are also provided in Table A1 in Appendix A.

2.2.1. Country of Origin

To obtain a general perspective on the dataset, the publications are investigated at the country level. In the analyzed articles, 34 countries are identified, with their geographical distribution shown in Figure 2. China and the United Kingdom are leading countries in our sample papers, which is in line with the general trends in terms of the countries with most contributions to the *Safety Science* journal [40].

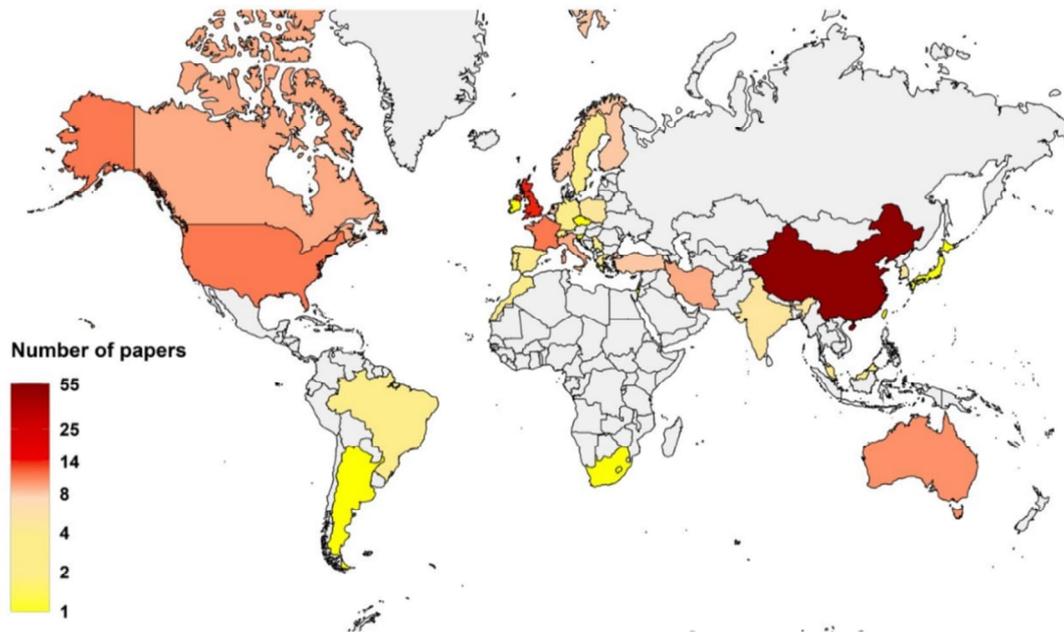


Figure 2. Geographical distribution of the articles in the data sample, period 2010–2019.

2.2.2. Stages of a System Life Cycle

The stages of a system life cycle considered in a paper can vary based on the aim of the study, the author's point of view, and the industrial application domain. For instance, the life cycle phases of offshore wind power systems include resource extraction, component manufacturing, construction, transportation, operation, maintenance, and disposal [42]. In another article [43], the stages of a life cycle are steel fabrication and raw material extraction, shipbuilding, operation, maintenance, and end of life. Since, in this study, there is a broad range of articles in different settings and industries, we adopt a more generic categorization for the stages of a system life cycle. Therefore, in the context of this article, four major categories for a system's life cycle are considered: *design*, *manufacturing/construction/development*, *operation*, and *decommissioning*. This categorization is based on a study by Kafka [44], in which design, manufacturing, operating, and decommissioning are mentioned as stages. The reason why we combine three words ('manufacturing', 'construction', and 'development') for the second stage is that different industries use different terms for the implementation of the design. For instance, in a study with a focus on the aviation industry [45], the term 'manufacturing' is used while another article concerning the construction industry used the term 'construction' [46]. Although different terms are used in these two example papers, their stages of the system life cycle refer to the implementation of the design, which is a phase after design and before operation. It should be noted here that articles focusing on the maintenance activities are grouped in the operation stage, because such work is commonly considered a major part of the operation stage [47].

As can be seen in Figure 3, in our dataset, 131 papers focused on the *operation phase*, while only one paper focused on *decommissioning* in which the author proposes a risk assessment method for the ship recycling sector [48].

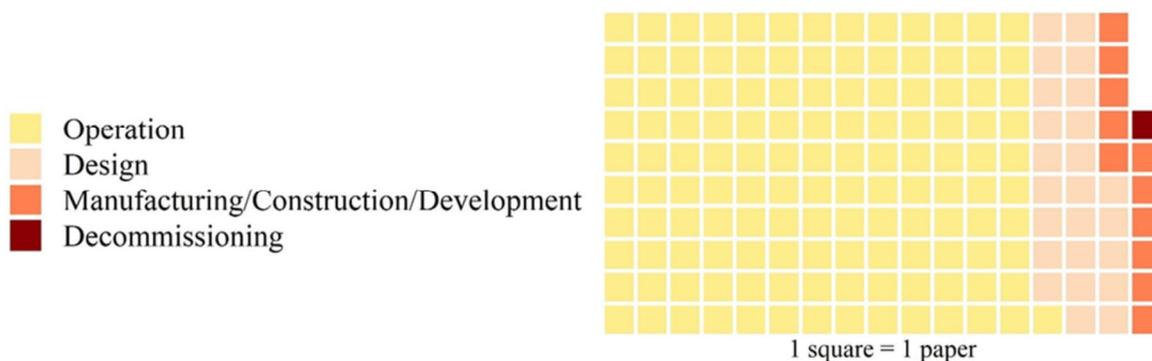


Figure 3. Waffle chart, distribution of the papers in terms of stage of the system life cycle.

2.2.3. Industrial Application Domain

The analysis of the industrial application domain shows that 44 of the papers applied an existing model or proposed a novel model for general application. Aside from this category, 12 industries are identified (Appendix A). Maritime and aviation are the first and second most prevalent industries in the sample with 28 and 16 articles, respectively. The petrochemical, robotics, and energy industries each have one paper. The distribution of articles in terms of the industrial application domain is shown in Figure 4.

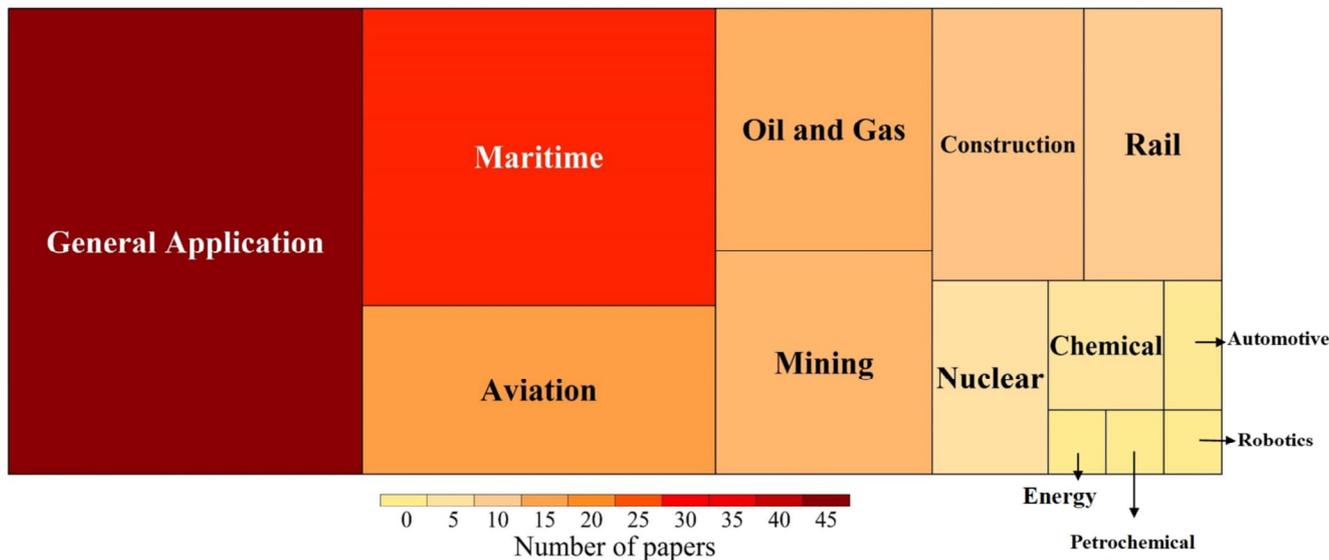


Figure 4. Distribution of papers considering industrial application domain.

2.2.4. Model Type/Approach

The categories adopted for classifying the model types are first defined based on the proposed categorization by Lim et al. [49]. Together with the categorization emerging from the articles in our sample, a slightly different categorization is adopted, in which 10 model types are defined. The categories are *hazard/accident analysis method*, *fuzzy approach*, *mathematical modeling*, *data analysis and data mining*, *Bayesian approach*, *simulation*, *statistical analysis*, *analytic hierarchy process (AHP) method*, *artificial intelligence technique*, and *other* (also mentioned in Table A1 in Appendix A).

The reason for defining the hazard/accident analysis method category is two-fold. First, the sample papers that are considered in this category are mapped with the list in the literature review paper by Wiene et al. [50], which presents a comprehensive list of accident analysis methods available in the literature. Second, although hazard analysis and accident analysis have a different focus (proactive vs. reactive), these model types are similar in nature and can be considered in one category. Hazard analysis is a way to discover potential forms of harm, their effects, and causal factors [51]. The accident analysis method is used to identify the reasons why an accident occurred and to prevent future accidents [50]. Additionally, according to a common view on safety management, the safety of a system should be ensured through both safety audits and inspections, as well as tracking and analysis of accidents [52]. Thus, we assign one category for all the hazard/accident analysis methods. It is furthermore noted that this category encompasses methods for analyzing incidents or near misses as well. This is because, according to Wiene et al. [50], the term ‘near misses’ can be used interchangeably with incidents and act as a proxy for accidents. In their words, incidents or near misses mean “an undesired and unplanned event that did not result or only minimally resulted in a loss, damage, or injury, due to favorable circumstances. Were the circumstances different, it could have developed into an accident”.

The *fuzzy approach* can deal with vagueness in the meaning of linguistic variables in safety-related models, extending the binary or classical logic [53]. The *mathematical modeling* category includes papers in which models have a set of mathematical equations, while not falling under any other categories in which mathematical operations are used, such as the *fuzzy approach* or *Bayesian approach*. The *other* category includes model types that do not belong to any of the mentioned categories.

According to Figure 5, *hazard/accident analysis method* came as the most frequently used model type, followed by the *fuzzy approach*.



Figure 5. Number of publications in each model type.

2.2.5. Validation Approach

The papers are grouped in 7 categories with respect to the validation approach. In a review by Goerlandt et al. [54], following a paper by Suokas [55], the following categories are adopted for the validation approach: *reality check* (comparing the results of a model or a part of the model with real-world data), *peer review* (examination of the model by independent technical experts), *quality assurance* (examining the process behind the analysis), and *benchmark exercise* (comparing the model results with a parallel analysis either partially (partial model scope) or completely (full model scope)). In the current study, partial and complete benchmark exercises are considered as one category. Although these four categories are specifically proposed for validation of qualitative risk assessment (QRA), the authors believe these are meaningful also for the wider safety analysis literature. Indeed,

these methods can be found in the general modeling literature as a means of validation. For instance, *reality check* is used in system dynamics modeling [56] and human reliability analysis (HRA) [57]. An example of *benchmark exercises* can be found in cognitive modeling for educational psychology in a literature review on human reliability analysis [58]. *Expert opinion* (peer review) is employed for the validation of a decision support system (DSS) [59]. Lastly, a *quality assurance technique* is used to assess the quality of a mathematical model for the consequences of major hazards as a means of validation [60]. Combined, this indicates the adequacy of the mentioned four categories in the context of model-based safety analysis.

In the present work, three more categories are added to the above-mentioned validation approaches based on our findings in the sample papers, which are *validity tests*, *statistical validation*, and *illustration*.

Validity tests is a category comprising tests applied to the formulation of a model to build an argument for its validity, without comparing the model results to external empirical data [61]. Many validity tests can be found in operation research or system dynamics modeling [62], several of which can also be employed in general modeling practices. For instance, Schwanitz developed an evaluation framework for models of global climate change based on the experience of other modeling communities [63]. One relatively well-known example of a validity test is sensitivity analysis, in which the values of model parameters are varied and the corresponding changes in the results analyzed in terms of how well those changes align with experts' expectations or prior knowledge [64]. In our current study, any paper in which the validity of a model is tested quantitatively through the application of one or more specified tests is included this category. It is worth noting that model validation cannot be made entirely objectively, and that some part of this process is subjective [65]. However, if the dominant approach to validation is applying validity tests, the paper is considered in this category. As an example, in our sample papers, Mazaheri et al. [61] performed a sensitivity analysis for validating a Bayesian belief network, following ideas from Pitchforth and Mengersen [66].

The *statistical validation* category represents statistics-based quantitative methods, where the model performance is compared to external empirical data. This category includes but is not limited to tests of means, analysis of variance or covariance, goodness of fit tests, regression and correlation analysis, spectral analysis, and confidence intervals [64]. In statistical validation of engineering and scientific models, the focus is on the process of comparing the model prediction and experimental observations [67]. This method may at first sight appear to be similar to the reality check category. However, in statistical validation, the difference between model prediction and experimental observations is quantified through statistical metrics [67] as opposed to reality check, in which the difference is considered subjectively and primarily qualitatively. As an example in our sample, Ayhan and Tokdemir defined test cases to observe the predictive performance of their model as a means of validation [68]. In another paper, a measure of goodness-of-fit of the data is applied to validate the model [69].

Illustrations are sometimes presented when proposing a new safety analysis model or approach through a case study. In general, case studies are used to analyze new or ambiguous phenomena under real-world conditions in authentic contexts [70]. These are then used to build a conclusion that is drawn from the collected evidence and observed outcomes [71]. Nevertheless, there are different types of case studies, including illustrative or exploratory case studies [72], which have different aims, such as providing a description, testing a theory, or generating a theory [73]. In our present study, the illustration category denotes articles where an example case study is presented to show how the presented model works. Compared to other validation categories, illustrative case studies do not provide much confidence that the model provides correct or useful results. Instead, these merely show that a proposed model can indeed be applied, how this is done, and what results are obtained. As an example, in our sample papers, Yan et al. [74] applied their

developed fuzzy-set based risk assessment model to a rail transit project in China to an example case study.

The distribution of the sample papers in terms of the adopted validation approach (which is also an answer to research question 2) is discussed in Section 3.2.

2.2.6. Terminology Used for Validation

There is no consensus on the words used for validation, while there is a set of terminology that has been used interchangeably in the sample. These words are *validation*, *evaluation*, *verification*, *comparison*, *effectiveness*, *usefulness*, and *trustworthiness* (Table A1 in Appendix A). This issue is not limited to our study context, and it is well known that terminological inconsistency is common in safety [75], risk [5], and validation research [76]. Many definitions of risk and risk-related concepts exist. It has been argued that this results in a chaotic situation, which continues to present problems to academic and practitioner communities [77]. The unclear terminology presents a significant obstacle to a sound understanding of what model validation is, how it works, and what it can deliver [78]. The identified terminology used for validation in the sample data is further discussed in Section 3.4.

2.3. Reliability Check of the Extracted Data

Finally, it is acknowledged that, since the data are extracted from papers, in which all the required information is not explicitly stated, there is a methodological risk of the analyst's judgments subjectivity interpreting the results. That is, the person who extracts the data inevitably makes some judgments during the data retrieval process. Therefore, to assess the reliability of the retrieved data, an inter-rater reliability experiment is performed [79]. The following steps are executed: the first author extracted the data from the 151 articles. Then, the second author extracted the data from 15 randomly selected papers, i.e., 10% of the total number of papers, and recorded the results for each variable separately. Subsequently, the agreement between the responses of the two authors for the selected 15 papers is calculated through the Cohen Kappa index, which is the most popular index [80], using R programming language. Based on the gained Kappa score (0.887), it can be concluded that there is a very high level of agreement in the judgments of categorization [81]. Due to the subjectivity of many categorical scales, achieving perfect agreement is highly uncommon [80]. It is noted that the categorization of the adopted validation approach, which is the main focus of this paper, was always the same in the results of both authors, indicating that the inter-rater reliability of the data extraction is acceptable for our current purposes.

2.4. Data Analysis Method

In this study, all the data visualizations and statistical analysis tests are carried out using R programming language.

In Section 3.3, the correlation between validation and other variables in our dataset, including the year of publication, safety concept, model type/approach, country of origin, industrial application domain, and stage of the system life cycle is tested. The year of publication is an ordinal categorical variable, while others are nominal categorical variables. A new nominal categorical variable is added to the dataset called validation, which shows whether a paper is validated or not, so it has two levels: yes or no. To investigate whether there is a statistical correlation between validation and nominal variables, their statistical dependency is studied using Fisher's exact test. This is an alternative to Pearson's chi-square test of independence when the sample size is small [80,82]. The significance of the correlation is tested by computing the *p*-values. Furthermore, a stacked bar plot is used to show their contingency tables, which include the frequency distribution of the variables [80].

A separate section (Section 3.3.2) is dedicated to the relationship between validation and the year of publication, for which a Kruskal–Wallis test is performed [83]. This test is

the non-parametric equivalent of one-way ANOVA, and it is best for cases when there is one ordinal and one nominal variable.

3. Results

In this section, the answers to research questions proposed in Section 1.3 are provided.

3.1. Percentage of the Papers in Which the Models Are Attempted to Be Validated

In this section, the answer to research question 1 is investigated. Here, the articles are divided into two subgroups: those in which the models are not validated, and those in which they are. The data analysis shows that, in only 37% of the articles, a validation of the proposed or applied models is performed, while in 63% of the articles, no model validation is presented.

In the left plot in Figure 6, the total number of papers and the number of papers in each subgroup in each year are shown through a stacked bar plot. Each bar is divided into two parts, representing the subgroups, with the number of papers in which the models are validated shown in dark gray and the number of papers in which the models are not validated shown in light gray. As mentioned in Section 2.1, and as can be seen in the figure, there is an upward trend in the number of published papers from 2010 to 2019, with a significant spike in the number of articles in 2018 and 2019 compared to previous years.

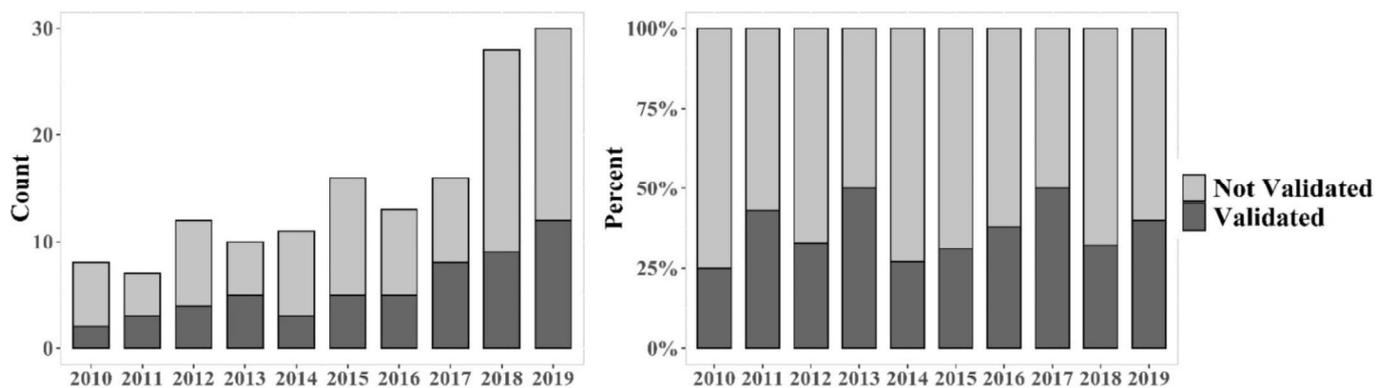


Figure 6. Count and the percentage of the papers in which the models are validated and the papers in which the models are not validated over the 10 years period.

In the right plot in Figure 6, the percentage of each subgroup is represented. The proportion of papers with validated models does not show a clear trend over the past ten years. For instance, in 2013 and 2017, about half of the authors attempted to validate their models in some way, while in 2012 and 2018, the percentage was 32%.

3.2. Approaches on Validation of Model-Based Safety Analysis

This section answers research question 2. As discussed in Section 2.2, the articles in which validation is performed are grouped into seven categories in terms of the adopted validation approach. Figure 7 shows the percentage of applied validation approaches in the sample papers as a pie chart. It is seen that 19.7% of the papers applied *benchmark exercises* to validate their models. For instance, Chen et al. compared their results with those of two other models: AHP and fuzzy weighted-average models as a means of validation [84]. Additionally, 7% of the papers applied a *reality check* approach, in which the output of the model is compared with the real-world data. The real-world data can be experimental results (e.g., [85]) or field data (e.g., [86]). In another approach, *peer review*, the model is examined by experts in that field. This approach is employed in 15.5% of the papers in which models are validated. Considering *quality assurance*, the approach examining the process behind the analysis, 2.8% of the papers applied this approach for validation.

The percentage of the other three validation approaches, which are *validity tests*, *statistical validation*, and *illustration*, are 18.3%, 12.7%, and 23.9%, respectively.

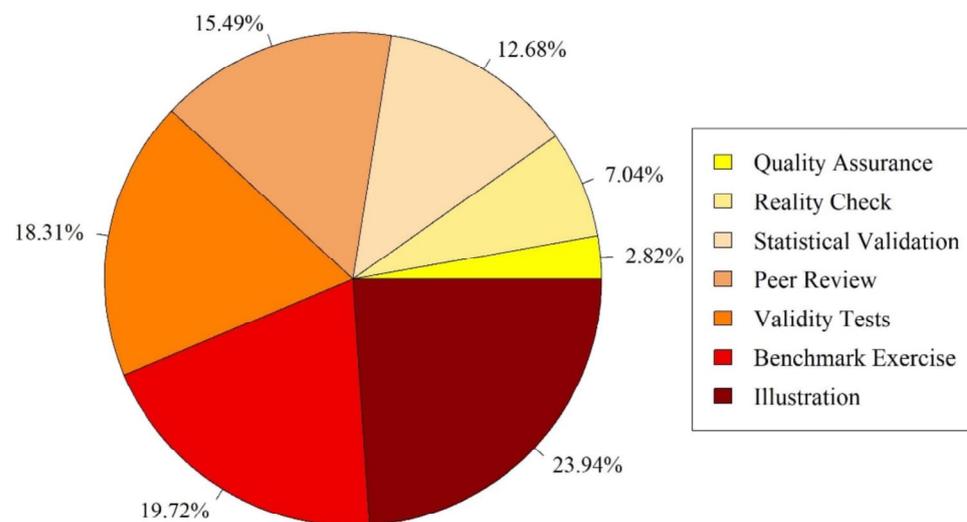


Figure 7. Pie chart-showing the distribution of articles in terms of the adopted validation approach, for cases where validation is performed.

It should be highlighted that some papers applied a mixture of these approaches to validate their models. In a paper by Mohsen and Fereshteh [87], the results of the proposed model are compared with a conventional method. Additionally, sensitivity analysis and expert opinions are used to validate the results. Thus, this work falls under the *benchmark exercise*, *validity tests*, and *peer review* categories, respectively. In another example, a three-step validation process is applied, in which the model development process is inspected (*quality assurance*), the sensitivity of results to changes in the model investigated (*validity tests*), and the model results compared with other approaches, such as FT and BN (*benchmark exercise*) [88].

In conclusion, it is seen that *benchmark exercise* and *illustration* are the most frequent validation approaches, while *quality assurance* is the least frequently adopted approach applied for validating model-based safety analyses reported in *Safety Science*.

3.3. Relationship between Validation and Other Variables

In this section, the correlation between validation and other variables, including year of publication, safety concept, model type/approach, country of origin, industrial application domain, and stage of the system life cycle, are investigated to find an answer to research questions 3, 4, 5, 6, 7, and 8. That is, it is studied whether validation has been more focused on in terms in relation to the above-mentioned variables.

3.3.1. Relationship between Validation and Safety Concept, Model Type/Approach, Country of Origin, Industrial Application Domain, and Stage of the System Life Cycle

This section answers research questions 4 to 8. As mentioned in Section 2.4, Fisher's exact test is used to test whether there is a significant statistical correlation between validation and other nominal variables, including safety concept, model type/approach, country of origin, industrial application domain, and stage of the system life cycle. The null hypothesis for each of the tests associated with the related research questions are mentioned in Table 2. The significance of the correlation is tested by computing the *p*-values. For *p*-values greater than the 0.05 significance level, we can conclude that no statistical correlation between the variables can be found, and that they are not dependent. The calculated *p*-value for each test is shown in Table 2.

Table 2. Correlation between validation and other nominal variables.

Variables	Research Question	Null Hypothesis	<i>p</i> -Value
Validation and safety concept	RQ 4	There is no correlation between validation and safety concept	0.4974
Validation and model type/approach	RQ 5	There is no correlation between validation and model type/approach	0.5437
Validation and country of origin	RQ 6	There is no correlation between validation and country of origin	0.5982
Validation and industrial application domain	RQ 7	There is no correlation between validation and industrial application domain	0.5953
Validation and stage of the system life cycle	RQ 8	There is no correlation between validation and stage of the system life cycle	0.6027

Based on the results (*p*-values), the null hypotheses cannot be rejected meaning that no correlation can be found between validation and the other investigated variables. Therefore:

- No relationship was found between how frequently validation was considered and models associated with particular safety-related concepts, including safety, risk, reliability, and resilience.
- No relationship was found between how frequently validation was considered and a specific model type/approach.
- No relationship was found between how frequently validation was considered and articles originating from a specific country.
- No relationship was found between how frequently validation was considered and a specific industry.
- No relationship was found between how frequently validation was considered and a specific stage of a system’s life cycle.

As mentioned in Section 2.4, stacked bar plots visualize contingency tables. In Figures 8–12, the stacked bar plots of validation and other variables are shown. The figures further confirm that no correlation can be found between validation and the other variables.

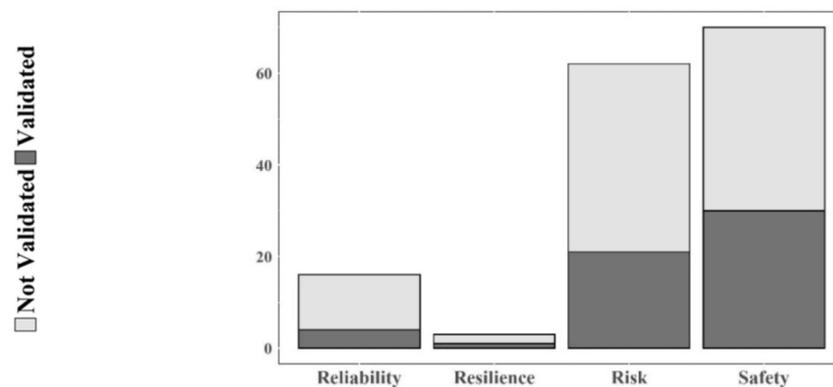


Figure 8. Stacked bar plot showing validation versus safety concept.

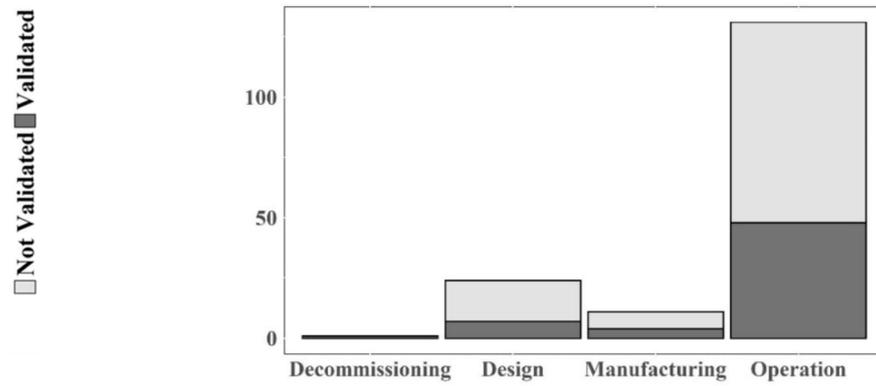


Figure 9. Stacked bar plot for validation versus stage of the system lifecycle.

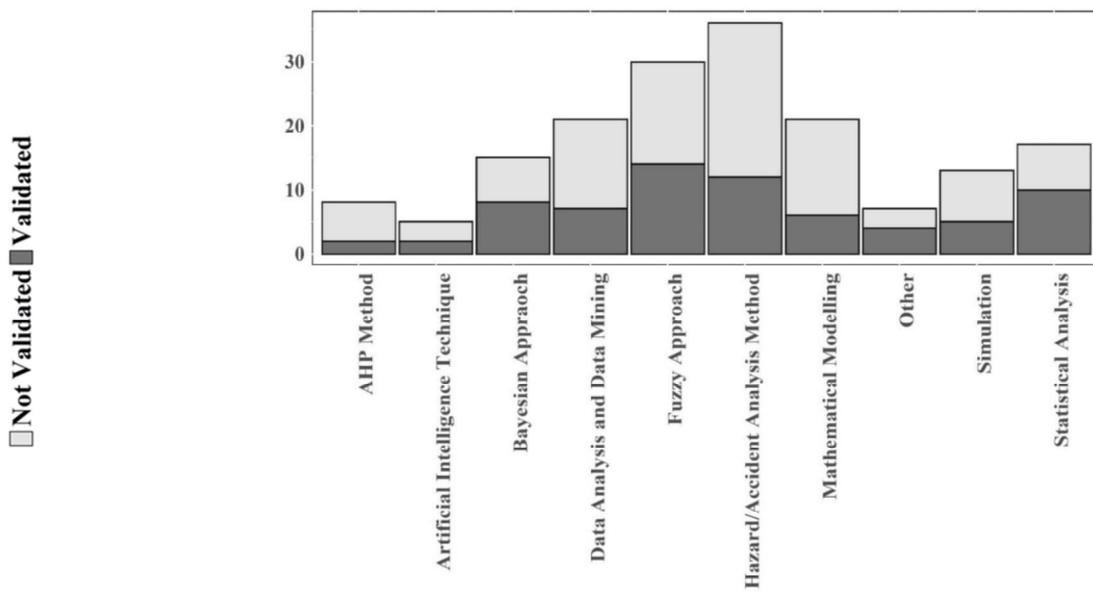


Figure 10. Stacked bar plot for validation and model type/approach.

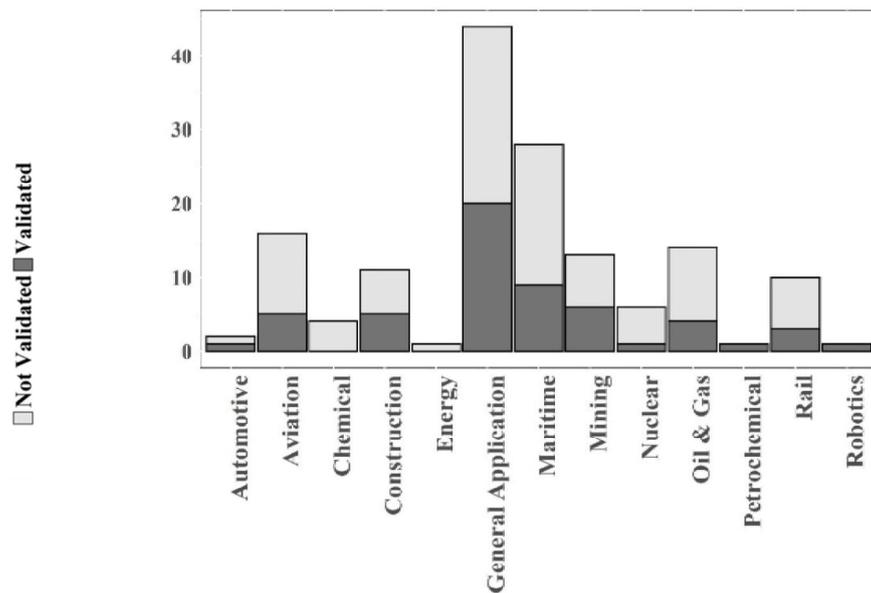


Figure 11. Stacked bar plot for validation and industry.

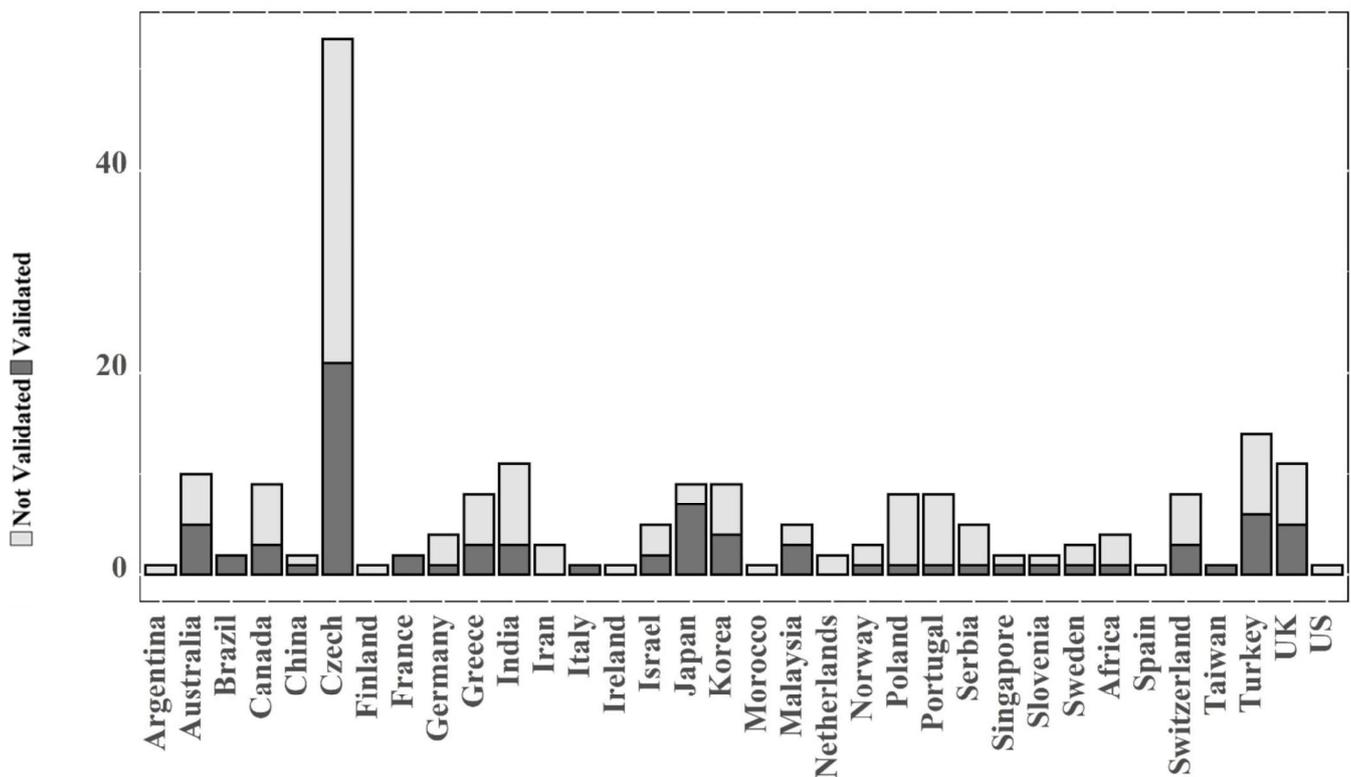


Figure 12. Stacked bar plot for validation and country.

3.3.2. Relationship between Validation and the Year of Publication

This section seeks an answer to research question 3. According to Figure 6, no trend can be observed in the relative number of papers in which the models are validated over the past 10 years in the sample. To confirm this observation, as described in Section 2.4, a Kruskal–Wallis test is performed to investigate whether there is a correlation between these two variables. The result of the test shows that there is no significant difference between the number of validated papers in different years. This confirms that no correlation can be found between the number of validated papers and the year of publication, and validation has not been more focused on in a specific year.

3.4. Terminology of Validation

This section answers research question 9. Having analyzed all the articles in the selected sample, the language of the validation in the model-based safety analysis was found to be inconsistent. The terms *validation*, *evaluation*, *effectiveness*, *verification*, *comparison*, and *usefulness* are used interchangeably in the selected papers. Furthermore, two articles in the sample apply the term *trustworthiness* [89,90]. There is also one paper [91] in which different terms, both *effectiveness* and *evaluation*, are used for validation throughout the article.

The distribution of the papers in terms of the terminology applied for validation is shown in Figure 13. The figure shows that, although a large variety in the validation-related terminology is found in the literature, *validation* is the most commonly used word in our sample.

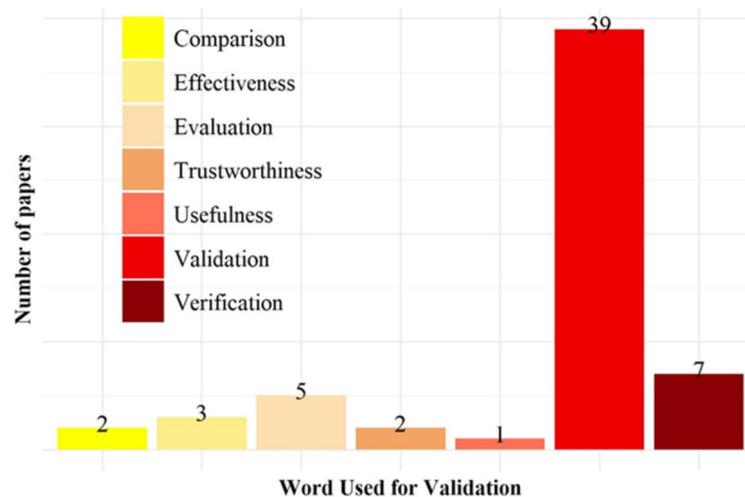


Figure 13. Distribution of papers in terms of the terminology used for validation.

4. Discussion

4.1. The Choice of Sub-Questions

In this study, the state of the practice in validation of model-based safety analysis in the academic literature is studied. To concretize this broad question, nine sub-questions are selected in Section 1.3. These sub-questions primarily aim to provide empirical evidence for arguments and claims in the academic community that validation is, in general, insufficiently considered in safety research, which contributes to a lack of evidence-based safety practices.

To the best of the authors' knowledge, no earlier work has systematically investigated validation in the context of model-based safety analysis in socio-technical systems; the focus of this work is exploratory and scoping in nature. Hence, the main purpose of the work is to better understand the extent of this problem of lack of attention to validation. Furthermore, we believe gaining insights into high-level trends and patterns in the issue of validation in relation to other aspects of model-based safety analysis can be useful to further advance this issue in the academic community and beyond.

In light of this, the percentage of articles in which the models are validated (RQ 1) and the trend over the past decade (RQ 3) are analyzed to scope the extent to which validation is considered and if temporal trends can be observed. Additionally, a better understanding of the identified validation approaches/methods (RQ 2) is useful, as it has been argued that it is not self-evident that validation exercises actually improve the model performance in relation to its intended use [92]. Closely related to this is the issue of the adopted terminology (RQ 9), which has been raised as an important foundational issue in safety science, because a different understanding of fundamental concepts can lead to different practical actions [93].

Furthermore, the relationships between validation and other aspects of model-based safety analysis are investigated to provide a broader exploratory understanding of the phenomenon (RQs 4 to 8). The underlying assumption of RQs 4 to 8 is that there are relationships between validation and *model type/approach*, *safety concept*, *stage of the system life cycle*, *industry*, and *country*, respectively. Through a series of statistical tests, these relationships are tested. From Section 3.3, it is, however, concluded that no relationship can be found between validation and a specific safety concept, model type/approach, industrial application domain, country of origin, or stage of the system life cycle. This suggests that the limited attention to validation is prevalent across the subdomains of safety research concerned with model-based safety analysis and thus in different academic communities working on different conceptual, theoretical, or methodological foundations and in various industrial application domains and countries.

If the results of some of the tests were affirmative, we could then investigate the reasons why validation is more prevalent in those areas in follow-up research to gain an understanding of why this is the case. Such investigations would require other research methods such as document analysis and interviews. Furthermore, the results could also be used as a basis for prioritizing research into the evidence of the effectiveness of validation practices, as considered in the next section.

4.2. Adequacy of the Applied Validation Approaches in the Investigated Sample

In our analysis, we made no judgment about the quality or effectiveness of the applied validation methods. We simply considered that, if the authors claimed that they have validated their models using any of the validation approaches of Figure 7, we considered those articles as indeed having validated the models.

As mentioned in Section 2.2, our sample contains seven categories of the adopted validation approaches. These are *reality check*, *peer review*, *quality assurance*, *benchmark exercise*, *validity tests*, *statistical validation*, and *illustration*. These categories are identified based on the approaches to validation as declared by the authors of the articles in our sample. Clearly, an important question is whether employing these methods improves the safety model and/or its results, i.e., whether these validation methods are, indeed, adequate. As argued by Goerlandt et al. [54], an inappropriate validation method may aggravate the problem and just add another layer of formalization by providing a false assurance of safety, through providing a seemingly adequate safety analysis, while this, in fact, is not the case [94]. Furthermore, performing validation work requires resources, such as time and money [95], so the effectiveness of such safety work should be questioned.

Although the identified validation approaches have been used in our investigated sample and other disciplines concerned with modeling, such as operations research and systems dynamics, they may not suffice to validate a model or its resulting outputs. In the wider literature on model validation, some of these methods are argued not to be adequate approaches to validation. According to Pitchforth and Mengersen [66], model validity is not simply a matter of a model's fit with a set of data but is a broader and more complex construct. The process of validating a model must go beyond statistical validation [64]. Oberkampf and Trucano [96] argue that reality checks are inadequate approaches to validation. They claim that "this inadequacy affects complex engineered systems that heavily rely on computational simulation for understanding their predicted performance, reliability, and safety". Therefore, we do not claim that the identified approaches to validation in our sample are adequate for model-based safety analyses. Indeed, we argue that there is a limited understanding of if, how, under what conditions, and to what extent the application of these validation approaches indeed improves the results. Therefore, it appears an important and fruitful avenue for future research to investigate the adequacy of these approaches.

One future research direction that may improve the practice of validation of model-based safety analysis is to develop and test a validation framework that encompasses different elements of a model in the validation process, not just a specific part of the model or its output. For instance, the model's underlying assumptions [16] and data validation [30] could be important parts of a more comprehensive model validation framework. In a study by Shirley et al. [97], full scope validation of a technique for human error rate prediction (THERP) method focuses on the internal components of the model rather than just the output of the model. Developing a validation framework for model-based safety analyses could help authors have a more thorough validation assessment, which may provide more confidence for safety practitioners in selecting and applying particular models.

Once the validation framework is developed, it should be tested to determine whether it improves the results, where aspects related to cost-effectiveness should be considered as well. We note here that "improving the results" concerns the aims and functions of a model-based safety analysis in relation to how this is intended to be used. This further suggests that a validation framework can have different functions, including but not limited to:

- Establishing confidence in a model;
- Identifying more hazards; and
- Improving the agreement of a model's output with empirical data.

Therefore, the validation framework should be tested to determine to what extent it satisfies its envisaged functions. To develop such a validation framework for model-based safety analysis (either a generic framework or one for a specific combination of model type, safety concept, and other relevant aspects), model validation frameworks developed in other scientific disciplines, such as environmental modeling or operation research, could be explored to see if and how these can be elaborated in a safety context. Nevertheless, due to the specific nature of the concepts for which models for safety analysis in socio-technical systems are built, which typically concern non-observable events or system characteristics, existing model validation approaches likely need to be modified.

4.3. Investigating the State of the Practice in Validation of Model-Based Safety Analysis among Practitioners

Based on the results of Section 3, it can be concluded that validation is not commonly performed in scientific work when proposing new model-based safety analyses or when applying them to new problems. Furthermore, acknowledging arguments for a need to strengthen the link between safety academics and safety practice [98], it is fruitful to dedicate future research to understand the state of the practice in validation of model-based safety analysis in practical safety contexts. In a study by Martins and Gorschek [99], the practitioners' perceptions on safety analysis for safety-critical systems are investigated. Their research indicates that should researchers focus not only on developing new models, but also on validation of those developed models, which could further culminate in increased trust in those models. More generally, they argue that more research should be dedicated to understanding how and why practitioners use specific approaches for eliciting, specifying, and validating safety requirements.

It would benefit both academics and practitioners to acquire qualitative evidence and empirical data regarding the validation of model-based safety analysis among practitioners. This can focus, for instance, on the merits and demerits of validation, their objectives in performing validation, the methods they use, and the challenges they face or may face in the process of validation. Their views on the function and effectiveness of the validation, i.e., whether validation indeed improves the model results, adds value for improving system safety, or if validation improves a model's credibility. This could inform the development of a framework for safety model validation. Finally, we believe that gaining more understanding of how practitioners see validation of model-based safety analysis in different industrial contexts can lead to further research directions and contribute to evidence-based safety practices.

4.4. Conceptual-Terminological Focus on Validation as a Foundational Issue

Another finding of Section 3 is that validation-related terminology in the academic literature on model-based safety analysis is not consistent. This issue of lack of terminological clarity in the safety and risk field has been raised by several authors [92,96]. Some attempts have been made to clarify the terminology of validation in other scientific domains. For instance, in an article by Finlay and Wilson [100], a list of 50 validity-related terms and their definitions in the field of decision support systems is provided. One reason why careful consideration of terminology is important is that there can be large differences in the way one conceptualizes and understands validation, which can, in turn, influence how one believes the validity of a safety model should be assessed [101,102]. When authors rely on a different understanding of the meaning of validation as a concept, this may be reflected in the terminology applied to refer to this idea. This appears plausible based on findings by Goerlandt and Montewka [102], who empirically investigated definitions of risk and the metrics that are used in the associated risk descriptions.

Amongst others, Aven has argued for the need to strengthen the foundations of safety science as a (multi-)discipline [93] by increased attention to issues such as meaning and implications of fundamental concepts underlying the discipline or its subdomains. Explicitly addressing such fundamental issues may strengthen the scientific pillars of safety science and ultimately improve safety practices. Considering the variety of implicit commitments in the approaches to validation taken in the articles in our investigated sample and the various options for what validation could do to “improve the results” of an analysis, as discussed in Section 4.2, giving explicit attention to validation as a concept could be a fruitful path for future scholarship.

4.5. Limitations of This Study and Further Future Research Directions

As this is, to the best of the authors’ knowledge, the first systematic study on the state of the practice in validation of model-based safety analysis, this research has several limitations.

First, we limited the scope of this research to a specific safety-related journal *Safety Science*. While, as discussed in Section 1.2, we believe this is a defensible choice as a basis for our exploratory and scoping analysis, we acknowledge that limiting the scope to *Safety Science* affects the results, such that they are not necessarily representative of all the literature on model-based safety analysis. For example, as mentioned in Section 2.2.1, articles originating from China and the United Kingdom occur most frequently in our sample. This follows the trend we observed in the safety science journal, in which the United Kingdom and China ranked first and third contributors, respectively [40], but this is not necessarily a good reflection of all academic work on model-based safety analysis. Likewise, the focus on the operation stage of the system lifecycle in our sample, as observed in Figure 3, should be understood from the fact that *Safety Science* was formerly published as the *Journal of Occupational Accidents* and has a legacy of having a significant focus on occupational safety [103].

Therefore, it may be fruitful to perform similar analyses for other journals where model-based safety analysis is proposed or applied with a focus on other stages of the system lifecycle [104], such as *Reliability Engineering and System Safety*, *Risk Analysis*, *Structural Safety*, and *Journal of Loss Prevention in the Process Industries*. For instance, performing this research in a journal with a focus on the *Design* phase rather than *Operation* phase of a system lifecycle could provide complementary insights into the state of practice of validation in academic work.

A second limitation is that this research is confined to articles published between 2010 to 2019. Extending this period could provide further insights into possible temporal developments.

Third, in this research, we only study the state of the practice in validation for model-based safety analysis. Validation has not been a significant research theme in safety science across problem domains [7]. Therefore, similar research in other areas in safety science, such as safety management systems or behavior-based safety, could also be beneficial.

5. Conclusions

In this paper, an analysis is performed of the relevant literature on model-based safety analysis for socio-technical systems, focusing on the state of the practice in validation of these models. Although lack of attention to validation in safety science has been raised in academia before, we aimed to provide empirical insights to understand the extent of this issue and to explore some of its characteristics. Nine research sub-questions are used to help characterize the extent, as well as possible trends and patterns in the state of the practice of model-based safety validation.

The analyses revealed that 63% of articles proposing a novel safety model or employing an existing model do not address validation in doing so. This shows that performing validation of model-based safety analysis is not a common practice in the considered sample. In this analysis, spanning a period of ten years (2010–2019), we could not find a systematically increasing or decreasing trend in the attention given to validation in the considered model-based safety analysis literature. Similarly, no correlation can be found

between validation and other investigated variables, including the safety concept, model type/approach, stage of the system life cycle, country of origin, or industrial application domain. Together, this suggests that the state of practice in validation is highly variable in the considered literature, and thus that the lack of focus on validation is prevalent across subdomains of safety science, across different communities working on different theoretical or methodological foundations, and in various industrial application domains.

In the remaining 37% of the articles, some form of validation is performed. Seven categories are identified: *benchmark exercise*, *peer review*, *reality check*, *quality assurance*, *validity tests*, *statistical validation*, and *illustration*. In our discussion, we argued that these approaches may not suffice to comprehensively validate a model, and that these different approaches in fact represent a variety of views on what function(s) validation can have in a safety analysis context. We furthermore argued that the terminological variety when referring to ‘validation’ as an activity may be based on significantly different, but often implicitly held, opinions of what validation means and what its purpose is in a context of model-based safety analysis. Therefore, we believe that increased academic attention to the meaning of validation as a concept in a safety analysis context may be a fruitful avenue for academic work. Ultimately, a focus on such foundational issues in safety science may strengthen the foundations of the discipline and could contribute to strengthening evidence-based safety practices in practical safety work.

Another way to improve the current situation could be to develop a validation framework, accounting for the function(s) of validation, the intricacies of the specific safety concepts addressed, and the model type, as well as procedural aspects of the model development and use. Once such a validation framework is developed, it would require testing to ascertain whether it improves the model’s results as intended and whether it does so in a cost-effective manner. We believe that such practice-oriented work would benefit from the earlier mentioned foundational focus on validation as a concept.

This work has several limitations, of which the scope limitation to the *Safety Science* journal is, arguably, the most significant one. This choice influences the results, so that they may not be representative of the wider literature on model-based safety analysis. In particular, the articles in our investigated sample focus primarily on the operation stage of the system lifecycle, which aligns well with the main focus in *Safety Science* but leaves the question open whether the situation is similar for other system lifecycle phases. Therefore, a future area of work would be to perform similar research for other journals with different focuses.

Overall, the authors hope that providing an understanding of the extent of the lack of attention to validation in model-based safety analysis and of some associated trends and patterns can provide some empirical grounding for earlier made arguments that validation would benefit from more academic work. We outlined some areas of future work, including a conceptual focus on the meaning and purpose of validation of model-based safety analysis, an improved understanding of validation practices in real-world organizational contexts, and practice-oriented work in developing and testing validation frameworks.

Author Contributions: Conceptualization, F.G. and R.S.; methodology, R.S. and F.G.; software, R.S.; validation, F.G. and R.S.; formal analysis, R.S. and F.G.; investigation, R.S. and F.G.; resources, R.S. and F.G.; data curation, R.S. and F.G.; writing—original draft preparation, R.S.; writing—review and editing, F.G. and R.S.; visualization, R.S. and F.G.; supervision, F.G.; project administration, R.S.; funding acquisition, F.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant number RGPIN-2019-05945 and the APC was funded by NSERC.

Institutional Review Board Statement: Not applicable.

Acknowledgments: The work in this article has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Variables and the associated categories.

Variables	Categories
Title of the paper	-
Name of the Author/Authors	-
Digital Object Identifier (DOI)	-
Safety Concept	<ul style="list-style-type: none"> • Safety • Risk • Reliability • Resilience
Year of publication	This ranges from 2010 to 2019.
Country of origin	<ul style="list-style-type: none"> <li style="width: 50%;">• Argentina <li style="width: 50%;">• Morocco <li style="width: 50%;">• Australia <li style="width: 50%;">• The Netherlands <li style="width: 50%;">• Brazil <li style="width: 50%;">• Norway <li style="width: 50%;">• Canada <li style="width: 50%;">• Poland <li style="width: 50%;">• China <li style="width: 50%;">• Portugal <li style="width: 50%;">• Czech Republic <li style="width: 50%;">• Serbia <li style="width: 50%;">• Finland <li style="width: 50%;">• Singapore <li style="width: 50%;">• France <li style="width: 50%;">• Slovenia <li style="width: 50%;">• Germany <li style="width: 50%;">• South Africa <li style="width: 50%;">• Greece <li style="width: 50%;">• South Korea <li style="width: 50%;">• India <li style="width: 50%;">• Spain <li style="width: 50%;">• Iran <li style="width: 50%;">• Sweden <li style="width: 50%;">• Ireland <li style="width: 50%;">• Switzerland <li style="width: 50%;">• Israel <li style="width: 50%;">• Taiwan <li style="width: 50%;">• Italy <li style="width: 50%;">• Turkey <li style="width: 50%;">• Japan <li style="width: 50%;">• United Kingdom <li style="width: 50%;">• Malaysia <li style="width: 50%;">• United States
Stage of the system life cycle	<ul style="list-style-type: none"> • Design Phase • Manufacturing/Construction/Development Phase • Operation Phase • Decommissioning Phase
Industrial application domain	<ul style="list-style-type: none"> <li style="width: 50%;">• Automotive Industry <li style="width: 50%;">• Mining Industry <li style="width: 50%;">• Aviation Industry <li style="width: 50%;">• Nuclear Industry <li style="width: 50%;">• Chemical Industry <li style="width: 50%;">• Oil and gas Industry <li style="width: 50%;">• Construction Industry <li style="width: 50%;">• Petrochemical Industry <li style="width: 50%;">• Energy Industry <li style="width: 50%;">• Rail Industry <li style="width: 50%;">• General Application <li style="width: 50%;">• Robotics Industry <li style="width: 50%;">• Maritime Industry
Model type/approach	<ul style="list-style-type: none"> <li style="width: 50%;">• Artificial Intelligence Technique <li style="width: 50%;">• Mathematical Modeling <li style="width: 50%;">• Accident Analysis Method <li style="width: 50%;">• Simulation <li style="width: 50%;">• AHP Approach <li style="width: 50%;">• Statistical Analysis <li style="width: 50%;">• Bayesian Approach <li style="width: 50%;">• Other <li style="width: 50%;">• Data Analysis and Data Mining <li style="width: 50%;">• Fuzzy Approach

Table A1. Cont.

Variables	Categories
Validation approach	<ul style="list-style-type: none"> • Benchmark Exercise • Reality Check • Peer Review • Quality Assurance
Word used for validation	<ul style="list-style-type: none"> • Validity Tests • Statistical Validation • Illustration • No Validation
	<ul style="list-style-type: none"> • Validation • Evaluation • Verification • Comparison
	<ul style="list-style-type: none"> • Usefulness • Trustworthiness • Effectiveness

References

- Dekker, S. *Foundations of Safety Science: A Century of Understanding Accidents and Disasters*; CRC Press LLC: Milton, UK, 2019; ISBN 978-1-351-05978-7. Available online: <http://ebookcentral.proquest.com/lib/dal/detail.action?docID=5746969> (accessed on 9 October 2021).
- Rae, A.; Nicholson, M.; Alexander, R. The State of Practice in System Safety Research Evaluation. In Proceedings of the 5th IET International Conference on System Safety, Manchester, UK, 18–20 October 2010. Available online: https://www.researchgate.net/publication/224218867_The_state_of_practice_in_system_safety_research_evaluation (accessed on 9 October 2021).
- Reiman, T.; Viitanen, K. Towards Actionable Safety Science. In *Safety Science Research*; CRC Press: Boca Raton, FL, USA, 2019; ISBN 978-1-351-19023-7.
- Guillaume, O.; Herchin, N.; Neveu, C.; Noël, P. An Industrial View on Safety Culture and Safety Models. In *Safety Cultures, Safety Models: Taking Stock and Moving Forward*; Gilbert, C., Journé, B., Laroche, H., Bieder, C., Eds.; Springer Briefs in Applied Sciences and Technology; Springer International Publishing: Cham, Switzerland, 2018; pp. 1–13, ISBN 978-3-319-95129-4.
- Aven, T. Foundational Issues in Risk Assessment and Risk Management. *Risk Anal.* **2012**, *32*, 1647–1656. [\[CrossRef\]](#)
- Goerlandt, F.; Khakzad, N.; Reniers, G. Special Issue: Risk Analysis Validation and Trust in Risk management. *Saf. Sci.* **2017**, *99*, 123–126. [\[CrossRef\]](#)
- Hale, A. Foundations of safety science: A postscript. *Saf. Sci.* **2014**, *67*, 64–69. [\[CrossRef\]](#)
- Robson, L.S.; Clarke, J.A.; Cullen, K.; Bielecky, A.; Severin, C.; Bigelow, P.L.; Irvin, E.; Culyer, A.; Mahood, Q. The effectiveness of occupational health and safety management system interventions: A systematic review. *Saf. Sci.* **2007**, *45*, 329–353. [\[CrossRef\]](#)
- Goncalves Filho, A.P.; Waterson, P. Maturity models and safety culture: A critical review. *Saf. Sci.* **2018**, *105*, 192–211. [\[CrossRef\]](#)
- Sulaman, S.M.; Beer, A.; Felderer, M.; Höst, M. Comparison of the FMEA and STPA safety analysis methods—A case study. *Softw. Qual. J.* **2017**, *27*, 349–387. [\[CrossRef\]](#)
- Goerlandt, F.; Kujala, P. On the reliability and validity of ship–ship collision risk analysis in light of different perspectives on risk. *Saf. Sci.* **2014**, *62*, 348–365. [\[CrossRef\]](#)
- Suokas, J.; Kakko, R. On the problems and future of safety and risk analysis. *J. Hazard. Mater.* **1989**, *21*, 105–124. [\[CrossRef\]](#)
- Amendola, A.; Contini, S.; Ziomas, I. Uncertainties in chemical risk assessment: Results of a European benchmark exercise. *J. Hazard. Mater.* **1992**, *29*, 347–363. [\[CrossRef\]](#)
- Laheij, G.M.H.; Ale, B.; Post, J.G. Benchmark risk analysis models used in The Netherlands. *Saf. Reliab.* **2003**, 993–999.
- Barlas, Y. Multiple tests for validation of system dynamics type of simulation models. *Eur. J. Oper. Res.* **1989**, *42*, 59–87. [\[CrossRef\]](#)
- Sargent, R.G. Verifying and validating simulation models. In Proceedings of the Winter Simulation Conference 2014, Savannah, GA, USA, 7–10 December 2014; pp. 118–131.
- Eker, S.; Rovenskaya, E.; Langan, S.; Obersteiner, M. Model validation: A bibliometric analysis of the literature. *Environ. Model. Softw.* **2019**, *117*, 43–54. [\[CrossRef\]](#)
- Le Coze, J.-C.; Pettersen, K.; Reiman, T. The foundations of safety science. *Saf. Sci.* **2014**, *67*, 1–5. [\[CrossRef\]](#)
- Casson Moreno, V.; Garbetti, A.L.; Leveneur, S.; Antonioni, G. A consequences-based approach for the selection of relevant accident scenarios in emerging technologies. *Saf. Sci.* **2019**, *112*, 142–151. [\[CrossRef\]](#)
- Li, Y.; Mosleh, A. Dynamic simulation of knowledge based reasoning of nuclear power plant operator in accident conditions: Modeling and simulation foundations. *Saf. Sci.* **2019**, *119*, 315–329. [\[CrossRef\]](#)
- Kulkarni, K.; Goerlandt, F.; Li, J.; Banda, O.V.; Kujala, P. Preventing shipping accidents: Past, present, and future of waterway risk management with Baltic Sea focus. *Saf. Sci.* **2020**, *129*, 104798. [\[CrossRef\]](#)
- Wybo, J.-L. Track circuit reliability assessment for preventing railway accidents. *Saf. Sci.* **2018**, *110*, 268–275. [\[CrossRef\]](#)
- Mikusova, M.; Zukowska, J.; Torok, A. Community Road Safety Strategies in the Context of Sustainable Mobility. In *Management Perspective for Transport Telematics*; Mikulski, J., Ed.; Springer International Publishing: Cham, Switzerland, 2018; pp. 115–128.
- Kirwan, B. Validation of human reliability assessment techniques: Part 1—Validation issues. *Saf. Sci.* **1997**, *27*, 25–41. [\[CrossRef\]](#)

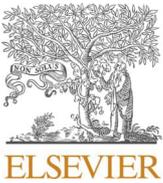
25. Hughes, B.P.; Newstead, S.; Anund, A.; Shu, C.C.; Falkmer, T. A review of models relevant to road safety. *Accid. Anal. Prev.* **2015**, *74*, 250–270. [CrossRef]
26. Wolkenhauer, O. Why model? *Front. Physiol.* **2014**, *5*, 21. [CrossRef]
27. Epstein, J.M. Why Model? *J. Artif. Soc. Social Simul.* **2008**, *11*. Available online: <https://www.jasss.org/11/4/12.html> (accessed on 9 October 2021).
28. Edmonds, B.; Le Page, C.; Bithell, M.; Chattoe-Brown, E.; Grimm, V.; Meyer, R.; Montañola-Sales, C.; Ormerod, P.; Root, H.; Squazzoni, F. Different Modelling Purposes. *J. Artif. Soc. Soc. Simul.* **2019**, *22*, 6. [CrossRef]
29. Kroes, P.; Franssen, M.; van de Poel, I.; Ottens, M. Treating socio-technical systems as engineering systems: Some conceptual problems. *Syst. Res. Behav. Sci.* **2006**, *23*, 803–814. [CrossRef]
30. Li, J.; Hale, A. Output distributions and topic maps of safety related journals. *Saf. Sci.* **2016**, *82*, 236–244. [CrossRef]
31. Reniers, G.; Anthonie, Y. A ranking of safety journals using different measurement methods. *Saf. Sci.* **2012**, *50*, 1445–1451. [CrossRef]
32. Amyotte, P.R.; Berger, S.; Edwards, D.W.; Gupta, J.P.; Hendershot, D.C.; Khan, F.I.; Mannan, M.S.; Willey, R.J. Why major accidents are still occurring. *Curr. Opin. Chem. Eng.* **2016**, *14*, 1–8. [CrossRef]
33. Gullo, L.J.; Dixon, J. *Design for Safety*; John Wiley & Sons, Incorporated: Newark, UK, 2018; ISBN 978-1-118-97431-5. Available online: <http://ebookcentral.proquest.com/lib/dal/detail.action?docID=5185085> (accessed on 9 October 2021).
34. Leveson, N.G. *Engineering a Safer World: Systems Thinking Applied to Safety*; Engineering Systems; The MIT Press: Cambridge, UK, 2012; ISBN 978-0-262-01662-9. Available online: <http://ezproxy.library.dal.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=e000xna&AN=421818&site=ehost-live> (accessed on 9 October 2021).
35. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* **2009**, *339*, b2535. [CrossRef]
36. Wee, B.V.; Banister, D. How to Write a Literature Review Paper? *Transp. Rev.* **2015**, *36*, 278–288. [CrossRef]
37. Li, J.; Goerlandt, F.; Reniers, G. An overview of scientometric mapping for the safety science community: Methods, tools, and framework. *Saf. Sci.* **2021**, *134*, 105093. [CrossRef]
38. Greenham, D. *Close Reading: The Basics*; Routledge: London, UK; Taylor & Francis Group: New York, NY, USA, 2019; ISBN 978-0-203-70997-9.
39. Bhattacharjee, A. *Social Science Research: Principles, Methods, and Practices*, 2nd ed.; Anol Bhattacharjee: Tampa, FL, USA, 2012; ISBN 978-1-4751-4612-7.
40. Merigó, J.M.; Miranda, J.; Modak, N.M.; Boustras, G.; de la Sotta, C. Forty years of Safety Science: A bibliometric overview. *Saf. Sci.* **2019**, *115*, 66–88. [CrossRef]
41. Brummett, B. *Techniques of Close Reading*; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2019; ISBN 978-1-5443-0525-7.
42. Huang, Y.-F.; Gan, X.-J.; Chiueh, P.-T. Life cycle assessment and net energy analysis of offshore wind power systems. *Renew. Energy* **2017**, *102*, 98–106. [CrossRef]
43. Dong, D.T.; Cai, W. A comparative study of life cycle assessment of a Panamax bulk carrier in consideration of lightship weight. *Ocean Eng.* **2019**, *172*, 583–598. [CrossRef]
44. Kafka, P. Probabilistic safety assessment: Quantitative process to balance design, manufacturing and operation for safety of plant structures and systems. *Nucl. Eng. Des.* **1996**, *165*, 333–350. [CrossRef]
45. Lu, Y.; Zhang, S.-G.; Tang, P.; Gong, L. STAMP-based safety control approach for flight testing of a low-cost unmanned subscale blended-wing-body demonstrator. *Saf. Sci.* **2015**, *74*, 102–113. [CrossRef]
46. Ding, L.; Zhang, L.; Wu, X.; Skibniewski, M.J.; Qunzhou, Y. Safety management in tunnel construction: Case study of Wuhan metro construction in China. *Saf. Sci.* **2014**, *62*, 8–15. [CrossRef]
47. Bağan, H.; Gerede, E. Use of a nominal group technique in the exploration of safety hazards arising from the outsourcing of aircraft maintenance. *Saf. Sci.* **2019**, *118*, 795–804. [CrossRef]
48. Garmer, K.; Sjöström, H.; Hiremath, A.M.; Tilwankar, A.K.; Kinigalakis, G.; Asolekar, S.R. Development and validation of three-step risk assessment method for ship recycling sector. *Saf. Sci.* **2015**, *76*, 175–189. [CrossRef]
49. Lim, G.J.; Cho, J.; Bora, S.; Biobaku, T.; Parsaei, H. Models and computational algorithms for maritime risk analysis: A review. *Ann. Oper. Res.* **2018**, *271*, 765–786. [CrossRef]
50. Wienen, H.; Bukhsh, F.; Vriezokolk, E.; Wieringa, R. *Accident Analysis Methods and Models—A Systematic Literature Review*; Centre for Telematics and Information Technology (CTIT): Enschede, The Netherlands, 2017.
51. Zahabi, M.; Kaber, D. A fuzzy system hazard analysis approach for human-in-the-loop systems. *Saf. Sci.* **2019**, *120*, 922–931. [CrossRef]
52. Stringfellow, M.V. *Accident Analysis and Hazard Analysis for Human and Organizational Factors*. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2010. Available online: <https://dspace.mit.edu/handle/1721.1/63224> (accessed on 9 October 2021).
53. Yazdi, M.; Nedjati, A.; Zarei, E.; Abbassi, R. A novel extension of DEMATEL approach for probabilistic safety analysis in process systems. *Saf. Sci.* **2020**, *121*, 119–136. [CrossRef]
54. Goerlandt, F.; Khakzad, N.; Reniers, G. Validity and validation of safety-related quantitative risk analysis: A review. *Saf. Sci.* **2017**, *99*, 127–139. [CrossRef]

55. Suokas, J. On the Reliability and Validity of Safety Analysis. Ph.D. Thesis, VTT Technical Research Centre of Finland, Espoo, Finland, 1985.
56. Peterson, D.; Eberlein, R. Reality check: A bridge between systems thinking and system dynamics. *Syst. Dyn. Rev.* **1994**, *10*, 159–174. [[CrossRef](#)]
57. Kirwan, B. Validation of human reliability assessment techniques: Part 2—Validation results. *Saf. Sci.* **1997**, *27*, 43–75. [[CrossRef](#)]
58. Boring, R.L.; Hendrickson, S.M.L.; Forester, J.A.; Tran, T.Q.; Lois, E. Issues in benchmarking human reliability analysis methods: A literature review. *Reliab. Eng. Syst. Saf.* **2010**, *95*, 591–605. [[CrossRef](#)]
59. Olphert, C.W.; Wilson, J.M. Validation of Decision-Aiding Spreadsheets: The Influence of Contingency Factors. *J. Oper. Res. Soc.* **2004**, *55*, 12–22. [[CrossRef](#)]
60. Vergison, E. A Quality-Assurance guide for the evaluation of mathematical models used to calculate the consequences of Major Hazards. *J. Hazard. Mater.* **1996**, *49*, 281–297. [[CrossRef](#)]
61. Mazaheri, A.; Montewka, J.; Kujala, P. Towards an evidence-based probabilistic risk model for ship-grounding accidents. *Saf. Sci.* **2016**, *86*, 195–210. [[CrossRef](#)]
62. Landry, M.; Malouin, J.-L.; Oral, M. Model validation in operations research. *Eur. J. Oper. Res.* **1983**, *14*, 207–220. [[CrossRef](#)]
63. Schwanitz, V.J. Evaluating integrated assessment models of global climate change. *Environ. Model. Softw.* **2013**, *50*, 120–131. [[CrossRef](#)]
64. Gass, S.I. Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis. *Oper. Res.* **1983**, *31*, 603–631. [[CrossRef](#)]
65. Barlas, Y. Formal aspects of model validity and validation in system dynamics. *Syst. Dyn. Rev.* **1996**, *12*, 183–210. [[CrossRef](#)]
66. Pitchforth, J.; Mengersen, K. A proposed validation framework for expert elicited Bayesian Networks. *Expert Syst. Appl.* **2013**, *40*, 162–167. [[CrossRef](#)]
67. Hills, R.; Trucano, T. *Statistical Validation of Engineering and Scientific Models: A Maximum Likelihood Based Metric*; Sandia National Labs.: Livermore, CA, USA, 2002.
68. Ayhan, B.U.; Tokdemir, O.B. Safety assessment in megaprojects using artificial intelligence. *Saf. Sci.* **2019**, *118*, 273–287. [[CrossRef](#)]
69. Valdés, R.M.; Comendador, V.F.; Sanz, L.P.; Sanz, A.R. Prediction of aircraft safety incidents using Bayesian inference and hierarchical structures. *Saf. Sci.* **2018**, *104*, 216–230. [[CrossRef](#)]
70. Phelan, S. Case study research: Design and methods. *Eval. Res. Educ.* **2011**, *24*, 221–222. [[CrossRef](#)]
71. Lee, S.-W.; Rine, D. Case Study Methodology Designed Research in Software Engineering Methodology Validation. In Proceedings of the Sixteenth International Conference on Software Engineering & Knowledge Engineering (SEKE'2004), Banff, AB, Canada, 20–24 June 2004; p. 122.
72. Hayes, R.; Kyer, B.; Weber, E. The Case Study Cookbook-Worcester Polytechnic Institute. Available online: https://zbook.org/read/9daf9_the-case-study-cookbook-worcester-polytechnic-institute.html (accessed on 9 October 2021).
73. Eisenhardt, K.M. Building Theories from Case Study Research. *Acad. Manag. Rev.* **1989**, *14*, 532–550. [[CrossRef](#)]
74. Yan, H.; Gao, C.; Elzarka, H.; Mostafa, K.; Tang, W. Risk assessment for construction of urban rail transit projects. *Saf. Sci.* **2019**, *118*, 583–594. [[CrossRef](#)]
75. Alpeev, A.S. Safety Terminology: Deficiencies and Suggestions. *At. Energy* **2019**, *126*, 339–341. [[CrossRef](#)]
76. Oberkampf, W.L.; Trucano, T.G. Verification and validation benchmarks. *Nucl. Eng. Des.* **2008**, *238*, 716–743. [[CrossRef](#)]
77. Kaplan, S. The Words of Risk Analysis. *Risk Anal.* **1997**, *17*, 407–417. [[CrossRef](#)]
78. Augusiak, J.; Van den Brink, P.J.; Grimm, V. Merging validation and evaluation of ecological models to ‘evaluation’: A review of terminology and a practical approach. *Ecol. Model.* **2014**, *280*, 117–128. [[CrossRef](#)]
79. Gwet, K.L. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement among Raters*, 4th ed.; Advances Analytics, LLC: Gaithersburg, MD, USA, 2014; ISBN 978-0-9708062-8-4.
80. Agresti, A. *An Introduction to Categorical Data Analysis*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2007; ISBN 978-0-471-22618-5.
81. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)]
82. McCrum-Gardner, E. Which is the correct statistical test to use? *Br. J. Oral Maxillofac. Surg.* **2008**, *46*, 38–41. [[CrossRef](#)]
83. Hecke, T.V. Power study of anova versus Kruskal-Wallis test. *J. Stat. Manag. Syst.* **2012**, *15*, 241–247. [[CrossRef](#)]
84. Chen, J.; Ma, L.; Wang, C.; Zhang, H.; Ha, M. Comprehensive evaluation model for coal mine safety based on uncertain random variables. *Saf. Sci.* **2014**, *68*, 146–152. [[CrossRef](#)]
85. Zhao, H.; Qian, X.; Li, J. Simulation analysis on structure safety of coal mine mobile refuge chamber under explosion load. *Saf. Sci.* **2012**, *50*, 674–678. [[CrossRef](#)]
86. Qingchun, M.; Laibin, Z. CFD simulation study on gas dispersion for risk assessment: A case study of sour gas well blowout. *Saf. Sci.* **2011**, *49*, 1289–1295. [[CrossRef](#)]
87. Mohsen, O.; Fereshteh, N. An extended VIKOR method based on entropy measure for the failure modes risk assessment—A case study of the geothermal power plant (GPP). *Saf. Sci.* **2017**, *92*, 160–172. [[CrossRef](#)]
88. Zhang, L.; Wu, S.; Zheng, W.; Fan, J. A dynamic and quantitative risk assessment method with uncertainties for offshore managed pressure drilling phases. *Saf. Sci.* **2018**, *104*, 39–54. [[CrossRef](#)]
89. Bani-Mustafa, T.; Zeng, Z.; Zio, E.; Vasseur, D. A new framework for multi-hazards risk aggregation. *Saf. Sci.* **2020**, *121*, 283–302. [[CrossRef](#)]

90. Zeng, Z.; Zio, E. A classification-based framework for trustworthiness assessment of quantitative risk analysis. *Saf. Sci.* **2017**, *99*, 215–226. [[CrossRef](#)]
91. Razani, M.; Yazdani-Chamzini, A.; Yakhchali, S.H. A novel fuzzy inference system for predicting roof fall rate in underground coal mines. *Saf. Sci.* **2013**, *55*, 26–33. [[CrossRef](#)]
92. Goerlandt, F.; Reniers, G. Prediction in a risk analysis context: Implications for selecting a risk perspective in practical applications. *Saf. Sci.* **2018**, *101*, 344–351. [[CrossRef](#)]
93. Aven, T. What is safety science? *Saf. Sci.* **2014**, *67*, 15–20. [[CrossRef](#)]
94. Rae, A.; Alexander, R.D. Probative blindness and false assurance about safety. *Saf. Sci.* **2017**, *92*, 190–204. [[CrossRef](#)]
95. Groesser, S.; Schwaninger, M. Contributions to model validation: Hierarchy, process, and cessation. *Syst. Dyn. Rev.* **2012**, *28*, 157–181. [[CrossRef](#)]
96. Oberkampf, W.L.; Trucano, T.G. Verification and validation in computational fluid dynamics. *Prog. Aerosp. Sci.* **2002**, *38*, 209–272. [[CrossRef](#)]
97. Shirley, R.B.; Smidts, C.; Li, M.; Gupta, A. Validating THERP: Assessing the scope of a full-scale validation of the Technique for Human Error Rate Prediction. *Ann. Nucl. Energy* **2015**, *77*, 194–211. [[CrossRef](#)]
98. Le Coze, J.-C. *Safety Science Research: Evolution, Challenges and New Directions*; Taylor & Francis Group: Milton, UK, 2019; ISBN 978-1-351-19022-0. Available online: <http://ebookcentral.proquest.com/lib/dal/detail.action?docID=5850127> (accessed on 9 October 2021).
99. Martins, L.E.G.; Gorschek, T. Requirements Engineering for Safety-Critical Systems: An Interview Study with Industry Practitioners. *IEEE Trans. Softw. Eng.* **2020**, *46*, 346–361. [[CrossRef](#)]
100. Finlay, P.; Wilson, J.M. Validity of Decision Support Systems: Towards a Validation Methodology. *Syst. Res. Behav. Sci.* **1997**, *14*, 169–182. [[CrossRef](#)]
101. Aven, T. The risk concept—historical and recent development trends. *Reliab. Eng. Syst. Saf.* **2012**, *99*, 33–44. [[CrossRef](#)]
102. Goerlandt, F.; Montewka, J. Maritime transportation risk analysis: Review and analysis in light of some foundational issues. *Reliab. Eng. Syst. Saf.* **2015**, *138*, 115–134. [[CrossRef](#)]
103. Goerlandt, F.; Li, J.; Reniers, G.; Boustras, G. Safety science: A bibliographic synopsis of publications in 2020. *Saf. Sci.* **2021**, *139*, 105242. [[CrossRef](#)]
104. Li, J.; Goerlandt, F.; Reniers, G. Mapping process safety: A retrospective scientometric analysis of three process safety related journals (1999–2018). *J. Loss Prev. Process Ind.* **2020**, *65*, 104141. [[CrossRef](#)]

Publication II

Sadeghi, & Goerlandt, F. (2023). Validation of system safety hazard analysis in safety-critical industries: An interview study with industry practitioners. *Safety Science*, 161. <https://doi.org/10.1016/j.ssci.2023.106084>



Validation of system safety hazard analysis in safety-critical industries: An interview study with industry practitioners

Reyhaneh Sadeghi^{*}, Floris Goerlandt

Dalhousie University, Department of Industrial Engineering, Halifax, Nova Scotia, Canada

ARTICLE INFO

Keywords:

Hazard analysis
Validation
System safety
Safety-critical industries
Practitioner's perspective

ABSTRACT

While many hazard analysis techniques exist, little empirical research has been dedicated to their use in industrial contexts, in particular concerning how practitioners validate hazard analyses. This raises questions about the accuracy, comprehensiveness, and credibility of safety analyses, and how practitioners consider this issue in relation to the overall system safety work. Acquiring qualitative evidence regarding the validation of hazard analysis among practitioners is important to support evidence-based safety practices. This paper qualitatively investigates the state of practice in hazard analysis and its validation for system safety among practitioners. Twenty semi-structured interviews were conducted with practitioners in safety-critical industries in North America. Feedback from practitioners indicates that only a limited number of hazard analysis methods are applied in industry, which are mainly based upon linear accident theory. It is also found that almost all practitioners perform some form of validation as they believe this type of safety work improves safety. Experts Reviews and benchmark exercises are the only methods reported for validating hazard analysis. In addition, practitioners highlighted several weaknesses of the current hazard analysis and hazard analysis validation practices, of which subjectivity is seen as the most important one. The authors discuss this in context of the emerging academic consensus that hazard analysis is inherently subjective, but that it can nevertheless be very useful especially when it relies on strong evidence. Also, several opportunities for organizations, regulatory bodies, and academic institutions are identified to improve the current state of the practice in both hazard analysis and hazard analysis validation.

1. Introduction

Hazard analysis is an integral part of system safety analysis. Developing a system free of hazards is an unrealistic objective of system safety (Stephans, 2004). Absolute safety is not possible since hazard sources are used within systems for desired system functions (Ericson, 2015). However, hazards should be identified, their associated risks assessed, and either eliminated or controlled to an acceptable level of risk (Vincoli, 2014). A plethora of hazard analysis techniques have been proposed and implemented across various industries, such as Fault Tree Analysis (FTA), Failure Mode and Effects Analysis (FMEA), and Systems-Theoretic Process Analysis (STPA). All of these methods should be applied rigorously, thoroughly, and systematically, in order to achieve comprehensive results (Dunjó et al., 2010).

An incomprehensive or flawed hazard analysis can result in a phenomenon that Rae and Alexander (2017) call “probative blindness.” This phenomenon means providing a false assurance that a system is safe

while, in reality, it is not. The results of a hazard analysis can be used as a basis for decision-making at different stages of a system's lifecycle. If hazard analysis is believed to be effective while it does not provide knowledge about the real problems, it can result in false assurance about the result of hazard analysis. Therefore, false confidence in the results may further lead to erroneous decisions, for instance in the system's design or operation stages.

Validation could be one way to address this concern. In other fields of study, such as operation research, validation processes are employed to deal with criticisms regarding the comprehensiveness and accuracy of an analysis (Eker et al., 2018). Although a significant amount of academic work has been published regarding hazard analysis methods, most of this focuses on proposing new analysis techniques. There is only little work published which focuses on the validation of these methods.

Some research has been performed in risk analysis validation, which is a closely related field to hazard analysis. For example, Goerlandt et al. (2017) presented a review focusing on the validation of safety-related

^{*} Corresponding author.

E-mail address: reyhaneh.sadeghi@dal.ca (R. Sadeghi).

<https://doi.org/10.1016/j.ssci.2023.106084>

Received 6 March 2022; Received in revised form 23 January 2023; Accepted 25 January 2023

Available online 3 February 2023

0925-7535/© 2023 Elsevier Ltd. All rights reserved.

quantitative risk analysis. Lathrop and Ezell (2017) have gone further by proposing a logical structure based upon the systems approach to address risk analysis validation. Also, Sadeghi and Goerlandt (2021) investigated the state of the practice in validation of model-based safety analysis in scholarly work, in which the hazard/accident analysis method is considered one of the model types/approaches.

In addition to the dearth of academic research, there is a lack of empirical work investigating the state of the practice regarding validation of hazard analysis among practitioners. Studies have investigated practitioners' views on closely related fields, such as the incident investigation (Dodshon & Hassall, 2017), the state of the art in verification and validation in Cyber-Physical Systems (Zheng et al., 2017), approaches on reliability and safety engineering for safety-critical systems (Singh & Singh, 2021), and the state of practice in verification and validation of software systems (Andersson & Runeson, 2002). However, to the best of the authors' knowledge, there appears to be no research literature that explores the state of the practice in validation of hazard analysis methods in the safety-critical industries among practitioners.

The scarcity of available empirical information on validation of hazard analysis work in safety-critical industries, along with the importance of hazard analysis as a basis for engineering design and operation management, is a key motivation to study practitioners' perspectives on hazard analysis validation. Improved knowledge about this can also contribute to diminishing the gap between academic safety science research and the actual work of safety practitioners (Reiman & Viitanen, 2019). Therefore, in-depth interviews were conducted to get the insights in the views of the practitioners in relation to the methods they use for hazard analysis as well as hazard analysis validation.

To ensure consistency and to clarify the scope and focus of the current research, the frequently appearing terms in this article are defined in Appendix A. Key definitions and concepts. As varying definitions for hazard analysis and the related terms coexists in literature (Kletz, 1999), the definitions given in Appendix A should be considered as a coherent stipulative basis for the aims of the current work, while no claims are made about their universal applicability. Note that when these terms are mentioned as part of direct quotes from interviewed practitioners, we kept the original meaning by the interviewees, to avoid misrepresenting the data.

The paper is organized as follows. The research methodology and data collection and analysis are presented in Section 2. Section 3 presents the results. Section 4 provides a discussion on the findings of the interviews, highlights avenues for future work, as well as the limitations of this research. Section 5 concludes.

2. Method and data

2.1. Scope

Although hazard analysis methods are used in many domains for many different purposes, such as security (Schmittner et al., 2014), in this research, the authors are solely concerned with its use in system safety. The scope of the study is not limited to a specific stage of a system life cycle (e.g. design); instead, the stages of a system lifecycle the interviewees have focused on when performing the hazard analysis are investigated. In addition, only safety-critical industries are targeted for this research. Saunders et al. (2013) defined safety-critical industries as those industries in which safety is paramount and where a failure or a malfunction has potentially catastrophic consequences, such as loss of life or serious injury. Frequently mentioned examples of such industries in literature are nuclear, oil and gas, chemical, aviation, rail, space and defense, maritime and automotive industries (Amberkar et al., 2001; Joubert & Feldman, 2017; Lowe et al., 2016; Lwears, 2012; Saunders et al., 2013; Singh & Singh, 2021). Therefore, the authors aimed to interview a mix of experts from these industries.

The participants in this research have industry-related experience in the field of system safety hazard analysis. The number of years of

experience is not considered as an inclusion/exclusion criterion; instead, interviewees were asked about their years of experience as part of the interviews. In addition, this research is limited to practitioners working in North America.

2.2. Participant recruitment

In this study, a semi-structured interview method is used to investigate the state of the practice in validation of system safety hazard analysis methods among practitioners, through which qualitative data is gathered. In a qualitative study, the aim is not to count opinions or interviewees but understand justifications, interpretations, and views of the participants. Therefore, sampling in this research is concerned with the richness of information (O'Reilly & Parker, 2013). The sampling method applied in this study is a combination of purposive and snowball sampling, which are two non-probability sampling techniques. In purposive sampling, the interviewees are selected because of the qualities they possess (Etikan et al., 2015). In snowball sampling, the participants are asked to recommend others they know who also meet the selection criteria (Bhattacharjee, 2012), so it works based on a referral approach. Determining the sample in this research started with purposive and continued with snowballing.

According to the criteria defined in Section 1.2, first, the term "system safety" was used to find prospective interviewees with related experience, based on which a list of prospective interviewees was prepared. A request was sent to ask them if they would like to participate in this research study. The list contained fifty-six people, of which thirty-four provided no responses even though a follow up message was sent to them. Twelve of them responded but were not interested in participating in this research. Using this initial list, only ten people responded positively and were interested in participating. In addition, four relevant industry groups in LinkedIn.com were identified. A research poster with general information about this research and the purpose of the study was posted on those groups and members were invited to participate in an interview. Two additional people accepted the invitation. Then, we relied on the initial participants to identify additional study participants (snowballing). Therefore, at the end of the interviews, the participants were asked to recommend others they know who also meet the selection criteria and ask them to send the email invitation to others with related work experience. Personal recommendations by our network were another way of snowballing. As a result of snowballing, 8 more people participated in an interview.

In terms of adequacy of the number of interviewees in qualitative research, Corbin and Strauss (2008) proposed that data gathering should be continued to reach a data saturation point where nothing new is being added to the data. It is essential that the steps taken to ensure saturation are made clear (Bowen, 2008). In this research, once the first five interviews were completed, they were transcribed and thoroughly read. Then, the preliminary categories were identified. As the interviews progressed, new ideas were identified, and categories were amended, accordingly. After performing 15 interviews, no new data was identified, and the same answers recurred. As a result, no new categories were also added to the already identified categories. However, 5 more interviews were performed to make sure that saturation of ideas had indeed occurred.

2.3. Questionnaire design

The questionnaire was structured into three parts and included 24 questions which were reviewed and approved by the authors' institutional Research Ethics Board (REB) under approval number 2021-5761. In Part I of the interviews, general information about interviewees and the companies they work in was collected. Part II gathered information about the hazard analysis methods they have been using. Part III, which is the main focus of this study, collected information specific to the validation of hazard analysis. See Appendix A for an overview of all

Table 1
Demographics of the interviewees.

Demographic information	Values and distribution (N, %)		
Region	Canada (14, 67 %)	USA (7, 33 %)	
Years of experience	≤10 (5, 25 %)	20–30 (7, 35 %)	≥30 (8, 40 %)
Highest education level	Bachelor (8, 40 %)	Master (11, 55 %)	PhD (1, 5 %)

Table 2
Breakdown of industries represented by interviewees.

Aerospace	Automotive	Aviation	Mining	Oil and Gas	Rail and Transit
6 (26 %)	2 (9 %)	4 (17 %)	1 (4 %)	3 (13 %)	7 (30 %)

interview questions.

2.4. Data collection and analysis

Participation in the interviews was voluntary, and these took place from November 2021 to January 2022 via Microsoft Teams, varying in length from 45 to 90 min. At the beginning of the interviews, the purpose of the study and a general explanation of interview questions were given to the participants. The interviewees were informed that all information obtained remains entirely anonymous, confidential, and secure. They were asked if they consent to record the interview through Microsoft Teams for subsequent qualitative analysis. Then, interview questions were discussed in detail.

Qualitative research is intended to generate knowledge grounded in human experience (Sandelowski, 2004). It is essential to analyze such data methodically to generate meaningful results. According to Braun & Clarke (2006), qualitative research is complex, and thematic analysis is a foundational method for such research through which patterns within qualitative data are identified, analyzed, and reported.

In this research, once the interviews were transcribed, they were thoroughly examined to gain an overall view of the gathered data. Then,

they are imported into NVivo, which is a software tool for analyzing qualitative data. In this software, the data was coded to identify common themes, including topics, ideas, and patterns of meaning that come up repeatedly and later categorized and analyzed.

2.5. Overview of interviewees

For this study, we sought views of system safety practitioners only. In total, 20 system safety practitioners from Canada and the USA participated in this research study offering insights in the current state of the practice in validation of system safety hazard analysis. More than half of the interviewees were from Canada while the rest were from the USA, see Table 1. In terms of the years of experience, the majority (75 %) of the interviewees are highly experienced in the field of system safety with more than 20 years of experience. Interviewees were asked about their level of education and their field of study. All interviewees have higher education mainly in an engineering field, including aerospace engineering, chemical engineering, industrial engineering, mechanical engineering, and electrical engineering.

The twenty participants represent 6 industry groups. The breakdown of participants’ industries is shown in Table 2. It should be noted that some of the interviewees were active in more than one industry. In such cases, all industries that the interviewee is actively working in are counted. For instance, one of the interviewees who is a system safety consultant is active in both the aerospace and rail and transit industries.

3. Results

The research results are presented in two sections. First, the results of part II of the interview questions are given, in which the practices and perspectives in hazard analysis are explained. This includes the adopted hazard analysis methods and the responsible unit for performing hazard analysis. Furthermore, motivations, weaknesses, and opportunities for improvement of hazard analysis are explained. Then, the results of Part III of the interview questions are given. This section focuses on the state of the practice in the validation of hazard analysis in safety-critical industries. It addresses the adopted methods for hazard analysis validation, validation definition, and motivations, challenges, weaknesses, and opportunities for improvement of hazard analysis validation.

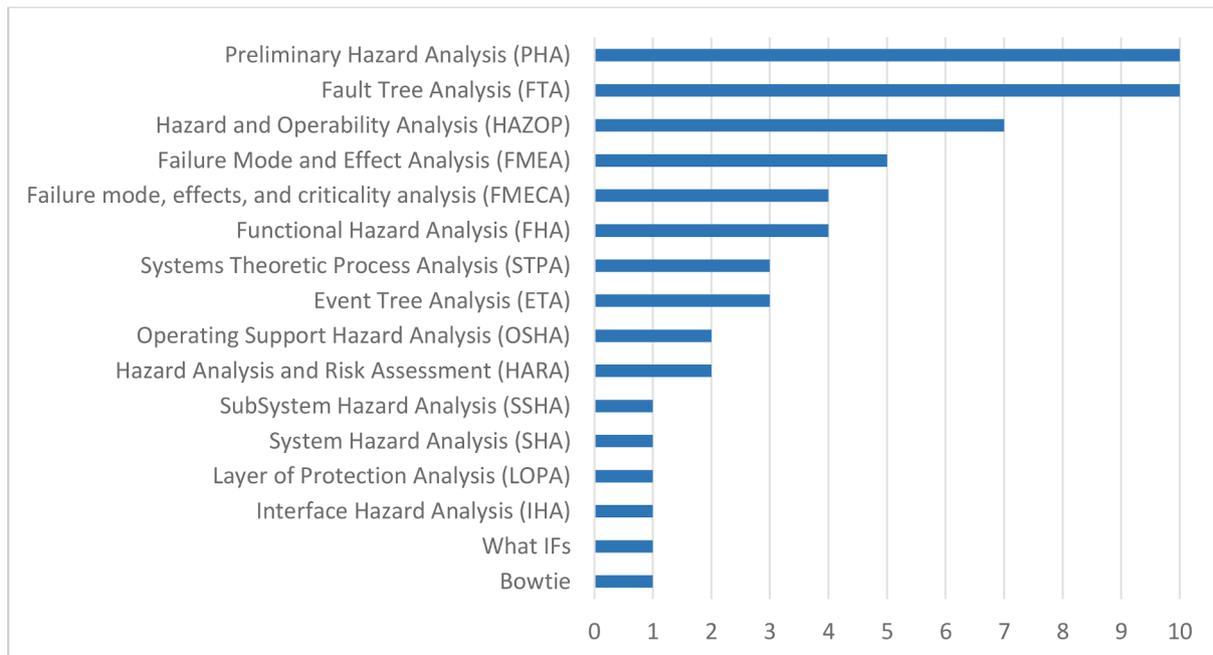


Fig. 1. Methods used for hazard analysis.

3.1. Hazard analysis: Practices and perspectives

3.1.1. Adopted hazard analysis methods

Although many different hazard analysis methods have been developed over the past forty years (Ericson, 2015), practitioners highlighted only a limited number of methods that are actually employed in their daily work. Fig. 1 shows the hazard analysis methods used and the number of times each method is mentioned by the interviewed practitioners. As can be seen, Preliminary Hazard Analysis (PHA) and Fault Tree Analysis (FTA) are the most frequently used hazard analysis techniques, followed by Hazard and Operability Analysis (HAZOP) and Failure Mode and Effect Analysis (FMEA).

Several interviewees highlighted that in almost all cases using one method would not suffice to achieve a high quality and comprehensive hazard analysis, so multiple methods are applied. PHA is commonly mentioned as the foundation for the whole analysis, and it is called by one of the practitioners “a black box analysis where you look at the whole system of interest.” Therefore, the analysis starts with applying a hazard analysis method, and then other analyses are performed to compare and extend their results. The further analyses will either feed into the already identified hazards or may detect new hazards. In addition, some practitioners perform further analyses to quantify the identified hazards.

One important point is how practitioners decide on the methods they use. Some of the interviewees work either in a consulting company or as a freelance consultant. In such cases, they first refer to the contractual agreement with the client to see if the contract dictates the standards or methods that should be used. If there are no contractor requirements, the decision on the selected methods is based upon the practitioners’ personal preference and experience. Other interviewees, who are employed in a company, highlighted personal preference and experience as the main determinants in selecting the method. The practitioners commonly agreed that the problem at hand needs to be analyzed accounting for a wide range of considerations, such as the complexity and the scale of the system, and the technologies involved in the system, based on which the practitioners decide on the method or methods that need to be used.

The preference for a given method can also be influenced by the applicable industry standards and best practices. Sometimes, the standards must be followed in order for a system to be allowed into service, and sometimes the companies prefer to follow the specific industry’s best practices to gain a competitive advantage. For instance, the HARA method has been mentioned by two practitioners (Fig. 1), both of whom are working in the automotive industry. The use of this method required by the applicable standard for the automotive industry.

Where standards allow more flexibility in the specific applied techniques, companies’ internal processes are also highlighted by some practitioners as an important factor influencing their decision. Several interviewees remarked that their decision to adopt a specific technique was based upon the processes the company laid out for them. Practitioners stated that they do not typically have a lot of flexibility in choosing the methods as the companies have invested significant financial resources to develop processes. So, when a new project starts, the processes that the companies already implemented will be used.

The stage of a system’s lifecycle is another item pointed out by practitioners as a factor that influences the choice of the method. One practitioner reported that “The approaches would be different for different stages of a system lifecycle. For instance, PHA is usually used for early concept design, and as the design progresses, HAZOP is used to look at specific scenarios.”

Practitioners were asked what stages of a system’s lifecycle they undertake hazard analysis for. A common opinion among interviewees was that this could vary for each project. Sometimes, only one stage is the focus of the project, e.g. design, and sometimes the analysis is performed for the whole system’s lifecycle. However, it is stated that hazard analysis should ideally account for the entire system’s lifecycle.

According to the interviewees, whether hazard analysis is performed throughout the system’s lifecycle very much depends on when the system safety team is involved in the analysis. If involved as soon as possible, for instance in early concept design, this allows safety engineers to influence the design. In their view, mitigating hazards when the design is finalized or when the system is already operational will lead to higher costs. Nevertheless, as reported, safety engineers do not always get involved from the early stages of a system lifecycle.

3.1.2. Units responsible for hazard analysis

All interviewees stated that in their organizations the system safety team works independently from the system engineering team. They all believe that a different group should perform hazard analysis than the group that designs a system, and the key is independence between these two groups. One practitioner explained this situation as follows: “the problem with these two teams working under the same umbrella is that they all have to have the same voice. The ideal way is that system safety reports directly to the top-level management.”

As reported by some practitioners, one factor that influences the independence between system safety and system engineering teams is the scale of a company. They pointed out that in large-sized companies these two teams tend to work independently. However, in small to medium sized companies, either someone in the system engineering team performs safety analysis, or safety analysis is outsourced to an external consultant.

Another suggested factor is the level of maturity in terms of the age of an organization. A few practitioners stated that in mature organizations, although the system engineering and safety engineering teams work in tandem, they are part of two independent entities. According to practitioners, in such organizations, the roles of each team have been clearly specified and communicated with the members of the teams. This attitude seems harder to achieve in young companies. As stated by an interviewee, “in young companies, a system safety engineer is not involved in the design and is not part of the design decisions. Once the design is complete, a system safety engineer is asked to review the design from the safety perspective, often without making any further changes to the approved design.”

3.1.3. Motivations/driving factors for performing hazard analysis

According to the interviewed safety practitioners, following items are the key drivers behind performing hazard analysis: safety, regulation, avoiding financial loss, and preventing reputational risks.

Safety is seen by several practitioners as the driving factor for performing hazard analysis. They asserted that safety is paramount, and hazard analysis is performed to support appropriate management of the system’s hazards. As mentioned by a practitioner “according to our safety culture, not only system safety engineers but all engineers in the company have to put safety in the center of everything.” An interesting quote: “because we work in a safety-critical industry, we have to do our due diligence to avoid harm whether it is asked for by an external party or not. The only difference would be the level of formalization.” This quote also supports the idea that safety is one of the main driving factors in performing hazard analysis.

Some practitioners stated that regulation drives hazard analysis. The approval of the systems by regulatory authorities, which sometime lead to certification, is contingent upon providing a body of evidence to support the analysis. One of the interviewees explained “we have to show evidence that diligent work has been done in identifying and controlling hazards so that we can get certified to sell or launch our system.” Those who highlighted this item believed that organizations are not necessarily concerned with building a safe system, but that hazard analysis is a regulatory requirement, so companies are required to perform them.

Avoiding financial loss is another motivation behind hazard analysis mentioned by some of the interviewees. It was highlighted that ideally, organizations should be concerned about safety; however, pragmatically

they are worried about money. So, they are performing different forms of hazard analysis and doing different types of tests for years to make sure that their systems run without any issue. If any incident happens, not only a considerable financial loss will happen due to accident and litigation costs, but it would impose a huge risk to their reputation. Preventing reputational risks was also mentioned by a few practitioners as a factor leading to a hazard analysis being performed.

3.1.4. Weaknesses and opportunities for improvement of hazard analysis

Two types of responses emerged from the interviews concerning the weaknesses of the hazard analyses they use. The first type of weaknesses concerns the hazard analysis methods, and the second type relates to the organizational approaches for performing and using hazard analysis.

Within the first category, four distinct aspects of the methodological weaknesses were highlighted: (i) the subjective nature of the methods, (ii) the methods being outdated, (iii) the lack of comprehensiveness, and (iv) the resource intensive nature of their application.

Several practitioners highlighted that the greatest weakness of the currently applied hazard analysis methods is that they are subjective. Two major reasons were reported for subjectivity: dependence on a facilitator and the lack of data. The former refers to the dependence of the analysis on the facilitator's experience and knowledge. It is also mentioned that having an experienced and knowledgeable facilitator who has comprehensive understanding of the mechanics of hazard analysis does not guarantee a comprehensive analysis. Subject matter experts (SMEs), who know the technical details of the system, are needed. Some practitioners also pointed out that there is a lack of clarity from regulators in terms of the required training and competency of the analysts who perform hazard analysis. It was furthermore highlighted that even when there is an experienced facilitator and a team of knowledgeable experts, it is hard to decide when an analysis should be ceased. As highlighted by one of the practitioners "you could always perform the analysis one more time, or there is always one more way you could perform the analysis." So, when to stop the analysis is a judgment call.

In terms of the latter, the lack of data, one practitioner addressed this issue as follows: "due to the lack of robust data, sometimes we do research to get probability data and sometimes the frequencies are decided subjectively. In either of these cases, however, the reliability data or failure data are not certain." This is confirmed by another practitioner who mentioned that the ideal data is captured from the system upon which an analysis is performed.

Another weakness mentioned by some of the interviewees is that the existing methods are outdated, in the sense that they fall short in identifying the hazards in today's complex systems. A commonly highlighted theme in this context is that the more systems involve automation, the more software and interfacing are added to systems. As systems become more integrated with software, the existing methods cannot effectively deal with their associated hazard. One example provided by a practitioner is that "think about a system that constantly changes its states, such as autonomous systems. How can you deal with such a system with a Fault Tree Analysis?".

Some practitioners stated that applying only one hazard analysis method does not lead to a comprehensive analysis. They asserted that two or more methods should be combined to make the comprehensive. One practitioner stated that sometimes they perform one more analysis as a cross check to see if an earlier performed hazard analysis has missed anything, aiming to be as comprehensive as possible. However, as stated by one practitioner "even using multiple methods does not guarantee that all the hazards are captured. There are always some levels of uncertainty in the analysis."

Another weakness highlighted by a few interviewees is that the current methods are resource-intensive, in the sense that they require a significant commitment of time and money. This is confirmed by a practitioner who mentioned that "there is not any type of analysis that can be done quickly to give engineers a quick idea of how to move

forward."

In terms of the weaknesses that relates to the organizational approaches to performing and using hazard analysis two items were highlighted. The first item is a lack of awareness about the importance of safety, in general, and hazard analysis, in particular, within the organizations. As stated by one practitioner "we, as system safety engineers, know the importance of hazard analysis but other engineers, such as system engineers, may not understand its importance. So, a lot of safety-related activities are done reactively because engineers do not have a safety mindset." It is also mentioned that this lack of awareness among engineers or even leaders of an organization can culminate in only focusing on the deadline of a project. As a result, such companies deal with hazard analysis as a document not as a rigorous analysis to support engineering decisions.

Occasionally, lack of awareness is compounded by poor communication and traceability among different teams within an organization. Some practitioners stated that poor communication and no traceability between the system safety and the system engineering teams result in a lack of comprehensiveness of the analysis. A lot of time can be spent on hazard analysis but without clear lines of communication and traceability, it cannot be ensured that the results are used, for example in the system design.

A next issue concerns the opportunities to improve the operational use of the current hazard analysis techniques. The practitioners' answers fall into 5 main categories: (i) developing new techniques, (ii) having experienced facilitators, (iii) issuing better guidelines, (iv) sharing information via a centralized database, and (v) educating people.

As mentioned above, one significant weakness of the currently used hazard analysis methods is that the existing methods are considered to be inadequate for today's complex systems. Therefore, practitioners expressed one main opportunity for improvement is to develop new techniques which are more suitable for dealing with these complex and software-intensive systems.

Having a good facilitator is another opportunity for practical improvement. As mentioned by a few practitioners, some characteristics of a good facilitator are being knowledgeable, experienced, and competent. One practitioner made the following comment "competency is about building the confidence over time." Facilitators have to understand the strong and weak points of each method so that they can choose the right method(s). One practitioner highlighted the importance of having younger engineers in the industry "having experienced system safety engineers is a key to success; however, it is also important to have younger engineers to bring a different perspective into the industry."

Another opportunity for improvement reported by practitioners is having a centralized safety risk database to share information with oversight by regulatory bodies. Organizations need to share their experiences about incidents and accidents, why they happened, what are the root causes. As quoted by a practitioner in rail and transit industry "in the UK, they have a very good, centralized repository of data which can be used to quantify failure rates. However, in Canada, there is no database where such information is being shared."

Issuing better guidelines for performing hazard analysis is also highlighted by some practitioners as another opportunity for improvement. One practitioner conveyed it as follows: "the standards are vague, and they can be interpreted in different ways. So, detailed guidelines on how to perform hazard analysis are required." Another practitioner expressed the need for a general written consensus on what would be the best practice under specific hazardous situations. For instance, having group conversations to reach a consensus about what would be the best practice when facing specific hazards and what would be the acceptable level of those hazard.

Another item that considered by a few practitioners as an opportunity for improvement is educating people. It is highlighted that safety courses, workshops, awareness sessions, and seminars are required, and they should not be limited to just university students. People who are working in the industry both in safety roles and non-safety roles, as well

as regulators need proper safety education.

3.2. State of the practice in the validation of hazard analysis methods

3.2.1. The definition of validation

In this subsection, the definitions of validation provided by practitioners in the context of hazard analysis are elaborated. Subsequently, their opinions regarding whether it is possible to achieve a correct result, are reported.

One clear pattern that emerged from the interviews is that an overwhelming majority of the safety practitioners defined validation as explained in the system engineering concept. Thus, they described validation as a way to ensure that a system works the way it is supposed to work. In this sense, validation is about developing test scenarios to make sure that the safety goals and functional safety requirements are met. All practitioners advancing this view referred to the V engineering model and asserted that they work in tandem with that. Far fewer practitioners provided a different definition, where validation is concerned with the comprehensiveness and correctness of a hazard analysis. One practitioner clarified that comprehensiveness and accuracy are two aspects of validation in the context of hazard analysis.

Only very few practitioners believed that it is possible to identify all the existing hazards. One stated that if the system safety engineers are informed and truly understand the newer techniques, such as STPA, every single hazard can be identified. However, most practitioners believed it is never possible to have a complete hazard analysis and there is always a margin of error. Several practitioners emphasized that hazard analysis is not an exact science. There are a lot of assumptions and judgments inherent in the process and people are a huge part of it. Safety engineering always tries to be rigorous enough to minimize uncertainty; however, considering the complex technologies, uncertainty will always be present.

Some practitioners referred to the Swiss cheese model (Reason, 1990) as a metaphor to explain that it is not possible to identify all hazards. They mentioned that different kinds of analyses are performed to add the layers of confidence, like the layers in the Swiss cheese model, but it is combinational events that are of concern. One practitioner told the following story “I have been responsible for a system that has been in operation since 1980. This system was handed over to me in 2018. I am still finding new hazards on this system. The new hazards mainly emerge as a result of interactions with other systems and the environment.”.

3.2.2. Adopted approaches and methods to validate hazard analyses

Practitioners were asked how they make sure that their assumptions, implementation steps, and results are accurate, comprehensive, and adequate for the purpose of the analysis. Only one interviewee mentioned that they do not have any form of validation. All others asserted that they validate their hazard analyses.

Almost all practitioners reported that they use independent reviews by experts as a means to validate their hazard analysis. Instead of relying just on one person and their expertise, the experiences of many types of engineers can be combined and a good analysis cannot be produced without their input. The safety engineers and system engineers have different perspectives, either side can get blinded to a certain extent. Therefore, having an independent review by a variety of experts sharpens the analysis, resulting in more accuracy and completeness.

Based upon the responses, each organization has a different review process. Several practitioners stated that they have a three-level peer review process: analysis by the hazard analyst, review by an independent person or a verifier, and approval by another independent person or an approver. The first two people usually have the same organizational level, whereas the person formally approving the analysis is someone from the management level. Ideally, the analyst, verifier, and approver are independent of the design team, and are involved in the project since the start, as stated by a practitioner. The reason given for this is that understanding the project history, such as the design decisions that were

made from the early steps of the project, is very important for having a comprehensive analysis.

In addition, during the creation of hazard analysis, workshops with SMEs are held to perform validation. As quoted by one of the practitioners “I do not wait for the analysis to be completed to send the results to stakeholders for review. I have been meeting with them constantly throughout the process.” Sometimes experts are from within the organizations, sometimes outside the organizations, and sometimes a combination of both. Workshops could take place with either a large group, or individual groups of stakeholders, users, maintainers, assemblers, or various groups associated with the client or the end-user. This depends on many factors, such as the types of hazards.

Sometimes external people are engaged to review and comment on the way hazard analysis is performed. For instance, one of the practitioners explained that once they engaged people from another industry to visit their operating units to get their impression about how well this unit performed hazard analysis compared to their industry: “we wanted to learn from them shamelessly and they brought a lot of good practices around hazard analysis to our company.” In cases when there is a client, customer, or an end-user, the result often gets submitted to them and then gets reviewed by their SMEs, as well. Another example by a practitioner was that once they complete their analysis, the results are sent to their client. Then, the client integrates this analysis into their own system-level analysis. So, the experienced engineers on the client’s side can also review the results and consider whether it makes sense or not. As one interviewee puts it: “there has to be consistency as you go up the hierarchical levels in the development.”.

Through these review sessions, practitioners not only aim to make sure that their analysis is comprehensive but also, they try to make their results credible. They focus on ensuring that the assumptions, uncertainties, execution steps, and results are clear, and are systematically communicated with the stakeholders, who were not involved in a hazard analysis. Therefore, everything is documented as they go through the analysis. The form of communication is at the discretion of the external parties, such as a client. They either get involved in the analysis during the whole process, or the results are presented to them once the analysis is completed, or the documents are submitted for review.

Another form of validation, reported by several practitioners, is the benchmark exercise, which refers to the comparison between the results of an analysis with parallel analysis. The benchmark exercise is used as a cross-check to see if anything is missed, and it provides confidence that things have been comprehensively thought of. Two types of benchmark exercises were reported: comparing the result of one hazard analysis with parallel analysis (or analyses), and industry reviews. The former was more common among practitioners. Sometimes they use more than one method, sometimes multiple methods, for an analysis because one method would not culminate in a comprehensive analysis. The same is confirmed by another practitioner who stated that they generally use a top-down and a bottom-up analysis and usually try to use a third method if the timeline of the project allows them to do so.

One of the practitioners working in a large company with different operating units reported that conducting two or more analyses in parallel for the same system is expensive. However, they compare the results of a hazard analysis of one operating unit to those of another unit as a means of validation. Two interviewees reported that they perform industry reviews to find a similar system in other companies with which to compare their hazard analysis. For instance, the following example was provided by a practitioner: “think of two tanks, they are going to have similar hazards. So, when the hazard analysis is performed for a new tank, I can just look at the results of another old tank and compare their results. For the most part, the results of the hazard analysis are going to be the same, except for those hazards related to the new technology.”.

When practitioners were asked about the extent to which validation has been integrated into hazard analysis, they mentioned that any form of validation of hazard analysis is not mandated by the applicable

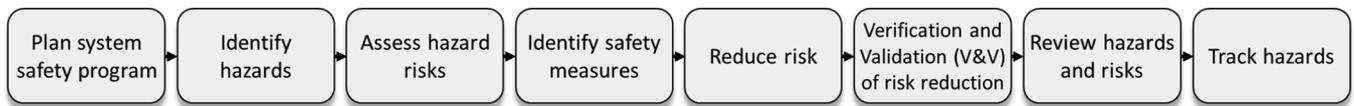


Fig. A1. System Safety Process adapted from Ericson (2015).



Fig. A2. Hazard-Accident/mishap relationship adapted from Ericson (2011).

standards for their industries and projects. They also mentioned that, to their knowledge, validation is not mentioned in any of the hazard analysis books or guidelines. However, in most cases, it is part of the organizations' internal processes. One of the practitioners made the following comment: "our peer-review process, the producer, checker, approver, is the process that is required by our organization." Sometimes, the validation approaches vary for each project that they work on. So, one thing that some of the practitioners do is that they create a program plan in which any form of validation is planned out at the early steps of the project. This helps them to clarify different steps of the analysis, including validation, with stakeholders.

3.2.3. Motivations, driving factors, key challenges, and barriers for validating hazard analysis

In the opinion of several practitioners, the motivations for validating hazard analysis mainly stem from the organization's internal policy. They asserted that the decision to perform validation or not depends on the company they work in. One practitioner stated that their management would like to do as little work as possible. Another practitioner mentioned that if there is a real issue and the leadership is sensing that, they do not take the risk of putting the system out with the issues it has. There is an effort to make safety a priority.

A few interviewees perform validation to ensure that their analysis is comprehensive, and everything is covered. As expressed by one practitioner "I would prefer to validate at least for my peace of mind to make sure that the analysis is complete." So, it also depends on the practitioners' experience to judge the level of validation needed.

Another driving factor reported by a practitioner is the level of novelty of the system. If a system is well-known and has been used extensively before, the level of validation would be lower because everyone is comfortable with the previous level of hazard analysis. When there is a new system or a new feature in an existing system, the level of

Table A1
Categories identified in the interviews.

Observation	Categories
Hazard analysis	<p>Adopted methods</p> <ul style="list-style-type: none"> ■ Preliminary Hazard Analysis (PHA) ■ Fault Tree Analysis (FTA) ■ Hazard and Operability Analysis (HAZOP) ■ Failure Mode and Effect Analysis (FMEA) ■ Failure mode, effects, and criticality analysis (FMECA) ■ Functional Hazard Analysis (FHA) ■ Event Tree Analysis (ETA) ■ Systems Theoretic Process Analysis (STPA) <p>Motivations and driving factors</p> <ul style="list-style-type: none"> ■ Safety ■ Regulation ■ Avoiding financial loss ■ Preventing reputational risks <p>Weaknesses</p> <p>Weaknesses concerns the hazard analysis methods</p> <ul style="list-style-type: none"> ■ The subjective nature of the methods ■ The methods being outdated ■ A lack of comprehensiveness ■ The resource intensive nature of their application <p>Weaknesses relates to the organizational approaches to performing and using hazard analysis.</p> <ul style="list-style-type: none"> ■ A lack of awareness about the importance of safety and hazard analysis ■ Poor communication and traceability among different teams within an organization <p>Opportunities for improvement</p> <ul style="list-style-type: none"> ■ Developing new techniques ■ Having experienced facilitators ■ Issuing better guidelines ■ Sharing information via a centralized database ■ Educating people
Validation of hazard analysis	<p>Adopted approaches</p> <ul style="list-style-type: none"> ■ Reviews by experts ■ Benchmark exercise <p>Motivations and driving factors</p> <ul style="list-style-type: none"> ■ The organization's internal policy ■ Increasing the comprehensiveness of the analysis ■ The level of novelty of the system <p>Key challenges and barriers</p> <ul style="list-style-type: none"> ■ Convincing stakeholders of the need to validate ■ Lack of competency ■ Schedule pressure ■ Lack of clear guidance ■ Budget limitations <p>Weaknesses</p> <ul style="list-style-type: none"> ■ Subjective ■ Resource-intensive <p>Opportunities for improvement</p> <ul style="list-style-type: none"> ■ Having formal processes on how to do validation ■ Increasing awareness about the value of validation ■ Having the top management support

validation would increase. As one practitioner stated in this context, there are no firm rules for when a system is considered sufficiently novel to change the approach to validation: “it is not written in stone; it is a judgment call.”

The most significant challenges/barriers to performing validation identified are grouped into five categories: (i) convincing stakeholders of the need to validate, (ii) lack of competency, (iii) schedule pressure, (iv) lack of clear guidance, and (v) budget limitations.

Several practitioners highlighted that they must convince stakeholders about the importance of validation. One interviewee explained this as follows “Sometimes it is difficult getting the stakeholders’ involvement since they do not understand why validation should be performed. The worst scenario is when validation is done, and nothing comes out of it. In such cases, they say this was a waste of time.” It was also reported that even when stakeholders grant access for performing validation, they do not realize that a certain level of experience in system safety is required for validation to be done correctly. Stakeholders often seem to think that any design engineer can validate the hazard analysis. However, practitioners highlight that it takes a certain type of experience to be able to perform a hazard analysis validation.

Lack of competency was another important barrier element to validation. Several interviewees reported that the system safety field is experiencing a worldwide shortage of competent resources. Although many young people are coming into the industry, there are also many experienced professionals retiring. This results in the loss of a lot of tacit knowledge, which is the wisdom, experience, insight, and competency of the experts. Effective transfer of tacit knowledge generally requires extensive time and regular interaction. One practitioner stressed that some companies hire new graduate students and assign them a project to start doing hazard analysis, while not sufficient time is taken to train novice engineers. Another practitioner explained that this lack of competency is not just limited to practitioners. Even government agencies lack the technical expertise necessary to guide the validation of increasingly complex, safety-critical industries.

Schedule pressures and budget limitations were also mentioned by several interviewees as factors hindering practitioners from performing validation. This becomes an especially significant issue if a change must be made to a system in a middle of a program. One of the practitioners highlighted that due to these challenges, a trade-off needs to be made regarding how many more hazards are expected to be obtained through performing validation, or how much the analysis may be improved. As quoted: “if I spend six more weeks on a hazard analysis to validate it, and as a result of these extra analyses only two more hazards are identified, which are not significant, it is not worth to spend time and money.”

A broad issue faced by several practitioners is the lack of clear guidance for how to validate hazard analyses. One of the practitioners stressed that even when validation is considered as part of the process, i. e. that it is an explicit task included the project schedule and budget, there are few details or guidelines on how validation is to be performed.

Only one practitioner stated that they did not face any challenge. He assumed the reason for this was that they have internal processes which are formalized and communicated with all relevant actors in the organization. He however indicated that this may not be the case in small and medium sized firms.

3.2.4. Weaknesses and opportunities for improvement of hazard analysis validation

The frequently mentioned weaknesses of the validation methods are that they are subjective and resource intensive.

As mentioned in Section 3.1.4, the greatest weakness of the hazard analysis methods is their subjectivity. This is also reported as the greatest weakness of the hazard analysis validation methods that practitioners are currently using. Practitioners explained that the people performing validation are a critical aspect of the work, and that the whole analysis relies upon them. So, for many interviewees, it all comes down to the skill, knowledge, and experience of the practitioner. If

people with the required knowledge or experience are not included in the work, they may not ask the right questions, or have a correct set of assumptions.

Another practitioner believed that validation in the form of structured brainstorming is not a real validation. He asserted that: “performing hazard analysis is fundamentally a recording of peoples’ opinions on various things that can go wrong. Validation is about reviewing those opinions to make sure that what have been said during hazard analysis is still valid from the perspective of the people who first expressed them.”

Another highlighted weakness is that validation of hazard analysis is an expensive and time-consuming process. One practitioner mentioned that “the schedule and budget are two competing factors up to the quality of a good, validated system.” As mentioned in Section 3.2.3, this situation is exacerbated if stakeholders do not understand the value of validation and consider it to be a waste of time, effort, and money. It is highlighted that there is always that push from companies to save money and spend less time on a project. Especially if then nothing of significance is indeed found in the validation process, they consider it as spending money and time to say that the primary analysis is fine.

The interviewed practitioners proposed a few opportunities for improvement of the current state of practice in validation of hazard analysis. These include: (i) having formal processes on how to do validation, (ii) increasing awareness about the value of validation, and (iii) having the top management support.

A common opinion among the interviewed practitioners is that a formal way of how to validate hazard analysis is required. They suggested having standards proposed by regulatory authorities, and standard processes, such as written work instructions. Many interviewees also highlighted frameworks proposed by academia could add significant value to improve the current situation in safety practice. One practitioner explained that they often follow an arbitrary peer-review process, and that people involved in their peer-review sessions do not necessarily have required experience or knowledge. So, it is suggested that a standard process which also clearly specifies the expertise required for the people involved in peer review would be beneficial.

A few practitioners highlighted that it is essential to enhance the visibility of validation and to educate the relevant actors in an organization about the importance of validation. One of the interviewees believed that enforcement through proceduralization does not bring in the desired results; however, education does. The different forms of education mentioned are workshops, discussion groups, and training courses. The practitioners believed that the primary responsibility for increasing this awareness about validation lays with industry, but many believed that also academia can play a significant role in this. One of the practitioners stressed that there should be more research done in terms of the validation concept and the formal processes on how to do validation: “Validation is an important topic, and there needs to be a lot more light brought to this field of study.”

Increasing awareness could also lead to top management support, which was mentioned by a few practitioners as an important opportunity for improvement. A leaders’ approach can instill commitment or indifference. One interviewee expressed it as follows: “the primary goal of a company is to make money. However, having a leadership who wants to do it right would be paramount. If leaders do not back you up, you could have the best safety team and you just crank out analysis.” If leadership considers safety analysis a priority, it will be written in the organization’s policies and communicated to the whole organization to be formalized.

3.2.5. Practitioners’ perspective on validation as a value driver

All interviewees consider validation a significant value driver and necessary to have a comprehensive hazard analysis. Validation helps practitioner to make sure that everything is covered, to see the gaps in the analysis, and to identify a mistake or a disagreement on assumptions. One practitioner stressed the importance of benchmark exercise as

follows: “it is a way to spot things that we have missed or have not explained thoroughly and clearly.” Another practitioner stated that through the peer review process different perspectives can be brought into the analysis: “we, safety people, have our perspective but when we present hazard analysis to a working group that has various areas of expertise, we get a lot of insight that we would not normally get. They can make those type of judgment calls that we cannot.” One example referred to system engineers who can deal with the system on a far more detailed level than safety engineers. So, their perspective can add considerable value to the analysis of safety engineers.

As mentioned in Section 3.2.4, validation is resource-intensive. This could also jeopardize the project’s budget and timing which might result in the whole program to be delayed. Practitioners, however, believed that companies will find it beneficial in the long run. Even if validation activities do not find anything, it is still valuable since it provides confidence in the analysis. One of the practitioners explained that validating the hazard analysis is not about having a green or red light. It is about creating layers of confidence to ensure that everything that could be captured, is captured.

4. Discussion

4.1. Reflection on hazard analysis

As reported by industry practitioners, most hazard analyses are performed using traditional methods, such as FMEA, FTA, and ETA. These techniques are based upon linear accident causality models, which are not well suited for incorporating complex and non-linear relationships between different elements of a system (Qureshi, 2008). Various authors have highlighted that these techniques have serious limitations in the analysis of modern complex systems (Dallat et al., 2019; Leveson, 2017; Qureshi, 2008). Leveson (2017) explained that since these traditional methods have been developed, the systems have witnessed dramatic changes, such as increased complexity; therefore, new methods are needed. This has been raised by some of the interviewed practitioners, as well (Section 3.1.4).

In response to the limitations of traditional methods, systems models have been proposed. These models consider the system a whole entity (Hollnagel & Goteman, 2004) and describe accidents not as simple chains of directly related physical and functional component failures, but rather as occurring through complex non-linear interactions among various factors (Baybutt, 2021). For instance, the STPA method, which is based on the systems theoretic accident model and processes (STAMP), extends the view of causality by including failures as a result of non-linear interaction between components in addition to linear interaction and component failures (Leveson, 2017). The STPA method can be found in Fig. 1 (Section 3.1.1), though, reported by only a few practitioners from the USA. No interviewed practitioner in Canada reported employing this method for hazard analysis.

A fundamental issue in the context of hazard analysis is the level of completeness of the analysis in terms of the identification of possible hazards, regardless of the theoretical foundation of the method (Baybutt, 2021; Dekker, 2019). There may exist some complicated hazards that are even beyond the capability of any current hazard analysis method to be identified (Baybutt, 2021). This supports the finding of the interviews as practitioners highlighted that the completeness and comprehensiveness of hazard analysis is under question (Section 3.1.4). To tackle this issue, practitioners often conduct multiple analyses. The idea that various techniques need to be applied to improve the results, has been empirically confirmed already a long time ago (Suokas, 1985; Harms-Ringdahl, 2001). However, as stated in interviews even using multiple methods does not guarantee that all relevant hazards are captured (Section 3.1.4).

Several practitioners highlighted the greatest weakness of the currently applied methods is that they are subjective (see Section 3.1.4). The issue of subjectivity has been discussed by many researchers (Aven

& Renn, 2009; Rosa, 1998; Solberg & Njå, 2012), where hazard analysis is believed to be inherently subjective. This is supported by different views on the ontological status of risk, associated with realism and constructivism.

The concept of risk in realism is based upon the idea that a certain state of the world can objectively be defined as risk, whereas since these states are not predetermined, they are uncertain (Rosa, 1998). In Rosa’s view, despite this ontological foundation, risk will move from an objective state (an ontological reality) to a subjective state (social construct) based on an assessor’s ability to identify, measure, and understand that state. Therefore, risks not only are shaped by objective states, but also by social factors. In constructivist views, risk does not exist per se, with the core concept of risk directly associated with the assessor’s knowledge about the situation and the ability to imagine a possible future state of affairs, which are inherently defined subjectively (Solberg & Njå, 2012). In their view, risk is not actually thought of in isolation, but rather it is connected to specified activities.

Although the dependence of the analysis on the facilitator’s experience and knowledge is mentioned by a few practitioners as one of the reasons of subjectivity (Section 3.1.4), it is not necessarily a weakness as their decisions can be supported by evidence. A risk analysis is a report on the uncertainties expressed by analysts; inherently subjective but rooted in evidence, which can be strong and compelling or weak (Aven & Guikema, 2011). This is what Kaplan (1997) called “evidence-based” risk assessment and decision-making. Based upon this idea, in addition to the analysts experience and knowledge, a “consensus body of evidence” is required to make a decision.

In Section 3.1.4, practitioners pointed out that the lack of clarity from regulators regarding the required training and competency of the analysts has worsened the issue of dependence on facilitators. This is supported by Provan et al. (2017), who state that the required knowledge and skills of safety professionals have not been specified. Usually, the selection of the analysis team is rather arbitrary. Thus, some criteria, e.g. who should be involved in the analysis, need to be in place for putting a team together. This lack of clear guidance is not just limited to this issue. Practitioners felt the need for detailed guidelines on how to perform a hazard analysis (Section 3.1.4). Therefore, issuing better and more detailed guidelines could alleviate some of the problems that practitioners face.

4.2. Reflection on the validation of hazard analysis

Practitioners defined validation either as (i) ensuring that a system works the way it is supposed to work and (ii) evaluating the comprehensiveness and correctness of a hazard analysis in (Section 3.2.1). The first definition is similar to how Engel (2010) defined validation, which is “evaluating a system or component during or at the end of the development process, to determine whether it satisfies specified requirements.” According to Engel, testing is a subset of validation, and the focus is on the safety requirements that are derived from the hazard analysis. All practitioners providing this definition referred to the V engineering model. According to this model, validation happens towards the end of the product development process, and the point of validation activities is to make sure the whole system, when integrated, works fine according to the safety requirements.

What the current research is concerned with is the validation of hazard analysis per se, including the assumptions, execution steps, and the results of the analysis. As defined in Section 1, validation means the process of ensuring that the hazard analysis is accurate, comprehensive, and credible. Thus, the second definition aligns more closely with how the authors defined validation at the outset of this research.

Independent reviews by experts and benchmark exercise are the only hazard analysis validation methods reported by interviewed practitioners (Section 3.2.2). These two methods were also reported as being used for the purpose of model-based safety analysis validation in academic work (Sadeghi & Goerlandt, 2021). While these methods are

useful and add value to the analysis, they are not without limitations. The quality of such validation practices, as currently applied, comes down to the skill, knowledge, and experience of the individual practitioner (Section 3.2.4). The lack of clear guidance on who should be involved in the validation process and how to actually perform validation exacerbate these challenges. According to (Balci et al., 2002), SMEs often provide unstructured, unclear, and deficient feedback due to the complexity of the systems as well as the copious amount of information that needs to be grasped. They also explained how eliciting, representing, and integrating SMEs' knowledge is a key challenge for performing validation activities.

Having formal processes on how to perform validation is highlighted as one of the opportunities for improving the current state of the practice in validation of hazard analysis (Section 3.2.4.) As mentioned in the introduction (Section 1), validation frameworks have been developed in other closely related fields, such as risk analysis (Lathrop & Ezell, 2017). A hazard analysis validation framework may alleviate some of the current challenges and help practitioners to have a structured validation assessment. Therefore, to confirm this assumption, a validation framework needs to be developed, and tested.

Validation adds another layer of analysis which requires the stakeholders' support as it demands extra budget, more time, and competent people to join the analysis. These raise challenges for practitioners to perform validation (Section 3.2.3.) The challenges are closely linked to the issues organizations deal with to manage safety. Based upon Rasmussen's model of practical drift (Rasmussen, 1997), there are three types of constraints on the system's operation which are economic, workload, and safety. More pressure on economic and workload push the system's operation closer to the safety boundary. Therefore, safety demands constant negotiation and compromises with other dimensions (Amalberti, 2013). One of the practitioners highlighted that due to schedule pressures and budget limitations, a trade-off needs to be made regards performing validation (Section 3.2.3). Often, such trade-offs are an integral part of organizations to remain operational and competitive. This does not mean that the trade-offs have no ramifications, as this could be a driver which can result in "drift into failure" (Dekker, 2011). Despite all challenges, embedding validation into the hazard analysis could be a way to identify errors in the analysis which could have subsequent effects on developing safer systems. However, this assumption needs to be tested and supported by evidence.

4.3. Safety education and awareness

Lack of awareness and the need for educating people has been highlighted often throughout the interviews, whether related to the general safety concept or more specific concepts of hazard analysis and validation. It was highlighted that safety courses, workshops, awareness sessions, and seminars are required, and they should not be limited to just university students. That is, people who are working in the industry both in safety roles and non-safety roles, as well as regulators need proper safety education.

Stephans (2004), raised 8 problem areas that need to be addressed to have an effective system safety program, one of which is education and training. He suggests a proper system safety education is adding system safety courses to the engineering programs core curriculum, so all engineers have at least a basic knowledge of the system safety objectives, concepts, and methods. Wassenhove et al. (2022a) highlighted that proper safety education should go beyond the basics and foundations of safety science, and that knowing the professional realities of the safety profession is necessary to create good safety education courses. In addition, continuous professional education is highlighted by Mkpat et al. (2018) as part of a proper safety education as it promotes continuous professional advancement.

The need for better training and awareness has been raised already a long time ago by researchers as an improvement opportunity (e.g. Kletz, 2001). The result of this work highlights that it is still an unresolved

issue based on practitioners' assertions. In this regard, the interview findings conflict with the commitment to safety as a fundamental concern for professional engineers in Canada and the USA (Andrews et al., 2019). This implies that there is room to improve the general understanding of the need for safety, hazard analysis, and validation in the engineering profession.

Having raised this as an opportunity for improvement, the authors nevertheless highlight that there is still lack of ample evidence to support the idea that safety awareness and education would have an actual positive safety effect. Although there has been some scholarly work on safety education, e.g., (Mkpat et al., 2018), more work needs to be done in this area to investigate the actual effect of safety awareness and education on safety. In addition, in a recent study by Wassenhove et al. (2022b), it is highlighted that there is a gap between the academic education and the practical reality of the safety profession, which raises questions about the value of such education.

One way to improve this would be the better dissemination of the ideas, experiences, and research findings among practitioners as well as safety scientists. Dissemination between practitioners can help them to learn from each other. This may prevent practitioners from repeating similar mistakes in their analysis. Learning from others' experiences is a great educational tool (Balci et al., 2002). This could happen in a form of a conference presentation, publication, or as suggested by a practitioner, knowledge sharing via a database (Section 3.1.4). A database for knowledge exchange seems legitimate; however, it is not entirely clear how that information can be relevant for different companies.

Dissemination between academia and industry requires an effective relationship between these two; however, the gap between them and a need for improvement has been highlighted before (Le Coze, 2019). Due to this gap, practitioners just rely on industry experiences and do not systematically use scientific evidence (Provan et al., 2019). This idea has been supported by the findings of this research. For instance, practitioners mainly rely on their experience, the applicable industry standards, or the organization's processes when it comes to choosing a hazard analysis method (Section 3.1.1). Not even a single practitioner mentioned that they search and try to find a new method that maybe more suitable for that specific situation. They are using the methods that they are most confident with, or their organizations laid out for them.

These all suggest that regulatory bodies, organizations, and academic institutions have to do much more to improve the current state of the practice. Regulators can create databases with clear guidance on how they should be used and provide funding to safety researchers to investigate the state of the practice in industry, find the problems, and provide solutions that are tested. Organizations can also dedicate more budget for safety-related activities, such as participation in conferences to share their experiences, and using new approaches by showing flexibility in their internal processes.

4.4. Limitations of this study

As with any research, this study is not without limitations. One limitation is using the snowballing as one of the sampling methods which may have biased the results to more established practitioners, as those will be the known practitioners to whom the other interviewees referred to. Restricting the scope of this research to North America is another limitation of this study. Future studies could perform similar research in other geographical areas where regulatory requirements and expectations in terms of professional practice may differ. This would help to provide a more thorough picture of the state of practice in validation of hazard analysis methods among practitioners worldwide.

As mentioned in Section 1.2, we aimed to interview practitioners in safety-critical industries. Frequently mentioned examples of such industries include nuclear Industry, oil and gas, chemical, aviation, rail, space and defense, automotive and maritime industries (Amberkar et al., 2001; Joubert & Feldman, 2017; Lowe et al., 2016; Lwears, 2012; Saunders et al., 2013; Singh & Singh, 2021). We could not interview any

practitioners from the nuclear and maritime industries.

It is interesting to highlight that, in general, the authors found it impossible to deduce differences between sectors from the responses given. Even those practitioners who were involved in more than one industry (refer to Section 3.5 Overview of interviewees) did not differentiate their answers based on their experience in different industries. However, further work in these safety-critical industries to broaden current findings and promote further comparison across industries would be insightful.

The fact that only 15 interviews were needed to reach saturation (moreover across industries) indeed is noteworthy, as it points to the fact that the problems are indeed well known and broadly shared. Given the relative paucity of empirical work on practitioner's views on hazard analysis practices, and the fact that academic work is often relatively detached from safety practices, documenting these can help bridge the science-practice gap (Le Coze, 2019).

5. Conclusion

This article presented the results of twenty semi-structured interviews with system safety practitioners working in North America to investigate the state of the practice in validation of hazard analysis in safety-critical industries. The interviews revealed that although many hazard analysis methods have been developed, only a limited number of them are employed in practice, which are mainly based upon the linear accident causality model. Only a few practitioners from the USA reported using a method based on systems theory. Practitioners highlighted that the methods they use are outdated and are not proper for a complete and comprehensive analysis of today's complex systems. In addition, subjectivity was raised as the greatest weakness of current methods. However, in the discussion section, it is argued that, based on an emerging consensus in the academic literature, hazard analysis is inherently subjective. This is supported by different views on the ontological status of risk: realism and constructivism. As long as hazard analysis can be supported by strong evidence, subjectivity is not necessarily a weakness, although interviewees raised concerns about the subjectivity.

One clear pattern that emerged from the interviews is that the majority of the safety practitioners defined validation based on the system engineering concept. Thus, validation was described as a way to ensure that a system works the way it is supposed to work. What the current research was concerned with is the validation of hazard analysis per se, meaning that to ensure that the assumptions, execution steps, and the results of the analysis are comprehensive and credible. It is found that almost all practitioners in some form validate their hazard analysis. Independent expert reviews and benchmark exercise are the only hazard analysis validation methods reported by the interviewed practitioners. In the discussion, it is argued that these approaches are not without limitations as their quality, as currently applied, comes down to the skill, knowledge, and experience of the individual practitioner. The lack of clear guidance on who should be involved in the validation process and

Appendix

Appendix A. Key definitions and concepts

Terms such as “hazard,” “risk,” “hazard analysis”, and “system safety” have the potential to cause confusion, as these are often used ambiguously. To ensure consistency and to clarify the scope and focus of the current research, this section aims to clarify the intended meaning of these frequently appearing terms, as used by the authors in the context of this work. Note that when these terms are mentioned as part of direct quotes from interviewed practitioners, we keep the original meaning by the interviewees, to avoid misrepresenting the data. The authors furthermore acknowledge that multiple definitions coexist in the literature, as acknowledged also e.g., in an influential glossary of key risk-related terms by the Society for Risk Analysis (Aven et al., 2018) and Kletz (1999). Hence, the definitions given here should be considered as a coherent stipulative basis for the aims of the current work, while no claims are made about the universal applicability of the below provided definitions. The following definitions, except for the validation definition, are adopted from the “Concise encyclopedia of system safety: definition of terms and concepts” by Ericson (2011) and “Hazard

how to actually perform validation exacerbates these challenges. Furthermore, although all practitioners believed that validation adds value to the analysis, and that it indeed improves operational safety, validation as an activity of safety work sometimes is justified mainly through demonstrated safety.

Some opportunities for improving the current state of the practice in validation of hazard analysis result from this research. First, there should exist better guidelines, for both hazard analysis and validation of hazard analysis, for practitioners in terms of who should be involved in the analysis and how the analysis should be performed. A validation framework embedded into a hazard analysis could improve the results, which can further improve operational safety and have subsequent effects on developing safer systems. However, what this validation framework covers and how, is important as a deficient or superficial framework can exacerbate probative blindness.

Second, regulatory bodies, academic institutions, and organizations all are responsible and have to play their roles in improving the current state of the practice. Regulators can create databases with clear guidance on how they should be used and provide funding to safety researchers to investigate the state of the practice in the industry, find the problems, and provide solutions that are tested and proved to be effective. Organizations can also dedicate more budget for safety-related activities, such as participation in conferences to share their experiences, and using new approaches by showing flexibility in their internal processes.

CRediT authorship contribution statement

Reyhaneh Sadeghi: Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Floris Goerlandt:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

The work in this article has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

The researchers wish to thank the practitioners who participated in an interview without whom this research would not have been possible.

analysis techniques for system safety” by Ericson (2015).

System- A system is an integrated group of interrelated and interacting components, including people (e.g., operators), products (e.g., technical artifacts), and processes (e.g., course of action) that aims to achieve a goal. A system is greater than the sum of its parts.

System Safety- System Safety is a discipline for developing safe systems. Through System Safety foreseeable accidents can be prevented by identifying hazards and eliminating or mitigating risks associated with these. The primary concern of system safety is the management of hazards, their identification, evaluation, elimination, and control. A simplified schematic representation of the overall process of system safety is illustrated in Fig. A1.

Hazard- Hazard is a condition that can potentially result in an accident (Fig. A2). Hazards are typically present in a system for different reasons, such as the use of hazardous system elements, the need for hazardous functions, or unknown design flaws.

Risk- Each hazard is associated with one or multiple risks, as constructed by an analyst. Risks can be expressed through possible accident scenario, often stated in terms of hazard likelihood and severity (Fig. A2).

Accident/Mishap- An occurrence that culminates in death, injury, damage, harm, or loss. An accident/mishap happens as a result of an actualized hazard. In system safety, the terms accident and mishap are synonymous and can be used interchangeably. Fig. A2 shows the hazard and accident/mishap relationship.

Hazard Analysis- The process of hazard analysis includes the identification of hazards and assessments of their associated risks, which are the second and third steps, respectively, in the system safety process as illustrated in Fig. A1. Thus, hazard analysis involves identifying hazards that exist within a system, their potential effects, and causal factors and assessing the risks presented by the identified hazards. This provides a basic foundation for system safety so that design measures can be further established to eliminate or mitigate the identified hazards.

Validation- Before defining the term “validation”, as intended in this research, it is important to highlight the scope of this work. This research is solely concerned with the hazard analysis part of a system safety program. As defined, hazard analysis includes identification of hazards and assessment of their associated risks, steps 2 and 3, respectively, of the system safety process (Fig. A1). As a result, the scope of hazard analysis validation, as intended here, is also restricted to the hazard analysis phase of a system safety program. This excludes the wider scope or other steps of the system safety process, such as safety measures identification (step 4 in Fig. A1), risk reduction (step 5 of Fig. A1), and V&V of risk reduction (step 6 in Fig. A1).

In light of this, in this research and in the context of hazard analysis, validation is taken to mean the process of ensuring that a hazard analysis is accurate, comprehensive, and credible. Accuracy focuses on assessing whether the analysis and its results are correct and free of errors (Sargent, 2014). Comprehensiveness concerns the adequacy of the scope, assumptions, implementation steps, and results of the analysis in line with the stated purpose of the analysis (Goerlandt et al., 2017). Credibility refers to the extent to which stakeholders and decision-makers can trust and use the results of an analysis (Sargent, 2014).”.

Appendix B. Interview questions

Part I: Collects general information about interviewees and the companies they work in:

- 1.1. What country (countries) is your company located in?
 - 1.2. How many employees does your company have?
 - 1.3. What is the industry of your company?
- If the company is active in different industries:

- which sector the interviewee work in?

- 1.4. How many years of experience do you have in the risk and safety field?
 - 1.5. What is your job title? And what kind of activities you are involved in?
 - 1.6. What is your level of education?
- Diploma Bachelor's degree Master's Degree Ph.D.
- 1.7. What is your field of study?

Part II: Gathers information about the implemented hazard analysis models:

- a. What hazard analysis method(s) do you currently (or until most recently) use in your organization?
- b. Why did you choose this method? How did you make sure that you chose a correct method?
- c. What stage(s) of the system life cycle are you undertaking the hazard analysis method for?
- d. Is there any specific team/department in your organization who is responsible for implementing and maintaining the hazard analysis method? What is the name of this team/department?
- e. What are the driving factors in implementing the hazard analysis method?
- f. What are the weaknesses of current hazard analysis method?
- g. What are the “opportunities” to improve the current hazard analysis method?
 2. Part III: Questions about validation approach.
 3. How do you (and your colleagues/team) make sure that the assumptions, implementation steps, and results are correct/adequate for the purpose of identifying all hazards?
 4. Do you believe that it is possible to achieve a correct result?
 5. How do you make the results credible to stakeholders? (For instance, the managers or the engineers who use the results of the hazard analysis for the design?)
 6. Do you believe that these activities (related to both assessment and assurance) are effective (or add value to your hazard analysis method)?

7. To what extent have these activities been integrated into the hazard analysis process?
8. What are the motivations/driving factors in performing these activities?
9. What are the key challenges/barriers to perform these activities?
10. What are the weaknesses of these activities?
11. What are the “opportunities” to improve these activities?
12. How do you describe “validation” in the context of hazard analysis?

Appendix C. . List of categories identified in the interviews

(See Table A1).

References

- Amalberti, R., 2013. Navigating Safety: Necessary Compromises and Trade-Offs—Theory and Practice. (1st ed.). Springer Dordrecht. <https://doi.org/10.1007/978-94-007-6549-8>.
- Amberkar, S., Czerny, B.J., D'Ambrosio, J.G., Demerly, J.D., Murray, B.T., 2001. A Comprehensive Hazard Analysis Technique for Safety-Critical Automotive Systems. 2001-01-0674. <https://doi.org/10.4271/2001-01-0674>.
- Andersson, C., Runeson, P., 2002. Verification and validation in industry—A qualitative survey on the state of practice. Proceedings International Symposium on Empirical Software Engineering 37–47. <https://doi.org/10.1109/ISESE.2002.1166923>.
- Andrews, G.C., Shaw, P., McPhee, J., 2019. Canadian professional engineering and geoscience: Practice and ethics (Sixth edition.). Nelson.
- Aven, T., Guikema, S., 2011. Whose uncertainty assessments (probability distributions) does a risk assessment report: The analysts' or the experts'? Reliab. Eng. Syst. Saf. 96 (10), 1257–1262. <https://doi.org/10.1016/j.res.2011.05.001>.
- Aven, T., Ben-Haim, Y., Andersen, H.B., Cox, T., Drogue, E. L., Greenberg, M., Guikema, S., Kröger, W., Renn, O., Thompson, K.M., Zio, E., 2018. Society for Risk Analysis Glossary. 9.
- Aven, T., Renn, O., 2009. On risk defined as an event where the outcome is uncertain. J. Risk Res. 12 (1), 1–11. <https://doi.org/10.1080/13669870802488883>.
- Balci, O., Nance, R.E., Arthur, J.D., Ormsby, W.F., 2002. Expanding our horizons in verification, validation, and accreditation research and practice. Proc. Winter Simul. Conf., 1, 653–663 vol.1. <https://doi.org/10.1109/WSC.2002.1172944>.
- Baybutt, P., 2021. On the need for system-theoretic hazard analysis in the process industries. J. Loss Prev. Process Ind. 69, 104356 <https://doi.org/10.1016/j.jlp.2020.104356>.
- Bhattacharjee, A., 2012. Social Science Research: Principles, Methods, and Practices. Digital Commons University of South Florida.
- Bowen, G.A., 2008. Naturalistic inquiry and the saturation concept: A research note. Qualitative Research : QR 8 (1), 137–152. <https://doi.org/10.1177/1468794107085301>.
- Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>.
- Corbin, J., Strauss, A., 2008. Basics of qualitative research techniques and procedures for developing grounded theory. (3e [ed.] / Juliet Corbin, Anselm Strauss.). SAGE.
- Dallat, C., Salmon, P.M., Goode, N., 2019. Risky systems versus risky people: To what extent do risk assessment methods consider the systems approach to accident causation? A review of the literature. Saf. Sci. 119, 266–279. <https://doi.org/10.1016/j.ssci.2017.03.012>.
- Dekker, S., 2011. Drift into failure from hunting broken components to understanding complex systems. Ashgate.
- Dekker, S., 2019. Foundations of Safety Science: A Century of Understanding Accidents and Disasters. CRC Press LLC.
- Dodshon, P., Hassall, M.E., 2017. Practitioners' perspectives on incident investigations. Saf. Sci. 93, 187–198. <https://doi.org/10.1016/j.ssci.2016.12.005>.
- Dunjó, J., Fthenakis, V., Vilchez, J.A., Arnaldos, J., 2010. Hazard and operability (HAZOP) analysis. A literature review. J. Hazard. Mater. 173 (1), 19–32. <https://doi.org/10.1016/j.jhazmat.2009.08.076>.
- Eker, S., Rovenskaya, E., Obersteiner, M., Langan, S., 2018. Practice and perspectives in the validation of resource management models. Nat. Commun. 9 (1), 5359 <https://doi.org/10.1038/s41467-018-07811-9>.
- Engel, A., 2010. Verification, validation, and testing of engineered systems. Wiley.
- Ericson, C., 2011. Concise encyclopedia of system safety definition of terms and concepts. Wiley.
- Ericson, C., 2015. Hazard Analysis Techniques for System Safety, (2nd ed.). Wiley.
- Etikan, I., Musa, S.A., Alkassim, R.S., 2015. Comparison of convenience sampling and purposive sampling. Am. J. Theor. Appl. Stat. 5(1), Article 1 <https://doi.org/10.11648/j.ajtas.20160501.11>.
- Goerlandt, F., Khakzad, N., Reniers, G., 2017. Validity and validation of safety-related quantitative risk analysis: A review. Saf. Sci. 99, 127–139. <https://doi.org/10.1016/j.ssci.2016.08.023>.
- Harms-Ringdahl, L., 2001. Safety analysis: Principles and practice in occupational safety, (2nd ed.). Taylor & Francis.
- Hollnagel, E., Goteman, Ö., 2004. The Functional Resonance Accident Model. Proceedings of Cognitive System Engineering in Process Plant.
- Joubert, C.G., Feldman, J.A., 2017. The effect of leadership behaviours on followers' experiences and expectations in a safety-critical industry. S. Afr. J. Econ. Manag. Sci. 20 (1), 1–11. <https://doi.org/10.4102/sajems.v20i1.1510>.
- Kaplan, S., 1997. The Words of Risk Analysis. Risk Anal. 17 (4), 407–417. <https://doi.org/10.1111/j.1539-6924.1997.tb00881.x>.
- Kletz, T.A., 1999. Hazop and Hazan: Identifying and assessing process industry hazards, (4th ed.). Institute of Chemical Engineers.
- Kletz, T.A., 2001. Learning from accidents, (3rd ed.). Gulf Professional Publishing.
- Lathrop, J., Ezell, B., 2017. A systems approach to risk analysis validation for risk management. Saf. Sci. 99, 187–195. <https://doi.org/10.1016/j.ssci.2017.04.006>.
- Le Coze, J.-C., 2019. Safety Science Research Evolution. CRC Press LLC, Challenges and New Directions.
- Leveson, N.G., 2017. Rasmussen's legacy: A paradigm change in engineering for safety. Appl. Ergon. 59 (Pt B), 581–591. <https://doi.org/10.1016/j.apergo.2016.01.015>.
- Lowe, A., Hayward, B., Branford, K., 2016. Leadership in safety critical industries: Project Report 1 (2016:11).
- Lwears, R., 2012. Rethinking healthcare as a safety-critical industry. IOS Press 41 (Suppl 1), 4560–4563. <https://doi.org/10.3233/WOR-2012-0037-4560>.
- Mkpat, E., Reniers, G., Cozzani, V., 2018. Process safety education: A literature review. J. Loss Prev. Process Ind. 54, 18–27. <https://doi.org/10.1016/j.jlp.2018.02.003>.
- O'Reilly, M., Parker, N., 2013. 'Unsatisfactory Saturation': A critical exploration of the notion of saturated sample sizes in qualitative research. Qualitative Research : QR 13 (2), 190–197. <https://doi.org/10.1177/1468794112446106>.
- Provan, D.J., Dekker, S.W., Rae, A.J., 2017. Bureaucracy, influence and beliefs: A literature review of the factors shaping the role of a safety professional. Saf. Sci. 98, 98–112. <https://doi.org/10.1016/j.ssci.2017.06.006>.
- Provan, D.J., Rae, A.J., Dekker, S.W., 2019. An ethnography of the safety professional's dilemma: Safety work or the safety of work? Saf. Sci. 117, 276–289. <https://doi.org/10.1016/j.ssci.2019.04.024>.
- Qureshi, Z., 2008. A Review of Accident Modelling Approaches for Complex Critical Sociotechnical Systems.
- Rae, A., Alexander, R., 2017. Probative blindness and false assurance about safety. Saf. Sci. 92, 190–204. <https://doi.org/10.1016/j.ssci.2016.10.005>.
- Rasmussen, J., 1997. Risk management in a dynamic society: A modelling problem. Saf. Sci. 27 (2), 183–213. [https://doi.org/10.1016/S0925-7535\(97\)00052-0](https://doi.org/10.1016/S0925-7535(97)00052-0).
- Reason, J.T., 1990. Human error. Cambridge University Press.
- Reiman, T., Viitanen, K., 2019. Towards Actionable Safety Science. In: Safety Science Research. CRC Press, pp. 203–222. <https://doi.org/10.4324/9781351190237-13>.
- Rosa, E.A., 1998. Metatheoretical foundations for post-normal risk. J. Risk Res. 1 (1), 15–44. <https://doi.org/10.1080/136698798377303>.
- Sadeghi, R., Goerlandt, F., 2021. The State of the Practice in Validation of Model-Based Safety Analysis in Socio-Technical Systems: An Empirical Study. Safety (Basel) 7 (4), 72. <https://doi.org/10.3390/safety7040072>.
- Sandelowski, M., 2004. Using Qualitative Research. Qual. Health Res. 14 (10), 1366–1386. <https://doi.org/10.1177/1049732304269672>.
- Sargent, R.G., 2014. Verifying and validating simulation models. 118–131. <https://doi.org/10.1109/WSC.2014.7019883>.
- Saunders, F.C., Gale, A., Sherry, A., 2013. Understanding Project Uncertainty in Safety-critical Industries. PMI Global Congress, Istanbul.
- Schmittner, C., Ma, Z., Smith, P., 2014. FMVEA for Safety and Security Analysis of Intelligent and Cooperative Vehicles. 282–288. https://doi.org/10.1007/978-3-319-10557-4_31.
- Singh, P., Singh, L.K., 2021. Reliability and safety engineering for safety critical systems: an interview study with industry practitioners. IEEE Trans. Reliab. 70 (2), 643–653. <https://doi.org/10.1109/TR.2021.3051635>.
- Solberg, Ø., Njå, O., 2012. Reflections on the ontological status of risk. J. Risk Res. 15 (9), 1201–1215. <https://doi.org/10.1080/13669877.2012.713385>.
- Stephans, R., 2004. System Safety for the 21st Century: The Updated and Revised Edition of System Safety 2000, Vol. 28. John Wiley & Sons.
- Suokas, J., 1985. On the reliability and validity of safety analysis. VTT Technical Research Centre of Finland. Dissertation [Dissertation].
- Vincoli, J.W., 2014. Basic Guide to System Safety (3rd Edition). Wiley.
- Wassenhove, W., Foussard, C., Dekker, S.W., Provan, D.J., 2022a. A qualitative survey of factors shaping the role of a safety professional. Saf. Sci. 154, 105835 <https://doi.org/10.1016/j.ssci.2022.105835>.
- Wassenhove, W., Foussard, C., Denis-Remis, C., 2022b. A case study on the Industrial Risk Management (IRM) post-master academic education program of MINES Paris PSL University. Saf. Sci. 151, 105733-. <https://doi.org/10.1016/j.ssci.2022.105733>.
- Zheng, X., Julien, C., Kim, M., Khurshid, S., 2017. Perceptions on the state of the art in verification and validation in cyber-physical systems. IEEE Syst. J. 11 (4), 2614–2627. <https://doi.org/10.1109/JSYST.2015.2496293>.

Publication III

Sadeghi, & Goerlandt, F. (2023). A proposed validation framework for the system theoretic process analysis (STPA) technique. *Safety Science*, 162. <https://doi.org/10.1016/j.ssci.2023.106080>



Discussion

A proposed validation framework for the system theoretic process analysis (STPA) technique

Reyhaneh Sadeghi^{*}, Floris Goerlandt

Dalhousie University, Department of Industrial Engineering, Halifax, Nova Scotia, Canada



ARTICLE INFO

Keywords
STPA
Validation
Risk
Hazard

ABSTRACT

Validation is a prominent challenge in the domain of risk management in general, and hazard analysis in particular. Practitioners have highlighted a lack of clear guidance on how to perform validation of hazard analyses, who should be involved, and when to stop the validation process. Aiming to contribute to addressing this issue, this study proposes a validation framework for the System Theoretic Process Analysis (STPA) technique, based on foundational concepts in risk analysis and prior theoretical work on validation in related disciplines. STPA, which is a hazard analysis technique based on System-Theoretic Accident Model and Processes (STAMP) accident causality model, is selected due to its increasing popularity in different industries, and because no validation frameworks have yet been proposed for this technique. The proposed STPA validation framework aims to support a systematic assessment of the analysis's comprehensiveness, accuracy, and credibility. It consists of a set of theory-based concepts that are elaborated as guide questions, each focusing on different aspects of STPA. The framework employs a formative approach, i.e., it aims to help stakeholders systematically reason about the analysis and advise on improvements or further elaboration. To develop this framework, theoretical validation concepts in the pertinent literature in risk science, social science, and operations research, system dynamics, and simulation modeling disciplines have been used. It is recognized that the proposed framework should be further tested to confirm its practical usefulness, and it should be investigated whether it indeed improves the hazard analysis in terms of the envisioned functions.

1. Introduction

STPA is a hazard analysis technique based on System-Theoretic Accident Model and Processes (STAMP), an accident causality model based on systems theory (Leveson, 2004b). STAMP has three main components. First, as opposed to the traditional methods, STAMP considers safety a control problem, meaning that safety is about imposing constraints on the system's behavior rather than preventing failures (Leveson, 2012). Hence, not only component failures but also accidents resulting from component interactions are considered in the analysis (Leveson, 2004a). Second, STAMP considers systems as hierarchical structures where each level controls the activity of the level beneath it (Leveson, 2004b). To determine the required control action, a process model is used (Fleming et al., 2013) which is the third component of STAMP. A process model is defined as a representation of the state of the controlled processes which are kept updated through feedback control loops (Leveson, 2017).

STPA has gained increasing popularity for hazard analysis with

application in different industries (Patriarca et al., 2022). A theoretical analysis of the adequacy of various risk assessment methods in light of the tenets of accident causation in socio-technical systems according to Rasmussen's (1997) systems risk framework, also highlights STPA as one of the few currently available techniques which align with a systems view (Dallat et al., 2019). Nevertheless, some limitations of STPA have been identified which need to be addressed to facilitate a wider application in industry or to be widely recommended by regulatory authorities. Lack of formalism (Dakwat & Villani, 2018), dependence on available information and those who perform it (Harkleroad et al., 2013), its time-consuming nature (Patriarca et al., 2022), and use of abstraction for managing the complexity of a system (Baybutt, 2021) are some of the limitations of STPA, as currently applied. These limitations make the validity of STPA a debatable issue.

Validation has been a topic of significant academic scrutiny in some fields of study, such as system dynamics (Barlas, 1996; Coyle & Exelby, 2000; Forrester & Senge, 1980) and operations research (Finlay & Wilson, 1987; Landry et al., 1983). However, although the lack of focus

^{*} Corresponding author.

E-mail address: reyhaneh.sadeghi@dal.ca (R. Sadeghi).

on validation in safety and risk science has been raised by some researchers (Aven, 2012; Goerlandt et al., 2017b; Habli et al., 2021), it has not been as frequently discussed explicitly in risk management field as other fields of studies. In a study by Sadeghi and Goerlandt (2021), empirical insights into the extent of this issue are provided. These authors showed that performing validation of model-based safety analysis is not a common practice in a selected subset of the academic literature. This includes hazard analysis techniques, which is one of the considered model-based techniques in this work. In another study by Sadeghi and Goerlandt (2022), interviews with system safety practitioners in North America showed that there is recognition of the importance of validation, and a desire to incorporate validation processes in executing hazard analyses. However, practitioners raised the lack of clear guidance and a formal validation framework as an important factor making validation a challenging task and pointed out that work towards this would be valuable.

In a recent review article on STAMP/STPA/CAST by Patriarca et al. (2022), STPA validation has been raised as an important issue that has been missing to a great extent from the reviewed papers. Articles have been published in which the results of STPA are compared with other hazard analysis methods through a case study to determine their comparative merit. For instance, Sulaman et al. (2019) qualitatively compared the results of FMEA and STPA methods using a case study research methodology to compare the effectiveness of the methods and investigate their differences. Also, some theoretical discussions on the validity of STPA have been made. For example, Hulme et al. (2022) studied the criterion-referenced concurrent validity of three systems-based methods, one of which is STPA.

In a paper by Valdez Banda et al. (2019), an initial evaluation framework for design and operational use of an STAMP-based Safety Management System (SMS) is proposed. This article highlights the importance of developing validation approaches for STAMP-based techniques, such as STPA. In addition, reviews by independent experts have been a common method to assess the validity of an STPA analysis (e.g. (Thomas et al., 2012)). However, to the best of the authors' knowledge, there has not been any work specifically focusing on formalizing such reviews into a comprehensive framework to systematically approach the validation of STPA in industry contexts.

To contribute to closing this gap, an STPA validation framework is proposed to systematically approach STPA validation. The purpose of this work is not to provide a final solution to STPA validation, but to propose a starting point in developing a formal validation framework for STPA. Through this framework, the comprehensiveness, accuracy, and credibility of STPA can be assessed. Comprehensiveness deals with the adequacy of the scope, assumptions, implementation steps, and results of the analysis to identify all relevant hazards, Unsafe Control Actions (UCAs), and loss scenarios. Accuracy focuses on assessing whether the analysis and its results are correct and free of errors (Sargent, 2013). Finally, validation is concerned with enhancing the credibility of an analysis (Collier & Lambert, 2019). Credibility refers to the extent to which stakeholders and decision-makers can trust and use the results of an analysis (Sargent, 2013).

The authors would like to advise readers who are not familiar with the STPA technique to read some key resources, such as the STPA handbook by Leveson & Thomas (2018), and the engineering a safer word book by Leveson (2012), before engaging with this article. The remainder of the article is structured as follows. Section 2 discusses the approaches to validation in related disciplines, on which the framework proposed in this study builds. The overall structure and assumptions underlying the proposed STPA validation framework are discussed in Section 3. Section 4 explains the developed framework in detail, focusing on the proposed validation tests for each step of STPA. A discussion is provided in Section 5. Finally, Section 6 concludes.

2. Approaches to validation in related fields of study

Section 2 consists of four sub-sections. In subsection 2.1, it is first explained how literature in other fields of study can form a basis for developing an STPA validation framework. Further in this subsection, the related fields that are selected and the reasons for this selection are explained. Then, subsection 2.2 to 2.4 explain how validation is approached in those related fields.

2.1. Validation in fields related to STPA

In a bibliometric analysis of the model validation literature, Eker et al. (2019) asserted that validation practices in different disciplines and application areas form separate knowledge clusters, with research within domains not citing each other. However, Eckerd et al. (2011) suggested that differences in validation practices of different modeling communities are superficial, and that these rely on common underlying concerns and principles. This creates an opportunity for knowledge exchange between academic communities. Some researchers have engaged in such bridge-building work. For example, Schwanitz (2013) drew on operations research and simulation modeling disciplines to develop a validation framework for integrated assessment modeling of global climate change. Pitchforth and Mengerson (2013) incorporated approaches from the fields of psychometrics and system dynamics to develop a high-level validation framework for Bayesian Networks. In our current study, we take a similar approach by analyzing and synthesizing validation practices in fields that can be argued to be relevant to the practice of hazard analysis, namely risk science, social science, and three other narrower areas of scholarship, which are operations research, system dynamics, and simulation modeling disciplines.

The literature on validation in risk science will be very relevant for developing an STPA validation framework, as the whole hazard analysis process can be framed with a risk management context. As defined by SRA (Aven et al., 2018), a hazard is "a risk source where the potential consequences relate to harm."¹ Hazards should be identified to specify their inherent and unique risks (Ericson, 2005). Hence it is imperative to review how validation is approached in risk science.

Literature on validation in social science can be also a useful base for developing an STPA validation framework. This relates to the realist/constructivist debate in risk science, where state-of-the-art views on the risk concept and hazard analyses consider these to be socially shared constructs (which refer to a possible reality) (Aven & Guikema, 2011; Goerlandt et al., 2017a; Rosqvist, 2010). This is further explained in Section 3.1. A constructivist approach interprets risk as a construct used in the present to refer to possible (future) realities, as a shared mental model by a group of experts and analysts (Goerlandt et al., 2017a). It is the expert judgments, as mediated through risk analysts, which are used throughout the whole process of risk assessment (Aven & Guikema, 2011; Redmill, 2002). If hazard and risk analysis are best understood as an expression shared by a group of experts and analysts, it can be approached as a social phenomenon. Hence, social science concepts regarding validation become meaningful to consider.

Because STPA relies on modeling a system as a safety control structure, it is plausible that insights from the operations research, system dynamics, and simulation modeling disciplines can also be helpful to build an STPA validation framework. In contrast to traditional hazard analysis techniques, in which accidents are considered the result of chain-of-event sequences, STPA (and the underlying STAMP theory)

¹ Note that this definition for 'hazard' is not exactly the same as the one adopted in STAMP and STPA, which defines this as "a system state or set of conditions that, together with a particular set of worst-case environment conditions, will lead to an accident (loss)". In STPA, hazards can be controlled (Leveson & Thomas, 2018). Nevertheless, if the lack of control is considered to be the 'risk source' in the SRA definition, these can be considered equivalent.

explain accident occurrence through inadequate control on the behavior of a system. STPA involves building a model of system components and their functional relationships and interactions through feedback control loops (Leveson & Thomas, 2018). This safety control structure is not a quantitative simulation or another type of mathematical model. It is a conceptual model to structure the analysts' knowledge and understanding of the system, which is subsequently used as a basis to systematically inspect unsafe control actions.

In addition, because of relying on a control structure as a basis for hazard analysis, STPA can be categorized as a model-based safety analysis. Model-based safety analysis can be built upon qualitative methods (Boolean formalisms such as fault trees or event trees) or quantitative methods (Transition systems such as Markov chains and Petri nets) (Abdellatif & Holzzapfel, 2020). In an article by Sadeghi and Goerlandt (2021), hazard analysis techniques which rely on representing the system in a type of model, are considered one of the model-based techniques.

Conceptually, STPA has similarities with system dynamics models, which are particular types of simulation models, as these aim to model complex dynamic systems through various feedback loop structures (Keys, 1988). In PhD dissertation by Dulac (2007), it is explained that system dynamics and STAMP have similarities, which are exploited to propose a dynamic risk management approach by combining STAMP safety control structures with system dynamic modeling principles. Hence, the validation literature in these modeling disciplines is also considered a useful foundation for developing an STPA validation framework.

2.2. Risk science

Although validation has been raised as an important issue in risk research (Aven, 2012; Goerlandt et al., 2017b; Rosqvist, 2010), there are very few studies explicitly focusing on this topic. Risk analysis validation concerns ensuring that a risk assessment accurately reflects the best available knowledge of the risk in question (Aven, 2017). Aven and Heide (2009) investigated to what extent risk analysis fulfills the scientific quality requirements of validity. They used four sub-criteria for the validity of risk analysis and discussed to what extent these are met in light of different perspectives of risk. They concluded that although risk analysis fulfills some of the basic scientific requirements, validity requirements are not in general met.

Lathrop and Ezell (2017) presented a logical structure to address validation from the perspective of using risk analysis for risk management. They proposed sixteen critical elements for the successful use of risk analysis for risk management, each assessed through a validation test. Goerlandt et al. (2017a) presented a review focusing on the validity and validation of safety-related Quantitative Risk Analysis (QRA), addressing its validity using three categories: conceptual, foundational, and pragmatic. They classified the approaches for implementing pragmatic validity to five categories following research by Suokas (1985), namely reality check, peer review, quality assurance, and complete and partial benchmark exercise.

2.3. Social science

Validity is an important concept in social science, which is concerned with the meaningfulness of measurements of concepts that are not directly accessible in empirical reality. In other words, validity addresses whether researchers are indeed measuring what they intended to measure (Drost, 2011). Trochim (2006) defined validation as "the best available approximation to the truth of a given proposition, inference, or conclusion." Therefore, the accuracy of the approximation or measurement plays a crucial role in the concept of validity as understood in social science research.

In general, four types of validity have been suggested by social scientists: statistical conclusion validity, internal validity, construct

validity, and external validity (Drost, 2011; Trochim, 2006). Trochim et al. (2015) proposed construct validity as the overarching category to frame the assessment of measurement quality, with all other validity types falling beneath it. They further divided construct validity into translation validity and criterion-related validity. The former addresses how well the measure is operationalized using two tests: face and content validity. The latter centers on the relationship of a measure to other independent measures, which falls into four categories: predictive, concurrent, convergent, and discriminant.

2.4. Operations research, system dynamics, and simulation modeling disciplines

Significant research has been conducted for the validation of operations research, system dynamics, and simulation models. In these modeling disciplines, validation is often defined as ensuring that the developed model is an accurate representation of the real-world problem (Forrester & Senge, 1980; Gass, 1983). A model is developed for a specific purpose and its validity is determined with respect to that purpose (Groesser & Schwaninger, 2012; Sargent, 2013). Researchers in these fields considered validation an important part of the modeling process (Balci, 1994; Barlas, 1996; Landry et al., 1983). For instance, Landry et al. (1983) proposed a validation framework for operations research models, called the modeling-validating process, in which the model validation process is embedded into the model development processes.

A plethora of model validation tests exist in the literature. Balci (1994) created a long list of validation techniques for simulation models which are categorized into informal, static, dynamic, symbolic, constraint, and formal techniques. The level of mathematical formality and complexity are the main distinguishing features of these categories, meaning that these two features increase moving from informal to formal techniques. Sargent (2013) described validation techniques commonly used in simulation model validation, including face validity, extreme condition validity, Turing test, and structured walkthroughs, a combination of which is used to validate parts of a model or a whole model. Barlas (1996) proposed a logical sequence for validating system dynamics models, which starts with testing the validity of the model structure, followed by tests of the model behavior. In general, there are significant overlaps in validation tests and ideas between these modeling fields.

3. Methodology

As mentioned in Section 2, the validation concepts and practices in risk science, social science, systems dynamics, simulation modeling, and Operations Research fields can be insightful for developing an STPA validation framework. To achieve this, a similar approach as used by Pitchforth and Mengersen (2013) is adopted. The aim is to develop a framework for validating STPA which consists of a series of well-grounded tests. The proposed tests are elaborated as guiding questions with focus on different aspects of STPA which are formatively assessed.

To select papers from the above-mentioned fields (Sections 2.2-2.4), the following steps were carried out. Articles on risk analysis validation were selected based mainly on the papers included in the Safety Science Special Issue "Risk analysis validation and trust in risk management" (Safety Science, 2017), and backwards snowballing, that is checking the references in those papers for relevant articles. Articles in social science validation were selected as found useful in another work on validation framework for expert-based models by Pitchforth & Mengersen (2013). In operations research, system dynamics, and simulation modeling disciplines, articles were obtained from a keyword search ("validation", "validity") in key journals addressing those model types, and further backwards snowballing of the identified articles.

Then, a list of all the identified papers was prepared. The identified papers were screened to investigate if they presented an elaborate set of

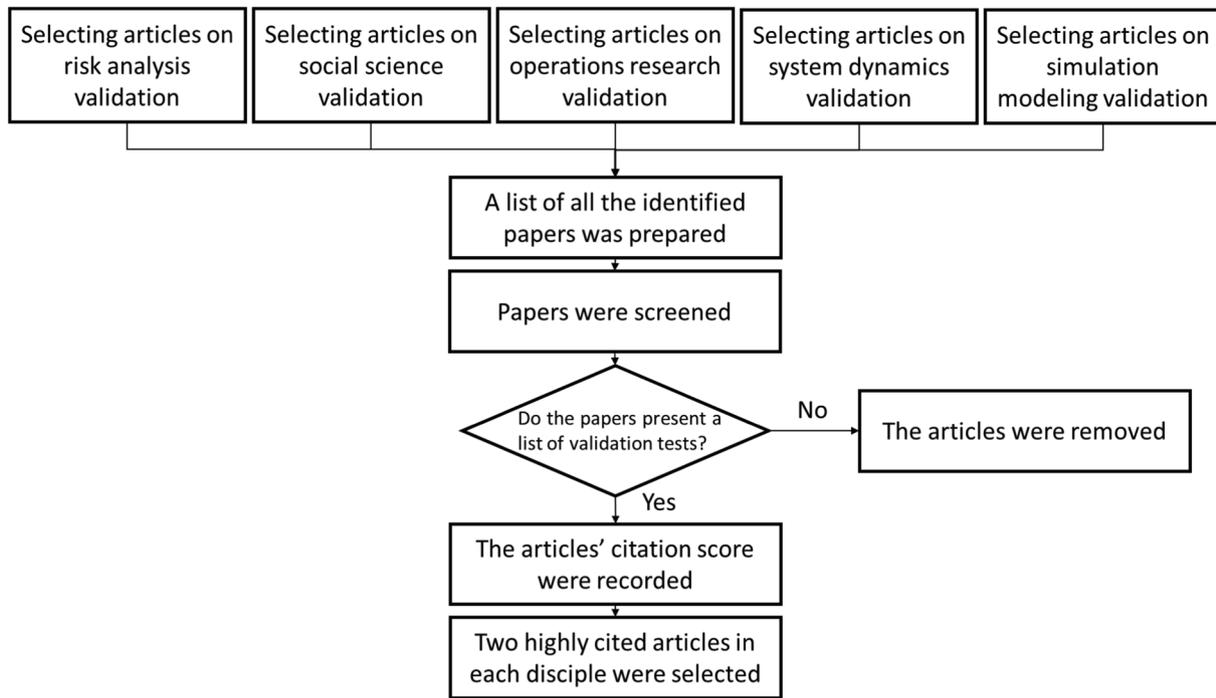


Fig. 1. The process of selecting papers.

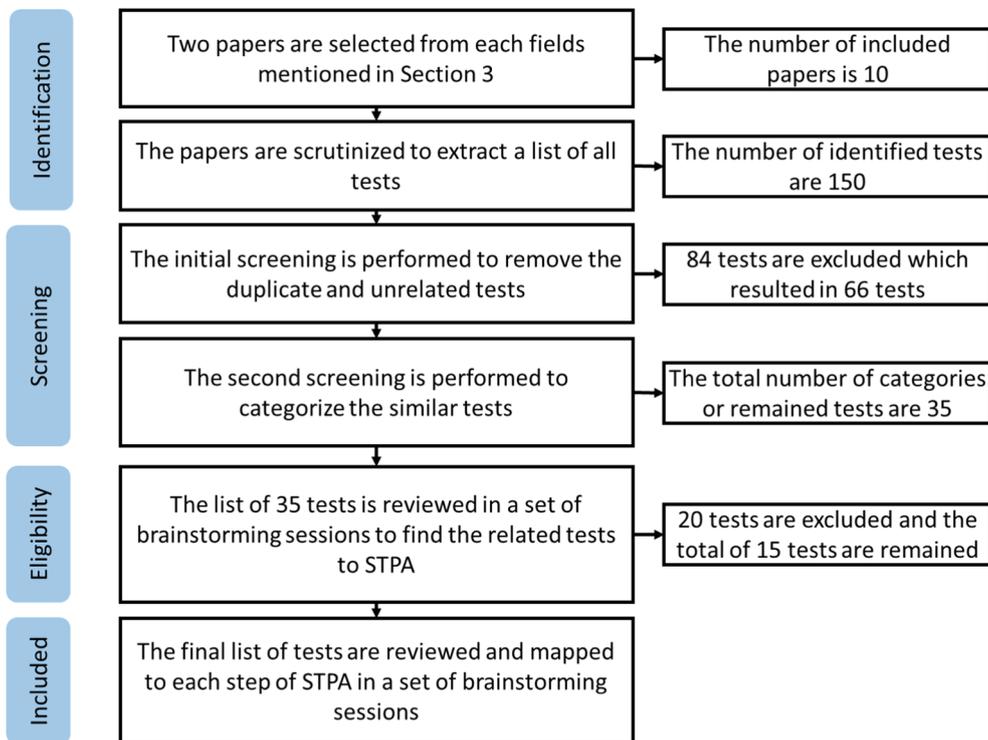


Fig. 2. The process of constructing the STPA validation framework based on PRISMA flow diagram.

validation tests. If not, they were removed from the list. If yes, their citation scores were checked using Scopus and recorded in March 2022. From this list, two highly cited articles in each discipline were selected to be included in the analysis. These articles, which often build on and integrate earlier work in their respective domains, are considered representative of the ideas and concepts underlying validation in those fields. Therefore, we consider these sufficient to serve as a basis for developing the STPA validation framework. The process of selecting the

papers is illustrated in Fig. 1.

The preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement is used to identify, screen, determine eligibility, and include tests for developing the validation framework (Moher et al., 2009). The flow diagram is shown in Fig. 2. The selected articles were thoroughly read to compile a list of validation tests using a close reading method (Brummett, 2019). The compiled list, which included 150 tests, was screened to identify, and eliminate the duplicate and

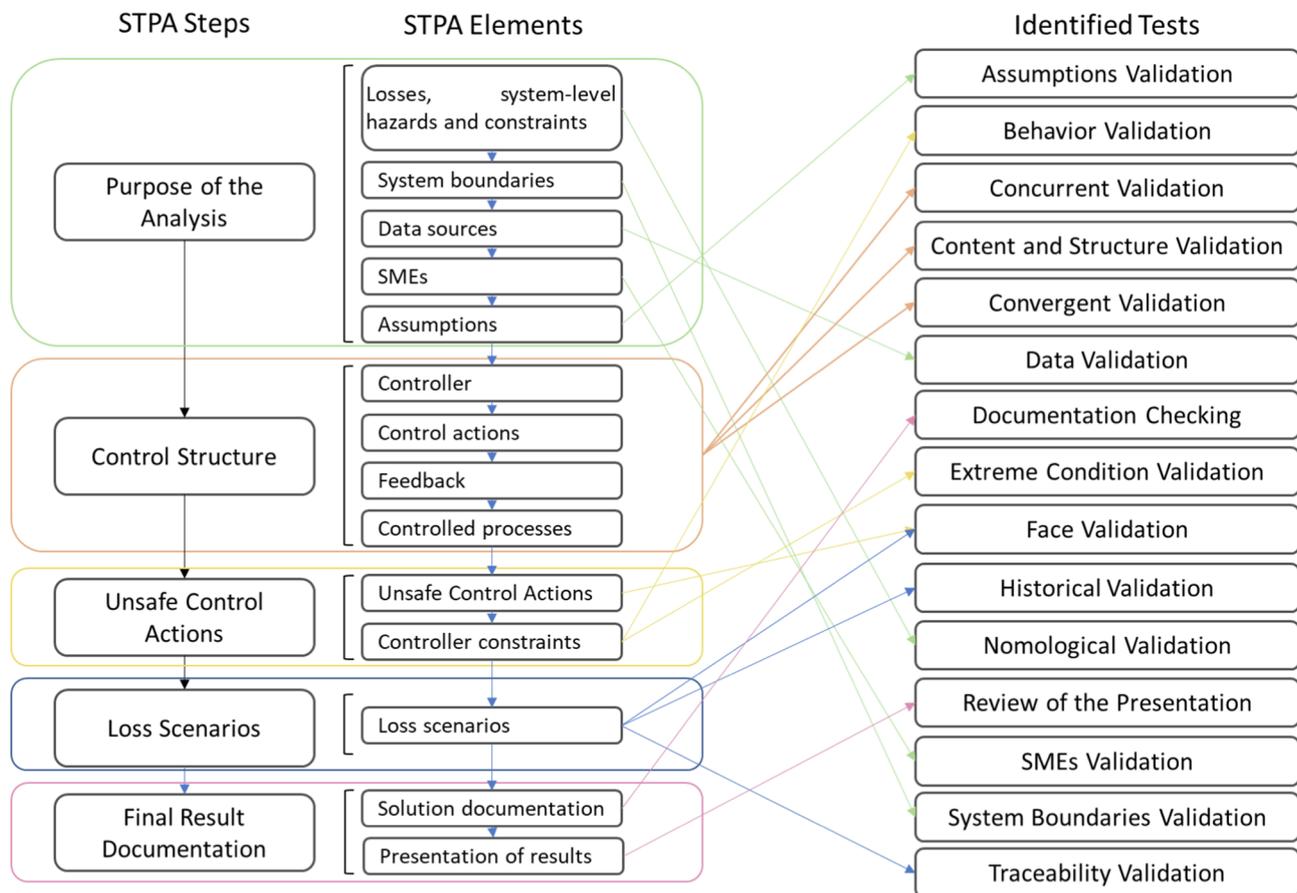


Fig. 3. The assigned tests to each element of STPA.

unrelated tests, e.g., statistical validation which require numerical data. Then, similar tests were classified into a set of categories. That is, there were some tests with different names, while their underlying concept and purpose were the same. Thus, they were classified as one category. For instance, independent peer review, expert review, and face validation were all categorized in one group and called face validation.

This resulted in a list of tests that could be applicable for STPA validation. These activities were carried out and recorded by the first author. The compiled list of tests was sent to the second author, who reviewed the list, and compared this with the tests in the original articles. Finally, the first and the second authors compared their findings in a series of brainstorming sessions, which showed a high level of agreement and led to a final list of tests to use as a basis for building an STPA validation framework, which is illustrated in Fig. 3 under the “Identified Tests” column. The reader is referred to Section 5 for a detailed explanation of each test.

Thereafter, the authors reviewed the process of STPA implementation based on its description in the STPA handbook (Leveson & Thomas, 2018), leading to an identification and listing of different elements of STPA. In total, 14 elements for an STPA analysis are identified, which are illustrated in Fig. 3 under the “STPA Elements” column. The reader is referred to Section 5 for a brief explanation and the STPA handbook (Leveson & Thomas, 2018) for detailed explanation of each STPA element.

Then, the authors considered what validation tests could be applied to what elements of STPA, through a series of brainstorming sessions. In those brainstorming sessions, the first and the second authors took one element in turn and reviewed the list of tests to see what the best solution for testing that element would be. Some of the tests and elements were easily matched. For instance, data validation was the best validation test that matches the data sources element. Other elements that

could not be matched with validation tests easily, needed more contemplation and discussion. For example, the reason why the “content and structure validation” test is assigned to the control structure is that when one makes a model of a system (in the context of STPA, building the control structure), it should be checked what elements in the system are included, how those are defined, and then how those elements are connected to each other. This is because we need to have confidence that our model of the system captures the relevant aspects of that system, without which, the model does not represent the system in its essential features.

As a heuristic, it was intended to assign at least three tests to each four steps of STPA (i.e. purpose of the analysis, control structure, unsafe control structure, and loss scenarios), to enable reviewers to have sufficient guide questions to systematically reflect on the accuracy and comprehensiveness of each aspect and step of the STPA analysis. As the validation of the final STPA results are more concerned with the credibility of an STPA analysis (for more explanation on credibility refer to Section 5.5), two tests are assigned. Combining the different elements of STPA and the identified tests in a logical structure, finally resulted in the proposed STPA validation framework.

It is important to highlight that the authors do not claim that the list of proposed tests is exhaustive. In addition, further thoughtful discussion on the appropriateness and completeness of the suggested tests, as well as phenomenological/empirical research focusing on validation practices specific to STPA analysis contexts, is welcomed. The proposed framework is based on the author’s knowledge and experience with STPA, the academic and professional literature on risk analysis, the literature on validation in risk research, as well as the literature on validation in the related scientific domains referred to in Section 2.

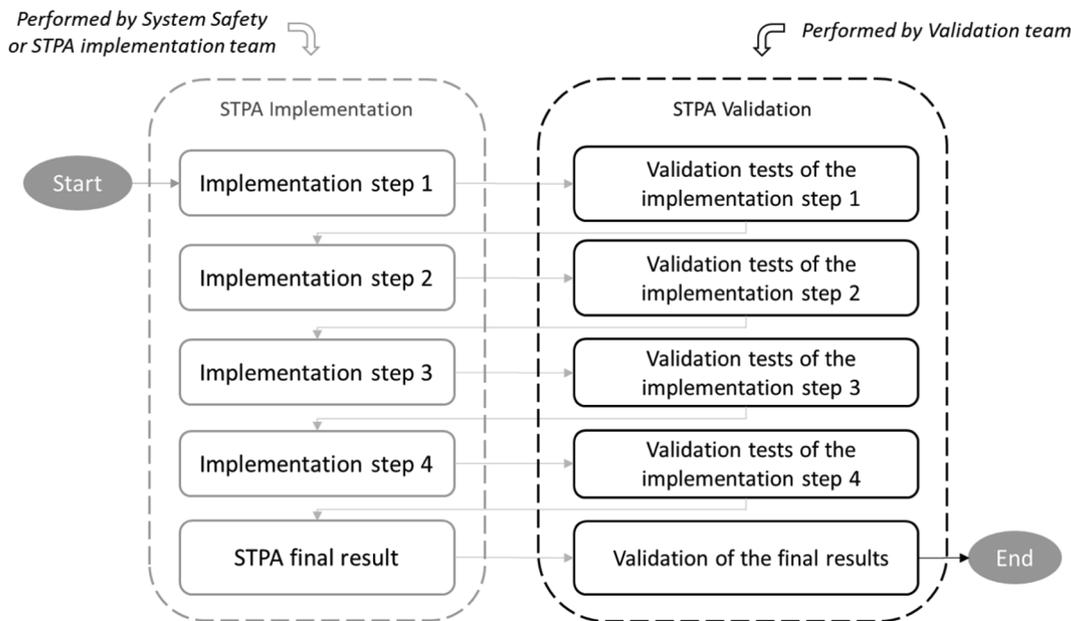


Fig. 4. Using the STPA validation framework in parallel with STPA implementation.

4. Overall structure and assumptions underlying the proposed STPA validation framework

This section is dedicated to the explanations of the overall structure and assumptions behind the proposed framework. For this, first, the conceptual foundations and assumptions are discussed in Sections 4.1. Then, two different applications of this framework are explained in Sections 4.2.

4.1. Conceptual foundations for and assumptions behind the proposed STPA validation framework

The proposed STPA validation framework assesses the validity as a judgment by an assessor, to increase the intersubjective agreement about the comprehensiveness, accuracy, and credibility of STPA between analysts, users, and stakeholders. This aligns with an understanding of hazard analysis as a subject-bound activity, where hazards are better understood as management-oriented social constructs than as objective realities (Dekker, 2019). In the risk research field, the topic of subjectivity of hazard and risk analysis has been discussed by many researchers, often by distinguishing between the realist and constructivist views of risk analysis (e.g. Bradbury, 1989). In general, risk realists assert that a certain state of the world can objectively be defined as risk, but even then such authors usually acknowledge that risk descriptions (especially in complex systems) are dependent on the stake of knowledge of analysts and experts, and hence social phenomena (Rosa, 1998). Risk constructivists assert that the core concept of risk itself refers to an assessor’s knowledge and uncertainty about the occurrence of an event, and hence is inherently subjective (Aven & Renn, 2009). Therefore, the subjective nature of hazard and risk analyses (descriptions of hazards and risks) is supported by different views.

Validation itself is also considered to be an inherently subjective

process. According to Barlas (1996), no validation can claim to be entirely objective as the assessors convey their judgments into the analysis. Even when using quantitative validation techniques, such as statistical validation, subjectivity is an integral part of the process. Barlas (1996) gives the example of determining a significance level (e.g., 0.05) by those who perform the analysis. Therefore, the proposed validation framework relies on subjective assessments by independent reviewers, focusing on the comprehensiveness, accuracy, and credibility of STPA, to increase the inter-subjective acceptance of the analysis.

The developed validation framework is formative, aiming to help peers and stakeholders reason about the analysis in a systematic manner and give advice for improvement or further elaboration. A formative process thus serves as an aid to thinking, rather than aiming to rate an analysis to be of a certain, quantitative standard (Busby and Hughes, 2006). According to Landry et al. (1983): “validation tests point to the areas where the possibility of some improvements exist.” Hence, the validation framework does not aim to lead to a binary accept/reject decision, or a numerical score to support such a decision. Instead, the primary goal is to enable a systematic critical reflection of STPA (or part thereof), including its limitations, errors, and areas in need of improvement.

4.2. Different applications of the STPA validation framework

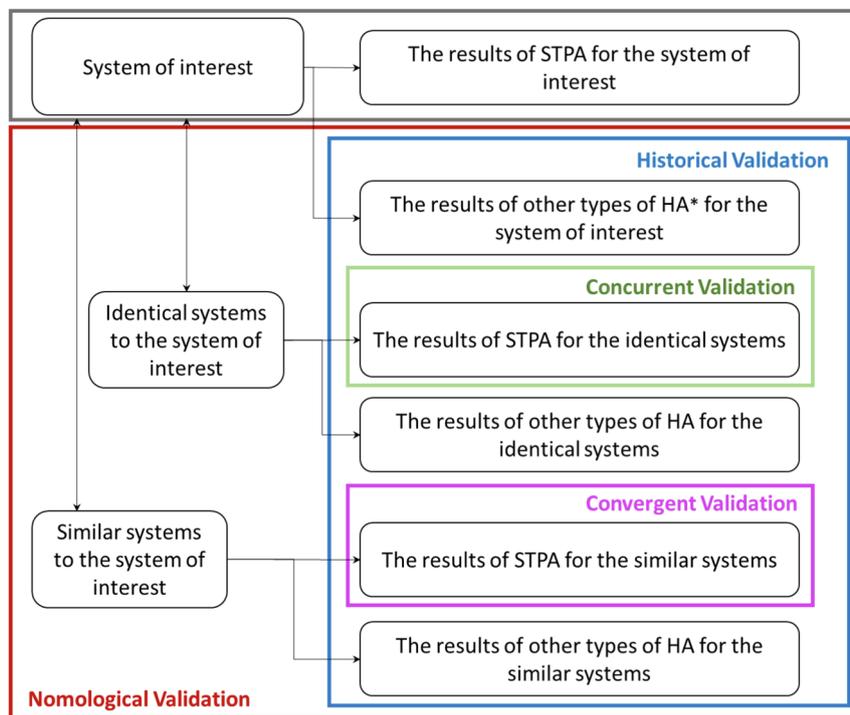
The proposed validation framework is primarily intended to be used in two types of processes: (1) in parallel with the STPA implementation (Fig. 4), or (2) after a complete STPA is performed (Fig. 5). These processes are discussed in Sections 4.2.1 and 4.2.2, respectively.

4.2.1. Performing validation in parallel with the STPA implementation

The process of using the STPA validation framework in parallel with STPA implementation is illustrated in Fig. 4. Here, STPA



Fig. 5. Performing validation using the STPA validation framework after STPA implementation.



* HA stands for Hazard Analysis

- The boxes within the gray line indicates the results of STPA for the system of interest that can be validated using the proposed STPA validation framework.
- The boxes within the red line shows the results of both STPA and other types of hazard analysis (using other techniques) for both identical and similar systems to the system of interest as well as the results of other types of hazard analysis for the System of Interest. These are all identified in the Nomological Validation step, serving as a basis for the Concurrent, Convergent, and Historical Validation steps.
- The box within the green line indicates the results of STPA for the identical systems to the System of Interest. This is used for the Concurrent Validation step.
- The box within the purple line indicates the results of STPA for the Similar systems to the System of Interest. This is used for the Convergent Validation step.
- The boxes within the blue line indicates all the analyses found in the nomological validation step which is used for the Historical Validation step.

Fig. 6. Mapping of the identified studies in the Nomological validation for the Historical, Convergent and Concurrent validation tests.

implementation and validation are performed in parallel, i.e., once an implementation step is done, the associated validation step is carried out. The idea of a parallel process is inspired by its widely adopted use in the simulation field (Landry et al., 1983; Law, 2014; Oral & Kettani, 1993). In a study by Landry et al. (1983), the modeling-validating process is proposed, in which the validation process is integrated into the modeling process. They suggested that it is better not to separate these two processes, to reduce the effects of compounding shortcomings throughout the analysis. Validating STPA in parallel with its implementation enables identifying possible errors during the execution of STPA, as soon as a specific step is complete. This way, errors or omissions are not carried along to subsequent analysis steps. This also aligns with the idea of validation in system engineering where validation is performed throughout the entire system's lifecycle to detect faults as early as possible (Engel, 2010).

In this form of utilizing the proposed framework, two separate teams lead the STPA implementation and validation processes. This is because experts who perform the analysis would likely not disagree with their

own judgments (Pitchforth & Mengersen, 2013), which can be related to a psychological phenomenon known as 'the IKEA effect' in which people overvalue their own creation (Norton et al., 2012). Hence, it is proposed to have two separate groups involved: an analysis implementation team and an analysis validation team. Some organizations have separate teams for system engineering, system safety, and validation, as practitioners believe that independence between these teams is of key importance to enable free and open constructive criticisms about the result of an analysis (Sadeghi and Goerlandt, 2023). In such cases, STPA validation is carried out by the validation team. However, if different teams do not exist, the system safety team can be divided into two groups to achieve an independent review. The first team then implements the STPA, while the second team validates the first team's analysis. Once implementation and validation of a given STPA step are finalized, the two teams meet and discuss their results before moving to the next STPA implementation step.

As part of this discussion, the level of agreement between these teams is assessed. If the level is low, the analysis is iteratively amended until an

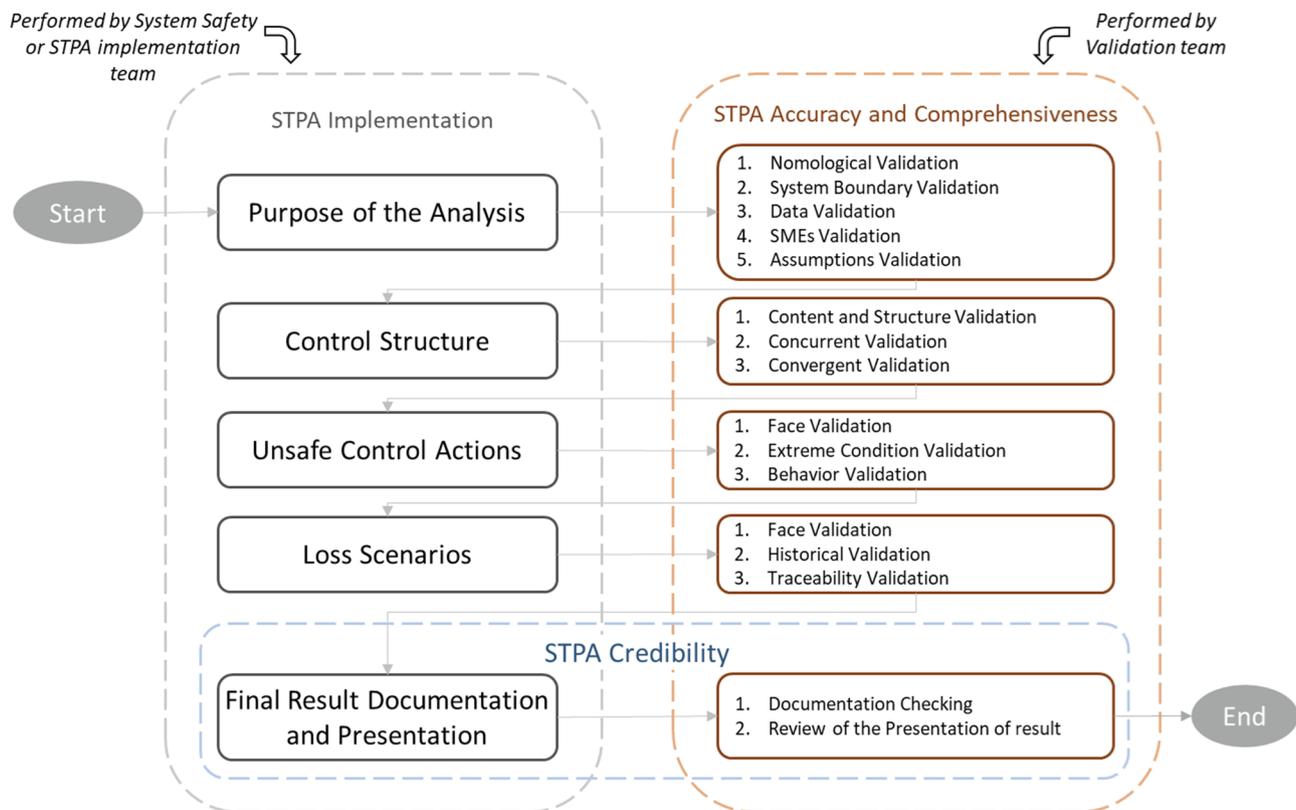


Fig. 7. Using the STPA validation framework in parallel with STPA implementation.

acceptable level is reached, which signals that validation can be ceased (Groesser & Schwaninger, 2012). Following the findings in research by Sadeghi and Goerlandt (2022), ceasing validation is essentially a judgment, which cannot be simply reduced to some quantitative criteria. In practice, given typical budget or schedule limitations, no matter what those quantitative criteria are, the validation will be ceased. Flexibility in the process to allow for expert judgments will be beneficial. Hence, it is proposed that the validation cessation should be decided in a brainstorming session between the two teams.

This iterative process of amending the analysis is considered an ideal way of performing validation, which is hypothesized to lead to the best results. However, due to practical limitations, such as time and/or resource availability, this process can be shortened or eliminated, provided that the disagreement is reported to the stakeholders and decision-makers. Such a disagreement is one of the key mechanisms to explicitly address uncertainties in STPA. This is important, as an explicit consideration of uncertainty has been widely argued in the risk science literature to be an essential aspect of risk assessment (e.g. Flage & Aven, 2009), and has also been discussed in a context of STPA in particular (e.g. Bjerga et al., 2016). Although it is argued that STPA reduces uncertainties compared to linear event-based techniques, it does not eliminate the uncertainties completely due to limitations in the evidence on which the analysis is based (Wróbel et al., 2018). Therefore, it is the responsibility of the analysts to make those uncertainties explicit and communicate them with stakeholders and decision-makers, if the STPA results are used in a risk assessment context (Goerlandt & Reniers, 2016).

4.2.2. Performing validation after the STPA implementation

Another way of validating STPA is using the proposed validation framework in a post hoc manner. That is, the framework can be used for an already existing and completed STPA. One example of this can be situations where an external certification body or a regulatory authority wants to validate the results of a company's hazard analysis. Another

case can be when a company outsources the process of hazard analysis to an external consultant, which is not uncommon in the industry (Sadeghi and Goerlandt, 2023). Application of the validation framework in a post-hoc manner can be also used for situations where the validation team is involved in later stages of the analysis, for instance if a company cannot spend more time on the validation steps as proposed in Section 3.2.1 due to schedule limitations. In such cases, as shown in Fig. 5, all the tests can be performed once the STPA implementation is done.

As can be seen in Fig. 5, once STPA implementation is completed, the results are handed over to the STPA validation team to perform the validation. In cases where the analysis and validation happen within the same organization by two different teams, once all the validation tests are carried out, the level of agreement between the two teams can be assessed. If a low level of agreement is reached, the analysis needs to be amended until reaching an acceptable agreement level. Then, the validation can be ceased. As mentioned in Section 3.2.1, if the agreement is low, and the company, for any reason, is not willing to continue the validation, the disagreements need to be reported to the stakeholders and the decision-makers.

Using the framework in a post-hoc manner is not without limitations. If the validation is performed after the STPA is implemented, there is a risk of not considering the validation results at all. If the validation starts too late, and some errors are found in the analysis, it may be too late for the STPA implementation team to influence the results of the analysis.

5. The STPA validation framework

This section consists of five sub-sections. Each of the first four sub-sections is associated with one of the four main steps of STPA, while the last sub-section relates to the final results of an STPA analysis. For reasons of brevity, it is assumed that readers are familiar with the STPA technique, so each sub-section only briefly outlines the related step. For further details, the reader is referred to the STPA handbook by Leveson & Thomas (2018), and the engineering a safer word book by Leveson

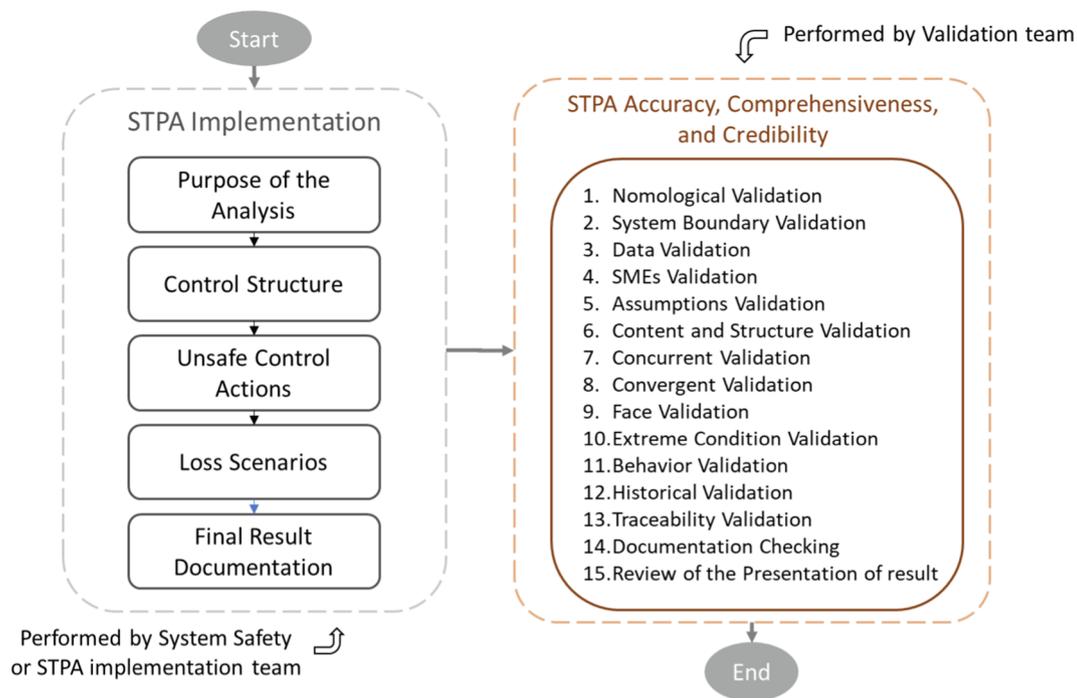


Fig. 8. Performing validation using the STPA validation framework after STPA implementation.

(2012).

For each step of STPA, the proposed validation tests and associated guide questions are elaborated in the following subsections. The guide questions are designed in three different ways: (1) extracted from literature and amended to be used in the context of STPA, (2) developed based on important notes in STPA handbook, or (3) developed based on the authors knowledge and experience in the field of hazard analysis, risk research, and validation, using the ideas captured in the validation tests extracted from the literature described in Section 2. In case the question is developed using the literature (case 1) or STPA handbook (case 2), the related research is cited.

The proposed framework suggests high level guide questions which may need to be further refined in practical use cases, or when applied in different industries. Hence, the authors do not claim that the list of guide questions is exhaustive, i.e. additional questions and further improvements may be needed. It is highlighted in Section 6. Discussion, why and how this proposed validation framework should be tested to ensure it achieves its objectives.

Figs. 7 and 8 are presented at the end of Section 5 to illustrate the overall process of using STPA validation framework in terms of the two different applications as explained in Sections 4.2.1 and 4.2.2, respectively. In addition, a summary overview of the complete validation framework is provided in Appendix A, showing different step of STPA, validation tests, the guide questions, and the associated validation functions.

5.1. Purpose of the analysis

5.1.1. Summary of the STPA step

STPA starts with specifying the purpose of the analysis, which has three main steps: (i) identifying losses (ii) identifying system-level hazards, and (iii) identifying system-level constraints (Leveson & Thomas, 2018). These steps form the foundations of the analysis, in which basic elements of the analysis, such as assumptions and system boundary, are also specified. Therefore, to ensure that the analysis has a solid foundation, careful consideration should be put into this step before moving to the next steps of the analysis.

5.1.2. Validation tests

5.1.2.1. Nomological validation. Pitchforth and Mengerson (2013) defined nomological validation as establishing confidence that a model domain fits within a wider domain based on the literature. In the context of STPA validation, through nomological validation, other STPA studies or other hazard analyses using other techniques (e.g., Fault Tree Analysis) for similar or identical systems in academic or grey literature (such as industry reports) are identified. The identified analyses can be used as a basis for comparison throughout the analysis. This forms a basis for a benchmark exercise, which has been identified as one of the most frequently used validation approaches in both academia (Sadeghi & Goerlandt, 2021) and industry (Sadeghi and Goerlandt, 2023). The following guide questions can be answered as the test of nomological validity for the purpose of the analysis step:

- Is there a similar/identical analysis in the existing literature or industry to support the result of this analysis?
- If yes, which identified analyses are nomologically adjacent and in which way, and which ones are distant, to the system of interest and its associated STPA? (Pitchforth & Mengersen, 2013)
- If not, have the STPA implementation team clearly explained why this analysis lies outside all current known research?

If there exist similar/identical analyses, they can be used to reflect on the comprehensiveness and accuracy of the identified losses, system-level hazards, and system-level constraints in the STPA step “purpose of the analysis”. Therefore, in addition to the above-mentioned questions, the following question is answered:

- If there exist similar/identical analyses, how similar are the identified losses, system-level hazards, and system-levels constraints?

As mentioned above, the identified studies through nomological validation can be used in other tests of the proposed validation framework, which are concurrent, convergent, and historical validation tests. Fig. 6 illustrates which identified studies can be used for each of the concurrent, convergent, and historical validation tests. For more

information on these tests, refer to [Sections 5.2 and 5.4](#).

5.1.2.2. System boundary validation. Before identifying system-level hazards, the system and the system boundary are identified in relation to the purpose and context of the analysis. System boundary distinguishes the system from its environment through which it is specified what elements the system has control over, and what elements it does not. It must be aligned with the purpose of the analysis (Forrester & Senge, 1980). As an example, it could be relevant for a specific analysis to include human actions within the system boundary (e.g. (Lordos et al., 2019)), while in other cases these could be excluded. Also, the system boundary validation test requires a conceptualization of the effects of changes in the system boundary (Forrester & Senge, 1980). To do so, the validation team evaluates how a change in the identified boundary by the implementation team would affect the identified losses, system-level hazards, and system-level constraints. The following questions are suggested for testing the comprehensiveness and accuracy of the defined system boundary:

- Is the system boundary explicitly defined and documented? (Lathrop & Ezell, 2017)
- Is the system boundary in line with the purpose of the analysis?
- Do the system designers or operators have control over all the identified elements included within the system boundary? (Leveson & Thomas, 2018)
- How would modifying the system boundary change the identified losses, system-level hazards, and system-level constraints? (Forrester & Senge, 1980)
- How would changes in the environmental context of the system affect the need for adjusting the system boundary to meet the purpose of the analysis?

5.1.2.3. Data validation. Unlike areas in which objective data are widely available and utilized, STPA does not rely on numerical input data; however, it does rely on a description of the analyzed system provided by human experts and by documents they produce (Harkleroad et al., 2013). For instance, in some cases where STPA is used to provide safety insights earlier in the concept development phase, it only requires a description of a system's components and their control relationships. Even in such cases, the quality of any source of input data, such as design documents and industry-related standards, plays a significant role in the results of STPA. Therefore, the sources of data are clearly specified and validated. The following questions are suggested as a test of data validation:

- What are the sources of input data?
- How reliable are the instruments (e.g., software) and processes (e.g., surveys) used for data collection and measurement (e.g., accident data)? (Balci, 1994)
- Are all sources of input data, including but not limited to design documents, industry-related standards, and historical data, up to date? (e.g., do they use the latest version of the design documents, or technical standards?)

5.1.2.4. Subject Matter experts (SMEs) validation. As mentioned above, STPA, similar to other safety analysis types, heavily relies on the knowledge and experience of the people involved in the analysis, both analysts and Subject Matter Experts (SMEs). According to Rosqvist (2010), the analysis is framed within analysts' and SMEs' mental models, which are not stable and cannot be examined. Expert judgment is raised as an important validation criterion in risk analysis (Aven & Heide, 2009), which is influenced by the methods of SME selection, elicitation, and combination (Lathrop & Ezell, 2017; Rae & Alexander, 2017a). SMEs selection concerns the process of identifying proper SMEs. Some selection criteria include the SME's knowledge of and experience

in the field of interest, the diversity of their backgrounds, and interest in the project (Cooke & Goossens, 2004). Furthermore, the number of experts involved in the analysis is an important decision since having multiple SMEs can mitigate the risk of a single expert's bias, but it may not be possible, for example, due to resource or time constraints (Boring et al., 2005).

SMEs elicitation and combination of their judgements are also performed systematically. Various elicitation and combination methods have been proposed in the literature (e.g. Cooke, 1991; Kaplan, 1992). SMEs' opinions can be elicited and aggregated using social or mathematical methods, or a combination of the two, but no single method is best in all circumstances (Clemen & Winkler, 1999). Rae and Alexander (2017a) suggested using a simple combination method as there exists no evidence that complicated methods, such as neural networks, have advantages over simple methods. Despite the choice of method, while performing elicitation, it is important to be mindful of the scrutability (being comprehensive and open to evaluation), fairness (experts providing a balanced view), and neutrality (not being prejudiced (Cooke & Goossens, 2004)).

Based on the above, to validate SMEs, the validation team investigates the SMEs selection, elicitation, and aggregation criteria and processes performed by the STPA implementation team. This test can be also used for the validation of SMEs who are selected for performing validation. Hence, the following questions are suggested to test SMEs validation:

- Is the process of SMEs selection, elicitation, and combination clearly and completely documented? (Lathrop & Ezell, 2017)
- Is the SMEs selection systematically conducted? Are the SMEs selection criteria reasonable considering the purpose of the analysis? Is the entire system covered with appropriate knowledgeable experts?
- Is the SMEs elicitation process systematically conducted? Is this process reasonable considering the purpose of the analysis?
- If more than one SME is involved in the analysis, are the results of SME elicitations combined in a meaningful way?

5.1.2.5. Assumption validation. Assumptions are critical in risk assessments and form an important part of the background knowledge (Flage & Aven, 2009). The assumptions are often simply presumed, while there may be alternative assumption sets that can have an important effect on the results (Lathrop & Ezell, 2017). Various methods have been proposed for the validation of assumptions. In this framework, the proposed assumption validation tests for STPA are adapted from a method proposed by Landry et al. (1983), rooted in ideas by Mason & Mitroff (1981).

This method has two main steps. First, it investigates whether all the identified assumptions are relevant. To do so, the STPA implementation and validation teams examine whether the opposite of any particular assumption would change the results of STPA. A 'no' answer indicates that the assumption is not very relevant to the problem situation. Second, the degree of importance and certainty of each relevant assumption is determined based on a judgment about their perceived impact on the analysis. For this, relevant assumptions are plotted on an importance-certainty graph.

This aims to help both teams to be in more agreement and gain a deep understanding of the assumptions on which the analysis will be based. In addition, this assumption validation test can lead to more accurate and comprehensive assumption sets. Therefore, to validate assumptions, in addition to the above-mentioned steps in creating an assumption importance-certainty graph, the following questions are answered:

- Are the assumptions fully identified, accurately described, understood, documented, and agreed upon? (Lathrop & Ezell, 2017)

- If the opposite of any particular assumption were true, would it have a substantial effect on the identified losses, system-level hazards, and system-levels constraint?
- Are the degree of importance and certainty of each relevant assumption credibly determined? (Landry et al., 1983)

5.2. Control structure

5.2.1. Summary of the STPA step

The next step of STPA consists of defining a hierarchical control structure. A control structure is a diagram that depicts the components of the system, and their functional relationships with feedback control loops (Leveson & Thomas, 2018). Different components of a control structure are controllers, control actions, feedback, and controlled processes. A controller issues control actions on a controlled process based on a control algorithm or procedure, which represent the controller's decision-making process and its underlying process models, i.e., the beliefs serving as a basis for those decisions. The definition of the control structure is a critical step in STPA since it is used as a guide for identifying and mitigating the Unsafe Control Actions (UCAs).

5.2.2. Validation tests

5.2.2.1. Content and structure validation. According to Eckerd et al. (2011), a model has content validity if the variables and parameters included in the model are an accurate representation of those variables and parameters in the real system. Through content validity, the adequacy of the elements included in a risk model are tested in relation to knowledge about the system and what is understood to be relevant in the real system (Goerlandt & Montewka, 2015). Hence, in STPA, the content validity test can be conducted for validating the elements included in the control structure.

In the context of Bayesian Networks, Pitchforth and Mengerson (2013) looked into both elements and the relationships between those elements in their test of content validity. Bollen (1989) defined content validation as a qualitative type of validation where the analysts judge whether the necessary structural relationships to satisfy the purpose of the analysis are included. There are also studies in which a structural validation test is proposed only for validating the relationships between elements of a model, whereas the validity of elements themselves is tested separately (Barlas, 1996; Forrester & Senge, 1980).

For the sake of clarity, in the proposed STPA framework, this test is denoted 'content and structure validation' as it aims to investigate the accuracy and comprehensiveness of both the elements included in the control structure, as well as their functional relationships. Hence, through this test, first the contents (what elements are in the model), and then the structure (how the elements are related) are validated. The suggested questions are:

- Does the created control structure include all relevant elements and system components? (Balci, 1994)
- Does the created control structure include all relevant functional relationships between these elements?
- Is the control structure an accurate representation of the system?
- Does the level of detail included in the control structure suffice for the purpose of the analysis?

5.2.2.2. Concurrent validation. In social science, concurrent validity refers to taking a similar, preferably, validated measure and comparing it with the outcome of the existing measures (Drost, 2011). Pitchforth & Mengerson (2013) used the concurrent validity concept in the context of Bayesian Networks (BNs), where it is taken to mean the possibility that a network or section of a network is identical to another network. Concurrent validity can also be used for the validation of a control structure in STPA. This can be defined as the possibility that a control structure

has the same content and structure as the control structure of an identical system. If other STPAs exist for an identical system (or subsystems), which would be identified in the nomological validation step, the developed control structure in those studies can be used as a basis for comparison. The suggested questions for performing concurrent validation are:

- Has any STPA for an identical system been identified in the nomological validation?
- If yes, is the developed control structure, including the controllers, controlled processes, control actions, and feedbacks, the same as the control structure in the identified STPA? If there are differences, why do these appear? (Pitchforth & Mengerson, 2013)

5.2.2.3. Convergent validation. Unlike concurrent validation, which investigates identical systems, convergent validation studies and compares the results of similar (not identical) systems (or sub-systems). In social science, convergent validation evaluates whether there is convergence across different measures of similar constructs (Drost, 2011). Strong correlations between different measures provides evidence of convergent validity. In Bayesian Networks, convergent validity investigates similarities in the structure, parametrization, and discretization of models for similar systems (Pitchforth & Mengerson, 2013). Based on these definitions, convergent validity of the control structure in STPA refers to investigating how similar the control structures are in the identified analyses of similar systems. That is, in this test, the identified nomologically adjacent analyses in the nomological validation test are used for performing the comparison (Section 4.1). The suggested questions are:

- Has any STPA for a similar system been identified when conducting nomological validation?
- If yes, how similar is the control structure to other control structures in the analysis of similar systems? If there are differences, why do these appear? (Pitchforth & Mengerson, 2013)

5.3. Unsafe control actions (UCAs)

5.3.1. Summary of the STPA step

This step of STPA consists of determining how the controlled system can get into a hazardous state and lead to accidents/losses. The control actions are reviewed to investigate how they can, in a particular context and worst-case environment, lead to a hazard. Controllers may issue UCAs by (i) not providing the control action, (ii) providing the control action, (iii) providing a potentially safe control action but too early, too late, or in the wrong order, (iv) providing the control action that lasts too long or stops too soon. Once UCAs and their causal factors have been identified, they are translated into constraints on the behavior of each controller (Leveson & Thomas, 2018).

5.3.2. Validation tests

5.3.2.1. Face validation. In the context of model validation (Section 2.4), face validation is defined as a peer review process where experts assess whether the model looks reasonable to them (Collier & Lambert, 2019; Sargent, 2013). Face validation has been also employed in social science, where it is taken to mean a subjective judgment on the operationalization of a construct, "at face value" (Drost, 2011; Trochim et al., 2015). Although face validity is one of the most commonly used tests for validation, it is considered the weakest form of validity, as analysts most likely would not disagree with their own analysis (Pitchforth & Mengerson, 2013). As mentioned above, this can be explained by a phenomenon known as the "Ikea effect", by which people overvalue their own creations (Norton et al., 2012). Having a separate, independent team of experts for performing validation would alleviate this issue to an

extent, as a positive outcome of a face validation test can be considered to increase the intersubjective agreement about the analysis. To perform face validation for UCAs, the validation team, who are knowledgeable in (aspects of) the domain of the studied system, reviews the identified UCAs and constraints to judge whether they appear reasonable and accurate. The suggested questions are:

- Are the identified UCAs logical and accurate from the validation team's perspective? (Sargent, 2013)
- Are there any other possible UCAs that have not been identified by the STPA implementation team?
- Are the identified UCAs accurately translated into constraints on the behavior of each controller? (Leveson & Thomas, 2018)

5.3.2.2. Extreme condition validation. In system dynamics (Section 2.4), extreme condition validation refers to assessing the validity of the model equations under extreme conditions (e.g., minimum, or maximum plausible input parameter values). Through this test, the plausibility of model results are evaluated against the knowledge of, or a judgment about, what would happen under a similar condition in the real system (Barlas, 1996). In the context of simulation models (Section 2.4), Sargent (2013) defined this test as the plausibility of the model outputs for any extreme or unlikely combination of levels of factors in the system, in relation to a reasonable expectation of how the real system would respond to such conditions. Based upon these definitions, in the proposed STPA validation framework, this test can be used to evaluate the plausibility of constraints, which are derived from the UCAs, under extreme conditions both within and outside of the system boundary. The suggested questions for the extreme condition validation are:

- Are the identified constraints plausible for extreme and unlikely interactions of components within the system? (Sargent, 2013)
- Are the identified constraints plausible for extreme and unlikely conditions in the system's environment?

5.3.2.3. Behavior validation. Model behavior testing has been widely applied for validating systems dynamics and simulation models (Section 2.4), mainly understood as comparing the model response to that of the real system in similar conditions (Balci, 1994; Barlas, 1996; Forrester & Senge, 1980). In STPA, there exists no quantitative model to link inputs and outputs. As mentioned in Section 2.1, the control structure is not a simulation or other type of mathematical model but is a conceptual model to structure the analysts' knowledge and understanding of the system, which is subsequently used to systematically identify unsafe control actions, and define constraints (Leveson & Thomas, 2018). Although the behavior validation test cannot be used to compare the behavior of the model with that of the real system, it can be used for validation of the identified constraints.

The behavior validation test is first suggested by Harkleroad et al. (2013) for STPA validation, who defined it as the comparison of the system's behavior with and without enforcement of the identified constraints on the behavior of each controller. This test can be easily used for systems that are in the operation phase; however, if STPA is implemented for a system in early concept design, for which there exists no real system, a simulation model of that system can be used for performing the comparison (Harkleroad et al., 2013). The suggested questions for the behavior validation test are:

- How does the enforcement of the identified constraints change the behavior of the system? (Harkleroad et al., 2013)
- Are the changes in the system's behavior as expected by the STPA implementation team?

5.4. Loss scenarios

5.4.1. Summary of the STPA step

In this STPA step, the loss scenarios, which describe the reasons why UCAs might take place in the system, are determined (Leveson & Thomas, 2018). For instance, scenarios are developed to explain how unsafe controller behavior and inadequate feedback and information can lead to UCAs. In addition to hazards that can occur through UCAs, hazards can also be caused by not executing or improperly executing a control action. Therefore, all these loss scenarios are investigated and elaborated.

5.4.2. Validation tests

5.4.2.1. Face validation. Face validation is proposed as one of the validity tests for UCAs (Section 4.3). This can also be used as a first validity test of the identified loss scenarios. Through this test, the validation team examines the comprehensiveness and accuracy of loss scenarios. Thus, the validation team reviews the identified loss scenarios, answering the below suggested questions:

- Are the identified loss scenarios logical and accurate from the validation team's perspective? (Sargent, 2013)
- Are all possible causal factors accounted for when identifying loss scenarios associated with the UCAs?

5.4.2.2. Historical validation. In modeling-oriented disciplines (Section 2.4), if historical data exists, the dataset is split into two datasets for building and testing the model, so it can be determined whether the developed model behaves as the system did according to historic data (Landry et al., 1983; Sargent, 2013). However, as mentioned earlier, STPA does not rely on numerical input data, but the conceptual idea of the validation test can be adapted. Through this validation test, the historical data, such as accident/incident data, of the studied system or identical or similar systems identified during nomological validation, is reviewed to ensure that the associated contributing factors are covered in the identified scenarios. Historical data would not be helpful if the studied system is considerably different from other existing systems. The suggested questions for historical validation are:

- Are the contributing factors of the previous incidents/accidents of the studied system covered in the identified scenarios? (Sargent, 2013)
- Are the contributing factors of the previous incidents/accidents of identical (sub-) systems identified in the nomological validation covered in the loss scenarios?
- Are the contributing factors of the previous incidents/accidents of the similar (sub-)systems identified in the nomological validation covered in the loss scenarios?

5.4.2.3. Traceability validation. Traceability should be maintained between various STPA outputs so changes can be made to the system without redoing the whole analysis. Every scenario must be possible to trace to one or more UCAs and each UCA must be traceable to one or more system-level hazards (Leveson & Thomas, 2018). A hazard can lead to one or more losses and each hazard should be possible to trace to the resulting losses. The traceability need not be a one-to-one relationship: a single constraint might be used to prevent more than one hazard, multiple constraints may be related to a single hazard, and each hazard could lead to one or more losses. As the last step of the loss scenarios validation, the validation team checks the traceability of all items, and makes sure that they are logical and unambiguously documented. Following questions can be used for this validation:

- Can the identified loss scenarios be traced to all relevant UCAs, hazards, and losses? (Leveson & Thomas, 2018)
- Can the identified losses in the first step of STPA be traced to all relevant hazards, UCAs, and scenarios? (Leveson & Thomas, 2018)
- Are the traceabilities properly documented?

6. Final results

In this proposed validation framework, a separate section is dedicated to the final results of the STPA, which consists of reviewing the final documentation and presentation of the results. This is important, because STPA analyses are used by design and/or operational teams to mitigate risks in design-focused work or to support operational risk management. The documentation is the responsibility of the STPA implementation team, and each step is documented once the associated step is completed. The completed documentation is handed over to the validation team for review. In addition to the documentation, a presentation for the stakeholders is prepared by the implementation team and handed over to the validation team along with other documents for review.

6.1. Validation tests

Unlike the previous validation steps, which are mainly concerned with the comprehensiveness and accuracy of STPA, the following validation steps are more concerned with the credibility of the analysis. Credibility refers to the extent to which stakeholders and decision-makers can trust and use the results of an analysis (Sargent, 2013). Busby and Hughes (2006) define credibility as the general notion of whether a risk assessment can be believed and trusted. Credibility is not just a factor of the analysis itself, but also of contextual factors which relates to the concept of trust in risk management. This relates to a wide array of social and perceptual factors beyond the quality of a risk assessment as such (Aven & Renn, 2010). Appropriate risk communication is critical for trust in risk management, in which the credibility of the analysis plays an important role. Therefore, validating the documentation and presentation of the STPA results are proposed in this section to support establishing this credibility.

6.2. Documentation checking

Documentation checking is conducted to ensure it is accurate, complete, and up-to-date (Balci, 1994). Transparency and sufficiency of documentation for a stakeholder review, where the stakeholders may not be familiar with the analysis technique, are also other aspects of documentation checking (Eddy et al., 2012; Lathrop & Ezell, 2017). This validation step also aligns with the idea of quality assurance which means examining the process behind the analysis (Goerlandt et al., 2017a). In the previous validation steps, the focus of the validation team was on a single step. In this step, however, the overall process is reviewed to ensure that it makes sense overall. Therefore, the validation team reviews the finalized documents and answers the following questions:

- Is the overall process behind the STPA implementation reasonably documented? (Sargent, 2013)
- Is the STPA documentation correct, clear, and complete? (Leveson & Thomas, 2018)
- Is the documentation in formats understandable to users and stakeholders who may not be knowledgeable about STPA? (Lathrop & Ezell, 2017)
- Are the sources of uncertainty clearly documented?
- Are the limitations of the analysis clearly documented? Can the limitations be justified with regard to the purpose of the analysis (Vergison, 1996)?

6.3. Review of the presentation of results

Communicating the analysis and its validation results are important for building the stakeholder's understanding of the complete analysis (Coyle & Exelby, 2000). Stakeholders are typically concerned about whether their interests are adequately considered in the analysis (Lathrop & Ezell, 2017). When presenting the results to stakeholders, how and where those identified interests are addressed in the analysis should be communicated to stakeholders, which aims to support the analysis being considered credible. The suggested questions are:

- Does the presentation include the appropriate information regarding where and how stakeholders' interests (identified in the problem situation) are included in the analysis? (Balci, 1994)
- Does the presentation clearly explain the sources of uncertainty? (Lathrop & Ezell, 2017)
- Does the presentation clearly explain the limitation of the analysis?

7. Discussion

As the developed STPA validation framework is rooted in theoretical concepts of scientific domains which can be considered to be closely related to hazard analysis, the authors believe that its proposal can contribute to enriching the literature on STPA and STAMP, and more generally the literature on validation of risk assessment. Its proposal follows earlier findings by Patriarca et al. (2022) and Sadeghi and Goerlandt (2022) that practically useful validation approaches are needed to further support the use of hazard analysis in industrial contexts. Furthermore, such frameworks can play an important role in collecting empirical evidence for the effectiveness of hazard and risk analysis techniques, which is scarce (Aven, 2012; Goerlandt et al., 2017a; Sadeghi & Goerlandt, 2021) but required to enable evidence-based risk practices (Hale, 2014).

Nevertheless, it is acknowledged that the proposed framework is an idealized process and that it builds on certain principles and commitments to foundational concepts in risk analysis, such as taking a constructivist stance in the realist-constructivist debate in risk research (Bradbury, 1989), understanding the risk concept through a close connection to uncertainty (Aven & Renn, 2009), and preferring an approach where analysis implementation and validation are performed in parallel by separate teams (Landry et al., 1983). While there appears to be a growing consensus on such risk-foundational issues in the wider risk research community (Aven, 2012; Aven & Zio, 2014), a healthy skepticism about the correctness and adequacy of such assumptions is warranted, especially given the wide range of problem domains STPA is targeted to be used in (Leveson & Thomas, 2018).

Validation is not often performed in academic work on model-based safety analysis (Sadeghi & Goerlandt, 2021); however, there is evidence to suggest that a desire to adopt validation exists among system safety practitioners (Sadeghi and Goerlandt, 2023). Landry et al. (1983) pointed to the need to determine the appropriate level of validation since a high level of validation will increase the associated costs. The proposed framework can be tailored to the available resources. For instance, if the implementation and validation are performed in parallel (see Section 3.2.1), and if contextual limitations do not enable the STPA implementation and validation teams to reach a high level of agreement in each step, it can be decided to stop validation and continue the next step of the STPA implementation. However, the low level of agreement is then reported, which is in line with uncertainty-based risk perspectives, to enable responsible consideration of the strength of evidence in the managerial decision-making stage.

The results of STPA can be used for different purposes, for instance for developing the system architecture, creating requirements, and creating test plans (Leveson & Thomas, 2018). The proposed validation framework does not consider any particular purpose for the STPA results. Its main intention is to propose a general structure for validation of

STPA, using a set of theory-based guide questions. The framework may need to be further elaborated if it is used for different purposes or aims of STPA, and guide questions may be tailored to the specific context of the analysis and additional questions formulated. Future research can also be directed towards further specifying aspects of the proposed framework, for instance proposing methods to construct a nomological map to compare STPA with hazard analyses of identical or similar systems or developing methods to assess the criticality and effects of assumptions.

The proposed validation framework aims to evaluate the comprehensiveness, accuracy, and credibility of STPA (Section 1), which are called the functions of the framework throughout this article. Each function can be linked to different types of safety work. According to Rae & Provan (2019), safety work consists of activities conducted in the name of safety, and is divided into four aspects: social, demonstrated, administrative, and physical safety work. The comprehensiveness and accuracy functions are at the core of the proposed STPA validation framework. These primarily aim to support safety-related decisions (a type of administrative safety work), and ultimately lead to physical safety work (i.e., operational work which would not occur if not for safety concerns). The credibility function is primarily concerned with demonstrated safety (showing safety to stakeholders) as it deals with ensuring that stakeholders can trust the results of an analysis. Out of all these activities, the safety of work, which means the absence of harm arising from operational work, can emerge (Rae & Provan, 2019). However, whether using the proposed STPA validation framework can indeed result in improvements to a STPA analysis, and to enhanced system safety or lower system risk, and under which conditions, need to be assessed. This is however beyond the scope of the current work.

The authors believe it to be a plausible hypothesis that the proposed formalized validation framework can indeed improve an STPA analysis, and the related issue of improving the safety of the system. To ensure that this framework can indeed achieve its direct envisioned goals in line with the above-mentioned functions, the proposed validation framework should be examined and tested. That is, it should be empirically investigated (e.g. through comparative case study research) whether the application of this framework indeed improves the comprehensiveness, accuracy, and credibility of an STPA analysis, and if so, under which conditions. Before engaging in such work, it is prudent to perform explicit research on the reasonableness of the proposed framework itself, i.e., triangulation research. The framework is developed based on theoretical notions of validation and tests rooted in the academic literature. It can be tested how far the framework aligns with what experts involved in reviewing STPA would do in absence of this framework, for instance through interview or case study research. This could confirm these theory-based tests or serve as a basis for further modifying the proposed framework.

As mentioned in Section 2, to develop the proposed theory-based validation framework, validation practices in fields that can be argued to be relevant to the practice of hazard analysis were selected and analyzed. The selected fields in this study are risk science, social science, operations research, system dynamics, and simulation modeling. It is plausible that other fields of research can also form a fruitful basis for developing an STPA validation framework. Thus, another research direction could be investigating and choosing other approaches to validation and developing a validation framework based on those fields. One possible field of study could be process evaluation which looks more into how a methodology is implemented in an organization, from a process point of view (Yin, 2017). The proposed validation framework in this study and the proposed tests focus mainly on what the STPA analysis looks like, rather than on how it is used in an organization. Such work could form a fruitful basis for scrutinizing and amending and/or extending the proposed validation framework through comparative research, which can point to gaps and limitations of the here proposed theoretical framework.

Building on knowledge on human performance in time-critical work, particularly that errors are more likely if the available time decreases

(Hall et al., 1982), it could be investigated how much resources should be allocated to validation of STPA analyses in relation to how much the analysis is improved. Such questions about the cost-effectiveness of validation as a type of safety work is a significant question for industry practitioners, as they report often facing challenges to convince clients or managers of the value of and need for validation, which requires organizational resources (Sadeghi and Goerlandt, 2023). Of course, this question can be extended in principle to how effective validation is from the perspective of the emergent safety of work, although this is a major conceptual and empirical challenge.

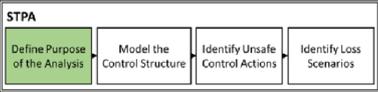
Similar to STPA implementation itself, the result of using this framework relies on the practitioners who use it. This can culminate in two issues: (1) the reliability of the results and (2) the validation team's bias affecting the outcome of the validation framework. In terms of the first issue, there is research suggesting that the validation of risk assessment can indeed be unreliable. For instance, Fabbri and Contini (2009) compared the results of peer reviews of different evaluators of a risk assessment, indeed finding a large variation. Even with a formalized framework, different experts may apply that framework differently, leading to variations in the validation outcome. Thus, a future research direction could be to investigate how much the use of this framework improves STPA depending on the characteristics of the practitioners applying it, e.g., in terms of their experience with using STPA or related hazard analysis techniques.

The validation team's potential biases can also have a negative impact on the validation results. For example, if the findings of STPA are new and are not in line with the validation team's knowledge and experience, disagreement between the implementation and validation teams may arise. This can be a common issue in STPA validation as earlier research has shown that STPA finds items that cannot be identified using traditional techniques (Arnold, 2009; Martínez, 2015). In a study by Bugalia et al. (2022), the findings of the STPA analysis were considered inappropriate by the stakeholders as they have never happened before. Thus, the choice of the validation team, for instance whether they are open to new ideas, can play a major role in the credibility the validation team gives to the results of STPA, and hence the interaction between these two groups. In addition, the STPA implementation team should be mindful while considering the validation team's comments. When disagreements arise, it is prudent to communicate them openly to decision makers and stakeholders, reflecting on the possible biases and disagreements arising because of the composition of, and relations between, the two teams. Thus, a future research direction would be investigating how the STPA validation team would use the framework in practice, and how the interaction between the two teams occurs.

The STPA technique can be used for different stages of a system's lifecycle (Leveson & Thomas, 2018), e.g. design and operational contexts. The generic STPA implementation steps for such different lifecycle stages are the same. Likewise, it is assumed that the proposed STPA validation framework can be used for different system lifecycle stages, with the same validation tests and guide questions. However, one of the future research directions would be testing this assumption. That is, the application of this framework for different stages should be tested to see if the tests and guide questions are applicable and useful for all stages. If not, the proposed framework needs to be more elaborated and tailored for each stage. In this sense, a possibly fruitful direction for future research can also be to understand how the validation of other STAMP-based tools, such as Systems Theoretic Early Concept Analysis (STECA) (Fleming, 2015), has been conceptualized and applied, and how such knowledge may be used to modify the here presented validation framework.

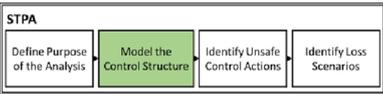
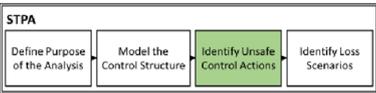
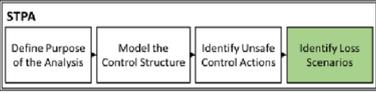
Finally, it also appears an important direction for future research on how (perhaps inappropriate) application of the proposed formal validation framework may affect the occurrence of probative blindness, i.e. the false assurance of the safety of a design or activity, where this assurance is not aligned with reality (Rae & Alexander, 2017b). Amongst

Table 1
The summary of the proposed STPA validation framework.

STPA Step	Main Activities	Validation Function	Validation Tests	Related Questions
1. <u>Define Purpose of the Analysis</u>	 <ol style="list-style-type: none"> 1. Identifying losses 2. Identifying system-level hazards <ol style="list-style-type: none"> a. Identify the system to be analyzed b. Identify the system boundary c. Identifying system states or conditions that will lead to a loss in worst-case environmental conditions 3. Identifying system-level constraints 4. Refine hazards 	Comprehensiveness and Accuracy	Nomological Validation	<ol style="list-style-type: none"> 1. Is there a similar/identical analysis in the existing literature or industry to support the result of this analysis? 2. If yes, which identified analyses are nomologically adjacent and in which way, and which ones are distant, to the system of interest and its associated STPA? 3. If not, have the STPA implementation team clearly explained why this analysis lies outside all current known research? 4. If there exist similar/identical analyses, how similar are the identified losses, system-level hazards, and system-level constraints?
			System Boundary Validation	<ol style="list-style-type: none"> 1. Is the system boundary explicitly defined and documented? 2. Is the system boundary in line with the purpose of the analysis? 3. Do the system designers or operators have control over all the identified elements included within the system boundary? 4. How would modifying the system boundary change the identified losses, system-level hazards, and system-level constraints? 5. How would changes in the environmental context of the system affect the need for adjusting the system boundary to meet the purpose of the analysis?
			Data Validation	<ol style="list-style-type: none"> 1. What are the sources of input data? 2. How reliable are the instruments (e.g., software) and processes (e.g., surveys) used for data collection and measurement (e.g., accident data)? 3. Are all sources of input data, including but not limited to design documents, industry-related standards, and historical data, up to date? (e.g., do they use the latest version of the design documents, or technical standards?)
			Subject Matter Experts (SMEs) Validation	<ol style="list-style-type: none"> 1. Is the process of SMEs selection, elicitation, and combination clearly and completely documented? 2. Is the SMEs selection systematically conducted? Are the SMEs selection criteria reasonable considering the purpose of the analysis? Is the entire system covered with appropriate knowledgeable experts? 3. Is the SMEs elicitation process systematically conducted? Is this process reasonable considering the purpose of the analysis? 4. If more than one SME is involved in the analysis, are the results of SME elicitations combined in a meaningful way?
			Assumption Validation	<ol style="list-style-type: none"> 1. Are the assumptions fully identified, accurately described, understood, documented, and agreed upon? 2. If the opposite of any particular assumption were true, would it have a substantial effect on the identified losses, system-level hazards, and system-levels constraint? 3. Are the degree of importance and certainty of each relevant assumption credibly determined?
2. Model the Control Structure	1. Controller			

(continued on next page)

Table 1 (continued)

STPA Step	Main Activities	Control Algorithm	Validation Function	Validation Tests	Related Questions
	b. Process Model 2. Control Actions 3. Feedback 4. Controlled Processes		Comprehensiveness and Accuracy	Content and structure Validation	1. Does the created control structure include all relevant elements and system components? 2. Does the created control structure include all relevant functional relationships between these elements? 3. Is the control structure an accurate representation of the system? 4. Does the level of detail included in the control structure suffice for the purpose of the analysis?
				Concurrent Validation	1. Has any STPA for an identical system been identified in the nomological validation? 2. If yes, is the developed control structure, including the controllers, controlled processes, control actions, and feedbacks, the same as the control structure in the identified STPA? If there are differences, why do these appear?
				Convergent Validation	1. Has any STPA for a similar system been identified when conducting nomological validation? 2. If yes, how similar is the control structure to other control structures in the analysis of similar systems? If there are differences, why do these appear?
3. <u>Identify Unsafe Control Actions</u>		1. List of Unsafe Control Actions and Casual Factors 2. Controller Constraints A controller constraint specifies the controller behaviors that need to be satisfied to prevent UCAs.	Comprehensiveness and Accuracy	Face Validation	1. Are the identified UCAs logical and accurate from the validation team's perspective? 2. Are there any other possible UCAs that have not been identified by the STPA implementation team? 3. Are the identified UCAs accurately translated into constraints on the behavior of each controller?
Extreme Condition Validation				1. Are the identified constraints plausible for extreme and unlikely interactions of components within the system? 2. Are the identified constraints plausible for extreme and unlikely conditions in the system's environment?	
Behavior Validation				1. How does the enforcement of the identified constraints change the behavior of the system? 2. Are the changes in the system's behavior as expected by the STPA implementation team?	
4. <u>Identify Loss Scenarios</u>		1. Identify scenarios lead to UCA	Comprehensiveness and Accuracy	Face Validation	1. Are the identified loss scenarios logical and accurate from the validation team's perspective? 2. Are all possible causal factors accounted for when identifying loss scenarios associated with the UCAs?
Historical Validation				1. Are the contributing factors of the previous incidents/accidents of the studied system covered in the identified scenarios? 2. Are the contributing factors of the previous incidents/accidents of identical (sub-)systems identified in the nomological validation covered in the loss scenarios? 3. Are the contributing factors of the previous incidents/accidents of the similar (sub-)systems identified in the nomological validation covered in the loss scenarios?	
Traceability Validation				1. Can the identified loss scenarios be traced to all relevant UCAs, hazards, and losses?	

(continued on next page)

- Arnold, R., 2009. A qualitative comparative analysis of SOAM and STAMP in ATM occurrence investigation. Lund University.
- Aven, T., 2012. Foundational issues in risk assessment and risk management. *Risk Anal.* 32 (10), 1647–1656. <https://doi.org/10.1111/j.1539-6924.2012.01798.x>.
- Aven, T., 2017. Risk analysis validation and trust in risk management: a postscript. *Saf. Sci.* 99, 255–256. <https://doi.org/10.1016/j.ssci.2017.08.009>.
- Aven, T., Ben-Haim, Y., Andersen, H.B., Cox, T., Drogue, E.L., Greenberg, M., Guikema, S., Kröger, W., Renn, O., Thompson, K.M., Zio, E., 2018. *Soc. Risk Anal. Glossary* 9.
- Aven, T., Guikema, S., 2011. Whose uncertainty assessments (probability distributions) does a risk assessment report: the analysts' or the experts'? *Reliab. Eng. Syst. Saf.* 96 (10), 1257–1262. <https://doi.org/10.1016/j.res.2011.05.001>.
- Aven, T., Heide, B., 2009. Reliability and validity of risk analysis. *Reliab. Eng. Syst. Saf.* 94 (11), 1862–1868. <https://doi.org/10.1016/j.res.2009.06.003>.
- Aven, T., Renn, O., 2009. On risk defined as an event where the outcome is uncertain. *J. Risk Res.* 12 (1), 1–11. <https://doi.org/10.1080/13669870802488883>.
- Aven, T., Renn, O., 2010. *Risk Management and Governance: Concepts, Guidelines and Applications*. Springer Berlin / Heidelberg. <http://ebookcentral.proquest.com/lib/dal/detail.action?docID=645899>.
- Aven, T., Zio, E., 2014. Foundational Issues in Risk Assessment and Risk Management. *Risk Anal.* 34 (7), 1164–1172. <https://doi.org/10.1111/risa.12132>.
- Balci, O., 1994. Validation, verification, and testing techniques throughout the life cycle of a simulation study. 53, 121–173.
- Barlas, S., 1996. Formal aspects of model validity and validation in system dynamics. *Syst. Dyn. Rev.* 12 (3), 183–210. [https://doi.org/10.1002/\(SICI\)1099-1727\(199623\)12:3<183::AID-SDR103>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1099-1727(199623)12:3<183::AID-SDR103>3.0.CO;2-4).
- Baybutt, P., 2021. On the need for system-theoretic hazard analysis in the process industries. *J. Loss Prev. Process Ind.* 69, 104356. <https://doi.org/10.1016/j.jlp.2020.104356>.
- Bjerga, T., Aven, T., Zio, E., 2016. Uncertainty treatment in risk analysis of complex systems: the cases of STAMP and FRAM. *Reliab. Eng. Syst. Saf.* 156, 203–209. <https://doi.org/10.1016/j.res.2016.08.004>.
- Bollen, K.A., 1989. *Structural equations with latent variables* (pp. xiv, 514). John Wiley & Sons. <https://doi.org/10.1002/9781118619179>.
- Boring, R. L., Gertman, D., Joe, J., Marble, J., Galyean, W., Blackwood, L., & Blackman, H., 2005. Simplified Expert Elicitation Procedure for Risk Assessment of Operating Events (INL/EXT-05-00433). Idaho National Lab. (INL), Idaho Falls, ID (United States). <https://doi.org/10.2172/911228>.
- Bradbury, J.A., 1989. The policy implications of differing concepts of risk. *Sci. Technol. Hum. Values* 14 (4), 380–399.
- Brummett, B., 2019. *Techniques of Close Reading*. SAGE Publications, Inc. <https://doi.org/10.4135/9781071802595>.
- Bugalia, N., Choudhury, S.R., Maemura, Y., Seetharam, K., 2022. A systems theoretic process analysis (STPA) approach for analyzing the governance structure of fecal sludge management in Japan. *Environ. Plan. B: Urban Anal. City Sci.* 49 (8), 2168–2194. <https://doi.org/10.1177/23998083221075639>.
- Busby, J.S., Hughes, E.J., 2006. Credibility in risk assessment: a normative approach. *Int. J. Risk Assess. Manage.* 6(4–6), 508–527. <https://doi.org/10.1504/IJRAM.2006.009542>.
- Clemen, R.T., Winkler, R.L., 1999. Combining probability distributions from experts in risk analysis. *Risk Anal.* 19 (2), 187–203. <https://doi.org/10.1111/j.1539-6924.1999.tb00399.x>.
- Collier, Z.A., Lambert, J.H., 2019. Principles and methods of model validation for model risk reduction. *Environ. Syst. Decis.* 39 (2), 146–153. <https://doi.org/10.1007/s10669-019-09723-5>.
- Cooke, R.M., Goossens, L.H.J., 2004. Expert judgement elicitation for risk assessments of critical infrastructures. *J. Risk Res.* 7 (6), 643–656.
- Cooke, R., 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. New York : Oxford University Press. https://web-p-ebcsohost-com.ezproxy.library.dal.ca/ehost/ebookviewer/ebook/ZTAwMHhuYV9fMjg4NTIyX19BTg2?sid=be80e70f-e2ba-4142-b552-1438b760db7e@redis&vid=0&format=EB&lpid=lp_169&rid=0.
- Coyle, G., Exelby, D., 2000. The validation of commercial system dynamics models. *Syst. Dyn. Rev.* 16 (1), 27–41. [https://doi.org/10.1002/\(SICI\)1099-1727\(200021\)16:1<27::AID-SDR182>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-1727(200021)16:1<27::AID-SDR182>3.0.CO;2-1).
- Dakwat, A.L., Villani, E., 2018. System safety assessment based on STPA and model checking. *Saf. Sci.* 109, 130–143. <https://doi.org/10.1016/j.ssci.2018.05.009>.
- Dallat, C., Salmon, P.M., Goode, N., 2019. Risky systems versus risky people: to what extent do risk assessment methods consider the systems approach to accident causation? A review of the literature. *Saf. Sci.* 119, 266–279. <https://doi.org/10.1016/j.ssci.2017.03.012>.
- Dekker, S., 2019. *Foundations of safety science: A century of understanding accidents and disasters*. CRC Press, Taylor & Francis Group.
- Drost, E.A., 2011. Validity and reliability in social science research. *Educ. Res. Perspect.* 38 (1), 105–124.
- Dulac, N., 2007. *A Framework for Dynamic Safety and Risk Management Modeling in Complex Engineering Systems* [Massachusetts Institute of Technology]. <http://sunnyday.mit.edu/safer-world/dulac-dissertation.pdf>.
- Eckerd, A., Landsbergen, D., Desai, A., 2011. The Validity Tests Used by Social Scientists and Decision Makers. 14.
- Eddy, D.M., Hollingworth, W., Caro, J.J., Tsevat, J., McDonald, K.M., Wong, J.B., 2012. Model transparency and validation: a report of the ISPOR-SMDM modeling good research practices task force-7. *Med. Decis. Making* 32 (5), 733–743. <https://doi.org/10.1177/0272989X12454579>.
- Eker, S., Rovenskaya, E., Langan, S., Obersteiner, M., 2019. Model validation: a bibliometric analysis of the literature. *Environ. Model. Softw.* 117, 43–54. <https://doi.org/10.1016/j.envsoft.2019.03.009>.
- Engel, A., 2010. *Verification, Validation, and Testing of Engineered Systems*. Wiley. <http://ezproxy.library.dal.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=e000xna&AN=329992&site=ehost-live>.
- Ericson, C., 2005. *Hazard Analysis Techniques for System Safety* (1. Aufl.). Wiley-Interscience.
- Fabbri, L., Contini, S., 2009. Benchmarking on the evaluation of major accident-related risk assessment. *J. Hazard. Mater.* 162 (2), 1465–1476. <https://doi.org/10.1016/j.jhazmat.2008.06.071>.
- Finlay, P.N., Wilson, J.M., 1987. The paucity of model validation in operational research projects. *J. Oper. Res. Soc.* 38 (4), 303–308. <https://doi.org/10.2307/2582053>.
- Flage, R., Aven, T., 2009. *Expressing and communicating uncertainty in relation to quantitative risk analysis (QRA). Theory & Application, Reliability & Risk Analysis*, p. 132.
- Fleming, C.H., 2015. *Safety-driven Early Concept Analysis and Development*. Massachusetts Institute of Technology.
- Fleming, C.H., Spencer, M., Thomas, J., Leveson, N., Wilkinson, C., 2013. Safety assurance in NextGen and complex transportation systems. *Saf. Sci.* 55, 173–187. <https://doi.org/10.1016/j.ssci.2012.12.005>.
- Forrester, J., Senge, P., 1980. Tests for building confidence in system dynamics models (pp. 209–228).
- Gass, S.I., 1983. Decision-aiding models: validation, assessment, and related issues for policy analysis. *Oper. Res.* 31 (4), 603–631.
- Goerlandt, F., Khakzad, N., Reniers, G., 2017a. Validity and validation of safety-related quantitative risk analysis: a review. *Saf. Sci.* 99, 127–139. <https://doi.org/10.1016/j.ssci.2016.08.023>.
- Goerlandt, F., Khakzad, N., Reniers, G., 2017b. Special issue: risk analysis validation and trust in risk management. *Saf. Sci.* 99, 123–126.
- Goerlandt, F., Montewka, J., 2015. A framework for risk analysis of maritime transportation systems: a case study for oil spill from tankers in a ship-ship collision. *Saf. Sci.* 76, 42–66. <https://doi.org/10.1016/j.ssci.2015.02.009>.
- Goerlandt, F., Reniers, G., 2016. On the assessment of uncertainty in risk diagrams. *Saf. Sci.* 84, 67–77. <https://doi.org/10.1016/j.ssci.2015.12.001>.
- Groesser, S.N., Schwanager, M., 2012. Contributions to model validation: Hierarchy, process, and cessation. *Syst. Dyn. Rev.* 28 (2), 157–181. <https://doi.org/10.1002/sdr.1466>.
- Habli, I., Alexander, R., Hawkins, R., 2021. Safety Cases: An Impending Crisis? 18.
- Hale, A., 2014. Foundations of safety science: a postscript. *Saf. Sci.* 67, 64–69. <https://doi.org/10.1016/j.ssci.2014.03.001>.
- Hall, R.E., Fragola, J., Wreathall, J., 1982. Post-event human decision errors: operator action tree/time reliability correlation (p. 48).
- Harkleroad, E., Vela, A., Kuchar, J., 2013. Review of Systems-Theoretic Process Analysis (STPA) Method and Results to Support NextGen Concept Assessment and Validation (ATC-427).
- Hulme, A., Stanton, N.A., Walker, G.H., Waterson, P., Salmon, P.M., 2022. Testing the reliability and validity of risk assessment methods in Human Factors and Ergonomics. *Ergonomics* 65 (3), 407–428. <https://doi.org/10.1080/00140139.2021.1962969>.
- Kaplan, S., 1992. 'Expert information' versus 'expert opinions'. Another approach to the problem of eliciting/ combining/using expert knowledge in PRA. *Reliab. Eng. Syst. Saf.* 35 (1), 61–72. [https://doi.org/10.1016/0951-8320\(92\)90023-E](https://doi.org/10.1016/0951-8320(92)90023-E).
- Keys, P., 1988. System dynamics: A methodological perspective. *Trans. Inst. Meas. Control* 10 (4), 218–224. <https://doi.org/10.1177/014233128801000406>.
- Landry, M., Malouin, J.-L., Oral, M., 1983. Model validation in operations research. *Eur. J. Oper. Res.* 14 (3), 207–220. [https://doi.org/10.1016/0377-2217\(83\)90257-6](https://doi.org/10.1016/0377-2217(83)90257-6).
- Lathrop, J., Ezell, B., 2017. A systems approach to risk analysis validation for risk management. *Saf. Sci.* 99, 187–195. <https://doi.org/10.1016/j.ssci.2017.04.006>.
- Law, A., 2014. *Simulation Modeling and Analysis* (5th edition). McGraw Hill.
- Leveson, N., 2004a. A systems-theoretic approach to safety in software-intensive systems. *IEEE Trans. Dependable Secure Comput.* 1 (1), 66–86. <https://doi.org/10.1109/TDSC.2004.1>.
- Leveson, N., 2004b. A new accident model for engineering safer systems. *Saf. Sci.* 42 (4), 237–270. [https://doi.org/10.1016/S0925-7535\(03\)00047-X](https://doi.org/10.1016/S0925-7535(03)00047-X).
- Leveson, N., 2012. *Engineering a Safer World: Systems Thinking Applied to Safety*. Cambridge, Mass : The MIT Press. https://web-p-ebcsohost-com.ezproxy.library.dal.ca/ehost/ebookviewer/ebook/ZTAwMHhuYV9fNDIxODE4X19BTg2?sid=e9969089-f149-426b-bb6e-776f0eca0b81@redis&vid=0&format=EB&lpid=lp_1&rid=0.
- Leveson, N., 2017. Rasmussen's legacy: A paradigm change in engineering for safety. *Appl. Ergon.* 59, 581–591. <https://doi.org/10.1016/j.apergo.2016.01.015>.
- Leveson, N., Thomas, J., 2018. *STPA Handbook*. https://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf.
- Lordos, G.C., Summers, S.E., Hoffman, J.A., De Weck, O.L., 2019. Human-machine interactions in apollo and lessons learned for living off the land on mars. *IEEE Aerospace Conference* 2019, 1–17. <https://doi.org/10.1109/AERO.2019.8741618>.
- Martínez, R.S., 2015. *System Theoretic Process Analysis of Electric Power Steering for Automotive Applications*. Massachusetts Institute of Technology.
- Mason, R., Mitroff, I., 1981. Challenging strategic planning assumptions: Theory, cases, and techniques. Wiley.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, 339, b2535. <https://doi.org/10.1136/bmj.b2535>.
- Norton, M.I., Mochon, D., Ariely, D., 2012. The IKEA effect: when labor leads to love. *J. Consum. Psychol.* 22 (3), 453–460. <https://doi.org/10.1016/j.jcps.2011.08.002>.

- Oral, M., Kettani, O., 1993. The facets of the modeling and validation process in operations research. *Eur. J. Oper. Res.* 66 (2), 216–234. [https://doi.org/10.1016/0377-2217\(93\)90314-D](https://doi.org/10.1016/0377-2217(93)90314-D).
- Patriarca, R., Chatzimichailidou, M., Karanikas, N., Di Gravio, G., 2022. The past and present of system-theoretic accident model and processes (STAMP) and its associated techniques: a scoping review. *Saf. Sci.* 146, 105566 <https://doi.org/10.1016/j.ssci.2021.105566>.
- Pitchforth, J., Mengersen, K., 2013. A proposed validation framework for expert elicited Bayesian networks. *Expert Syst. Appl.* 40 (1), 162–167.
- Rae, A., Alexander, R., 2017a. Forecasts or fortune-telling: when are expert judgements of safety risk valid? *Saf. Sci.* 99, 156–165.
- Rae, A., Alexander, R.D., 2017b. Probative blindness and false assurance about safety. *Saf. Sci.* 92, 190–204. <https://doi.org/10.1016/j.ssci.2016.10.005>.
- Rae, A., Provan, D., 2019. Safety work versus the safety of work. *Saf. Sci.* 111, 119–127. <https://doi.org/10.1016/j.ssci.2018.07.001>.
- Rasmussen, J., 1997. Risk management in a dynamic society: a modelling problem. *Saf. Sci.* 27 (2), 183–213. [https://doi.org/10.1016/S0925-7535\(97\)00052-0](https://doi.org/10.1016/S0925-7535(97)00052-0).
- Redmill, F., 2002. Risk analysis—a subjective process. *Eng. Manag. J.* 12 (2), 91. <https://doi.org/10.1049/em:20020206>.
- Rosa, E.A., 1998. Metatheoretical foundations for post-normal risk. *J. Risk Res.* 1 (1), 15–44. <https://doi.org/10.1080/136698798377303>.
- Rosqvist, T., 2010. On the validation of risk analysis—a commentary. *Reliab. Eng. Syst. Saf.* 95 (11), 1261–1265. <https://doi.org/10.1016/j.res.2010.06.002>.
- Sadeghi, R., Goerlandt, F., 2021. The state of the practice in validation of model-based safety analysis in socio-technical systems: an empirical study. *Safety* 7 (4), Article 4. <https://doi.org/10.3390/safety7040072>.
- Science, S., 2017. Risk analysis validation and trust in risk management. Part B 99, 123–256.
- Sadeghi, R., Goerlandt, F., 2023. Validation of system safety hazard analysis in safety-critical industries: an interview study with industry practitioners. *Saf. Sci.* 161, 106084. <https://doi.org/10.1016/j.ssci.2023.106084>.
- Sargent, R.G., 2013. Verification and validation of simulation models. *J. Simulat.* 7 (1), 12–24. <https://doi.org/10.1057/jos.2012.20>.
- Schwanitz, V.J., 2013. Evaluating integrated assessment models of global climate change. *Environ. Model. Softw.* 50, 120–131. <https://doi.org/10.1016/j.envsoft.2013.09.005>.
- Sulaman, S.M., Beer, A., Felderer, M., Höst, M., 2019. Comparison of the FMEA and STPA safety analysis methods – a case study. *Qual. J.* 27 (1), 349–387. <https://doi.org/10.1007/s11219-017-9396-0>.
- Thomas, J., de Lemos, F. L., Leveson, N., 2012. Evaluating the Safety of Digital Instrumentation and Control Systems in Nuclear Power Plants (Research Report NRC-HQ-11-6-04-0060; p. 66).
- Trochim, 2006. Introduction to Validity. Research Methods Knowledge Base. <https://conjointly.com/kb/introduction-to-validity/>.
- Trochim, W., Donnelly, J., Arora, K., 2015. Research Methods: The Essential Knowledge Base. In: ProtoView (Vol. 2, Issue 41). Ringgold Inc. <https://www.proquest.com/docview/1723086569/citation/AF664690EFB244EFPQ/1>.
- Valdez Banda, O.A., Goerlandt, F., Salokannel, J., van Gelder, P.H.A.J.M., 2019. An initial evaluation framework for the design and operational use of maritime STAMP-based safety management systems. *WMU J. Marit. Aff.* 18 (3), 451–476. <https://doi.org/10.1007/s13437-019-00180-0>.
- Vergison, E., 1996. A Quality-Assurance guide for the evaluation of mathematical models used to calculate the consequences of Major Hazards. *J. Hazard. Mater.* 49 (2), 281–297. [https://doi.org/10.1016/0304-3894\(96\)01746-3](https://doi.org/10.1016/0304-3894(96)01746-3).
- Wróbel, K., Montewka, J., Kujala, P., 2018. Towards the development of a system-theoretic model for safety assessment of autonomous merchant vessels. *Reliab. Eng. Syst. Saf.* 178, 209–224. <https://doi.org/10.1016/j.res.2018.05.019>.
- Yin, R.K., 2017. *Case Study Research and Applications: Design and Methods*. SAGE Publications.

Publication IV

Sadeghi, & Goerlandt, F. (2023). Reasonableness of a proposed System Theoretic Process Analysis (STPA) validation framework: An interview study. *Journal of Loss Prevention in the Process Industries*, 83, 105064–. <https://doi.org/10.1016/j.jlp.2023.105064>



Reasonableness of a proposed System Theoretic Process Analysis (STPA) validation framework: An interview study

Reyhaneh Sadeghi^{*}, Floris Goerlandt

Dalhousie University, Department of Industrial Engineering, Halifax, Nova Scotia, Canada

ARTICLE INFO

Keywords:

STPA
Validation
Reasonableness
Hazard analysis
STPA validation framework
Experts

ABSTRACT

Since its inception, the STPA technique has gained increasing popularity among researchers and industry practitioners. Nevertheless, the validity of its application has not yet received much scientific attention. Although some informal validation approaches have been used by STPA users, no formalized validation framework has been elaborated for practical use. This paper investigates the reasonableness of the recently proposed STPA validation framework, which includes 15 validation tests, each focusing on a specific step of an STPA analysis. To do so, STPA experts in both academia and industry were interviewed. First, it is investigated what approaches they have been using for validating an STPA analysis, the findings of which were categorized and mapped with the proposed validation framework. This aims to investigate the similarities and dissimilarities between the theory-based validation framework and the informal methods applied by experts in current practice. Then, the proposed framework was presented to the interviewees to seek their judgments about its reasonableness. Feedback from practitioners indicated that the proposed STPA validation framework has certain strengths, while several opportunities exist for further improvement. In particular, the findings indicate that most of the proposed theory-based tests have been already used by STPA experts in an unstructured manner. The experts appreciated the framework in that it provides clear guidance on how to validate each step of an STPA analysis systematically, and found some additional theory-based tests interesting for consideration in practice. The results also suggest that further research is needed to develop systematic techniques for performing each test to facilitate its application by STPA experts.

1. Introduction

System-theoretic Process Analysis (STPA) is a hazard analysis technique developed on Systems-Theoretic Accident Model and Processes (STAMP) as a theoretical foundation (Leveson, 2015). STPA includes the definition of accident scenarios covering design errors, such as software flaws, component interactions, and social, organizational, and management factors in the analysis, which cannot be equally well covered by traditional hazard analysis techniques based on linear accident causation theories (Dallat et al., 2019; Leveson, 2012). Since its inception in the early 2000s, STPA has been used in many industries for various applications, e.g. process industry (Baybutt, 2021; Sultana et al., 2019), maritime industry (Chaal et al., 2022; Ventikos et al., 2020; Wróbel et al., 2018), and aerospace industry (Fleming and Leveson, 2014). The use of STPA has shown promising results compared to traditional techniques, in terms of identifying more hazards (Arnold, 2009; Martínez, 2015).

In a scoping review, Patriarca et al. (2022) remarked that there has been an increase in the number of STPA publications in different industrial sectors, whereas in most of the reviewed articles, the validation of the STPA application and its results have not been discussed. Thus, the significant question is how the validity of an STPA application can be reasonably established. STPA experts have performed some form of validation, such as an unstructured expert review (Thomas et al., 2012). However, the lack of clear guidance has been raised by industry practitioners as an important factor making hazard analysis validation a challenging task, resulting in practitioners seeing significant value in developing a formal validation framework (Sadeghi and Goerlandt, 2023a).

In response to the lack of formalized framework for validating an STPA analysis, Sadeghi and Goerlandt (2023b) proposed a theory-based STPA validation framework, rooted in foundational concepts in risk analysis and prior theoretical work on validation in related disciplines, including risk science, social science, and operations research, system

^{*} Corresponding author.

E-mail addresses: reyhaneh.sadeghi@dal.ca (R. Sadeghi), floris.goerlandt@dal.ca (F. Goerlandt).

<https://doi.org/10.1016/j.jlp.2023.105064>

Received 10 February 2023; Received in revised form 24 March 2023; Accepted 15 April 2023

Available online 28 April 2023

0950-4230/© 2023 Elsevier Ltd. All rights reserved.

dynamics, and simulation modeling disciplines. This framework aims to support a systematic assessment of the STPA analysis's comprehensiveness, accuracy, and credibility. However, because the proposed framework has only been elaborated theoretically, it is important to perform explicit research addressing the reasonableness of the framework itself. Such work could confirm the proposed theory-based tests or serve as a basis for further modifying the proposed framework.

This article aims to evaluate the reasonableness of the proposed framework. Reasonableness is defined as the quality of being plausible or acceptable to a reasonable person. According to [van der Helm \(2006\)](#), being plausible is a subject-related notion and something can only be plausible when someone claims it to be. Consequently, reasonableness is determined through a reasonable individual's estimation. The aim is to investigate the extent to which the proposed framework is reasonable to a group of STPA users. This has two aspects: (i) whether the validation tests and ideas of the proposed framework are already used in practice, and (ii) whether certain theory-based validation tests, which are not yet used, could be applied in practice.

The remainder of this article is organized as follows. The summary of the proposed STPA validation framework by [Sadeghi and Goerlandt \(2023b\)](#) is outlined in Section 2. The research methodology and data collection and analysis are presented in Section 3. Section 4 presents the results. Section 5 provides a discussion of the findings of the interviews, addresses the limitations of the work, and highlights avenues for future work. Section 5 concludes.

2. Summary of the proposed STPA validation framework

This section provides a brief overview of the STPA validation framework proposed by [Sadeghi and Goerlandt \(2023b\)](#). Readers are advised to refer to that paper for more detailed information on the foundational concepts and the detailed definitions of each test, and the associated guide questions. In addition, the main aspects of the framework are illustrated in [Appendix B](#), for the different steps of an STPA analysis.

This framework draws on literature in risk science, social science, and operations research, system dynamics, and simulation modeling fields, to propose a formalized structure for validating an STPA analysis. A set of theory-based validation tests are proposed for the different STPA steps, which are further elaborated as guide questions (see [Appendix B](#)). These 15 validation tests can be either used in parallel with the STPA implementation ([Fig. 1](#)) or in a post-hoc manner ([Fig. 2](#)).

The proposed framework aims to assess the comprehensiveness, accuracy, and credibility of an STPA analysis through a judgment by an assessor to increase the intersubjective agreement among all parties involved in the analysis. Thus, it highlights the importance of having two independent teams, one in charge of STPA implementation and one in charge of STPA validation (the assessor(s)). The framework also does not take a binary reject/approve approach but rather aims to help peers and stakeholders reason about the analysis in a systematic manner, giving advice for improvement or further elaboration.

One assumption underlying this framework is that the decision on validation cessation cannot be simply reduced to some quantitative criteria, but that this decision should be made through a discussion between the STPA implementation and validation teams. In a parallel process ([Fig. 1](#)), this decision can happen after each step of validation, while in a post-hoc application ([Fig. 2](#)), it can happen once the validation is performed.

3. Method

3.1. Participant recruitment

For this research, semi-structured interviews were performed to seek STPA experts' judgments on the reasonableness of the proposed validation framework outlined in Section 2. The interview research methodology, which is a type of qualitative research, generates knowledge grounded in human experience ([Sandelowski, 2004](#)). This is deemed to be a suitable method for this research because, through interviews, in-depth knowledge of the actual validation practices among STPA

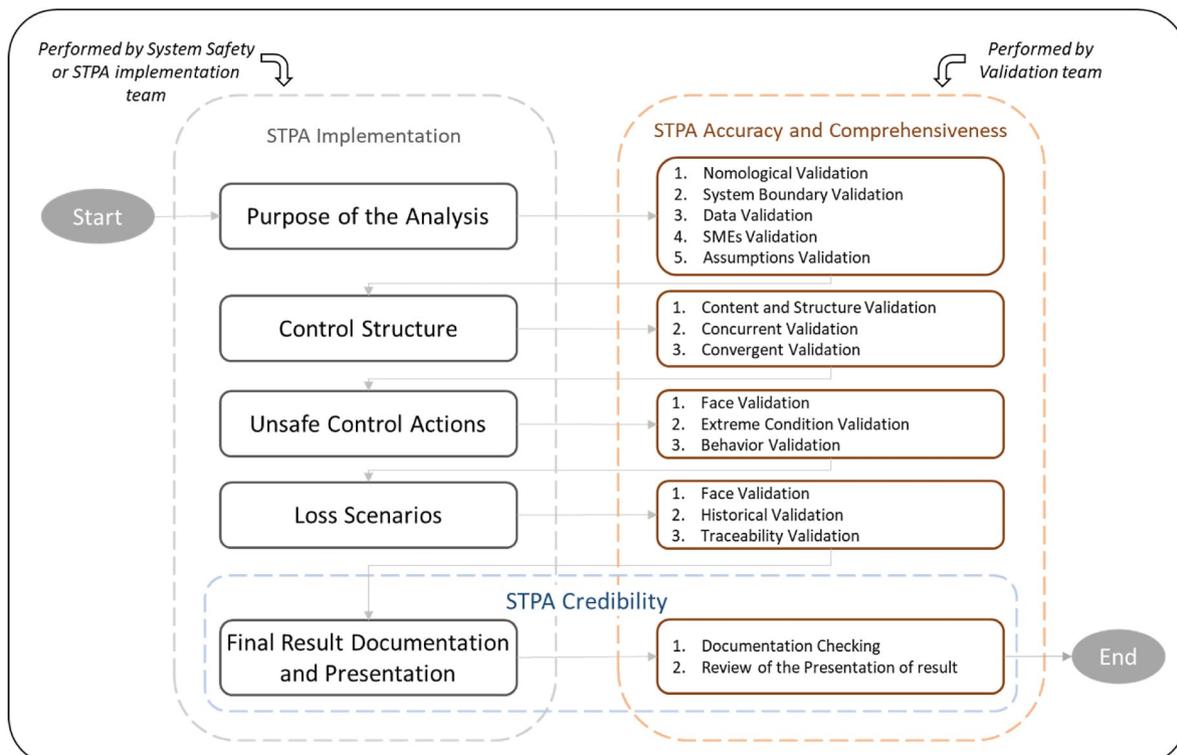


Fig. 1. Using the STPA validation framework in parallel with STPA implementation (adapted from [Sadeghi and Goerlandt, 2023b](#)).

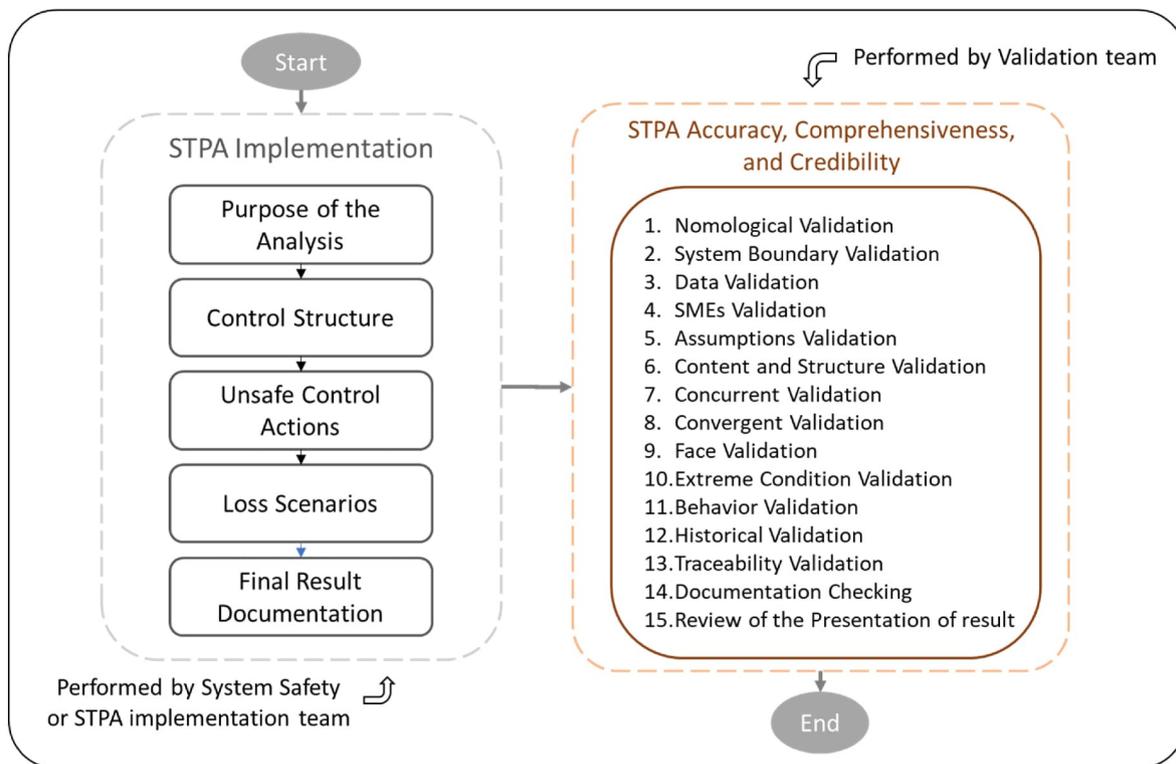


Fig. 2. Performing validation using the STPA validation framework after STPA implementation (adapted from Sadeghi and Goerlandt, 2023b).

experts can be gained.

For recruiting interviewees, the authors participated in the 10th European STAMP workshop and conference on September 29 and 30, 2022, and invited the participants of that workshop to take part in this study. The participants are researchers or industry practitioners with experience with various STAMP-related tools, mostly STPA, and are therefore considered a valid group to select participants from for obtaining insights into STPA validation. Four participants of this workshop were willing to take part in this research study. Furthermore, the first author participated in the 2022 Massachusetts Institute of Technology (MIT) STAMP workshop, which took place on June 6–10, 2022. Since the 2022 MIT STAMP workshop was held online, the authors could not reach out to all participants. However, some connections were made with STPA experts in attendance, to whom a request for participation was sent. A further five STPA experts were interested to participate in this research.

Finally, the recent scoping review on STAMP by Patriarca et al. (2022) was used to identify key academic authors active in STPA application and development. For this, the supplementary file provided along with the literature review by these authors is used, which contains a list of 321 documents. This list was reviewed and papers specifically focusing on STPA were selected. The authors of the selected papers were contacted and invited to participate in this research. Only three additional authors responded and were interested. A follow-up email was sent to the authors who had not responded to our email. Follow-up emails resulted in the participation of no further interviewees.

In total, thirteen STPA experts were interviewed, which was considered sufficient to generate insights on STPA validation by domain experts. The sampling technique in a qualitative study is different from a quantitative one as it aims to explore ideas and understand the reasoning for making judgments, rather than counting responses for a statistically representative sample (Gaskell, 2000). Thus, the decision on how many interviews suffice relates to the richness of the gathered information from interviews (Kuzel, 1992). As a rule of thumb, some researchers have suggested 10 to 20 interviews as a feasible number (Alam, 2020;

Sandelowski, 1995). This however is a suggestion, with a final judgment about sufficiency also depending on the data saturation, which is the situation where no significant new information, i.e., new insights or new themes, is identified from the interviewees (Bowen, 2008). In the current research, each interview was transcribed and analyzed after each interview (see Section 3.2), based on which it was apparent that saturation was reached after 10 interviews. Nevertheless, all thirteen interviews were conducted as these were already confirmed, as the additional information strengthens the findings.

The demographic information of the interviewees is summarized in Table 1. Seven participants are currently employed in academia, thus having experience with STPA mainly in research projects and from teaching STPA. Six interviewees are active in industry, either working in one company or providing consulting services to a wide range of industries. In terms of years of experience, five interviewees are highly experienced in the application of STPA with 10–15 years of experience, and eight interviewees have 5–10 years of experience. Interviewees were also asked about their level of education and their field of study. Seven interviewees have a master’s degree, and six have obtained a Ph. D., mainly in an engineering field.

3.2. Interview process

To perform this research, two in-depth interviews were performed with each expert, which took place in October and November 2022. The interviews were performed via Microsoft Teams, varying in length from 60 to 90 min for each interview session. Before starting an interview, the interviewees were asked to give consent to recording the interviews, to

Table 1
Demographics of the interviewees.

Demographic information	Values and distribution (N, %)	
Field	Academia (7; 54%)	Industry (6; 46%)
Years of experience	[5,10] (8; 62%)	[10,15] (5; 38%)
Highest education level	Master (7; 54%)	PhD (6; 46%)

facilitate subsequent transcription and analysis. The questionnaire was structured into three parts and was reviewed and approved by the authors' institutional Research Ethics Board (REB) under approval number 2021–5761. The detailed interview questions can be found in [Appendix A](#). The first and second parts of the interview questions were asked in the first interview session and the third part in the second interview. Below, the interview process followed for each interviewee is explained.

In part 1, questions about the interviewee's background were asked. In part 2, it was investigated whether the interviewee performs any form of validation for their STPA analyses and if they do, what STPA validation process they adopt, and what the steps and focus points are. More specifically, it is investigated what aspects of an STPA analysis need to be validated from the expert's point of view, even if they do not have a comprehensive validation process or framework to approach this systematically. To best leverage the insights and experiences of the expert, the interviewee was walked through the steps of a generic STPA analysis process one by one, starting from the "purpose of the analysis" (step 1) to "identifying loss scenarios" (step 4), and then "Final results documentation" (see [Figs. 1 and 2](#)). For each step, the related questions for each step, listed in [Appendix A](#), were asked, and an open discussion was held. At this stage, the developed STPA validation framework outlined in [Section 2](#), was not presented to the experts to let them use their own insights to answer the questions.

This part of the study seeks to understand why practitioners focus on those specific aspects of STPA and what they look for to establish validity. From this information, it is possible to reason back to the types of validation tests proposed in the validation framework outlined in [Section 2](#). To do this, after the first interview of each participant, the answers to each question were imported into the NVivo software ([QSR International Pty Ltd. NVivo, 2020](#)). For this, the interviews are first transcribed, and then patterns are identified within the obtained qualitative data, which is called thematic analysis ([Braun and Clarke, 2006](#)). The identified patterns were further categorized and mapped with the tests and proposed concepts in the theoretical validation framework of [Section 2](#). For instance, the way the interviewee decides to stop validation is compared to the way it is proposed in the validation framework. This revealed the similarities and dissimilarities between the validation framework developed based on theoretical foundations, and what STPA practitioners commonly do to develop (what they consider to be) valid analyses.

In the second interview session, the developed STPA validation framework was first presented to the interviewee. Then, the results of the mapping of the results from the first interview with the theoretical framework were presented to the interviewee. Further, the interviewee was walked through each test which they did not mention in their first interview, to investigate if such a test would nevertheless be reasonable to be used in practice from their point of view. Thus, the reasonableness of the whole framework and each test were assessed in the second interview. The questions of the second interview are listed in [Part 3 of Appendix A](#). The result of the second interview was also imported into the Nvivo software, similarly transcribing and analyzing patterns in the qualitative data, as for the first interview.

In addition to analyzing each interview separately, the data gathered from all interviews was combined and analyzed to obtain overall insights for the research study. As the interviews progressed, clear patterns in the data started to reveal. As mentioned in [Section 3.1](#), after ten interviews, no new themes were identified. Once saturation occurred in the gathered data, a final thematic analysis was performed to map the identified categories from all data to the proposed theory-based validation framework. It should be highlighted that in both parts of the analysis, the first author performed an initial thematic analysis of the results, which were then reviewed by the second author of this paper. The results showed a high level of agreement between the authors, and a discussion was held to find a consensus about the findings where interpretations differed.

4. Results

This section presents the results of this study, which is reported in two sections. In [section 4.1](#), the state of the practice among interviewees with respect to the validation of an STPA analysis is explained. This section further presents their judgments on each validation test of the theory-based validation framework outlined in [Section 2](#). [Section 4.2](#) explains the interviewees' judgments on the framework assumptions and proposed theoretical ideas.

4.1. Interviewees' views on the validation tests included in the proposed STPA validation framework

All interviewees stated that they consider validation an important part of an STPA analysis and strive to perform validation, although it may not be possible to do so in all cases. Some STPA experts, especially those working as consultants, highlighted that whether they perform validation depends on the project they work on and what their role in that is. For instance, if a project relates to a regulated industry, there are a lot of formalities, and it may be required to comply with a standard where validation is one of the mandatory requirements. Furthermore, concerning their roles in a project, for instance, if they facilitate the analysis, they are unlikely to be running a validation process. Or, if they are leading the analysis, whether validation is performed, and to what extent, depends on the project's available resources.

The interviewees were also asked whether they follow a structured or formalized validation process. One of the interviewees responded that "I have to be honest; I do not think it has been very good so far. You are asking an extremely hard question, which is good." Most of the interviewees explained that validation is mainly performed through an ad-hoc process and that they do not follow a formalized process, nor are guided by a systematic list of validation tests from which to choose. Thus, they highlighted that the quality of their existing validation practices is unknown.

In general, the interviewees' experiences and opinions with each validation test of the theory-based framework of [Section 2](#), are categorized into four groups: (1) The interviewee already applies this test, (2) The interviewee has not used this test before, but considers that it makes sense to use it in practice, (3) The interviewee has not used this test before but considers that it makes sense to use it in practice, with some caveats and limitations, and (4) The interviewee believes that this test does not make sense to be used in practice. [Fig. 3](#) summarizes the interviewees' opinions about each test, using the above-mentioned categories.

In general, all theory-based proposed validation tests have already been used in practice by at least one STPA expert. The most frequently used validation tests are 'Face Validation' and 'Content and Structure Validation' tests, which are already commonly applied by all 13 interviewees. It should be highlighted that even the tests that are already used in practice are not without limitations, as interviewees highlighted some of their limitations (see [sections 4.1.1 to 4.1.13](#) for the highlighted limitations by interviewees). Only one expert highlighted four tests as not making sense to be used in practice which are 'Nomological Validation', 'System Boundary Validation', 'Concurrent Validation', and 'Convergent Validation', and one interview highlighted 'Extreme Condition Validation' test as not making sense to be used in practice. The findings from the interviews concerning each test are discussed in turn below.

4.1.1. Nomological Validation

'Nomological validation' is one of the tests proposed for the 'Purpose of the Analysis' step, see [Fig. 1](#). Not only this test can be used for validation of the first step of STPA, but also the identified analyses through this test can be used for performing some of the proposed tests in the later steps of STPA, including Concurrent, Convergent, and Historical validation. Thus, it forms a basis for comparison or a benchmark exercise

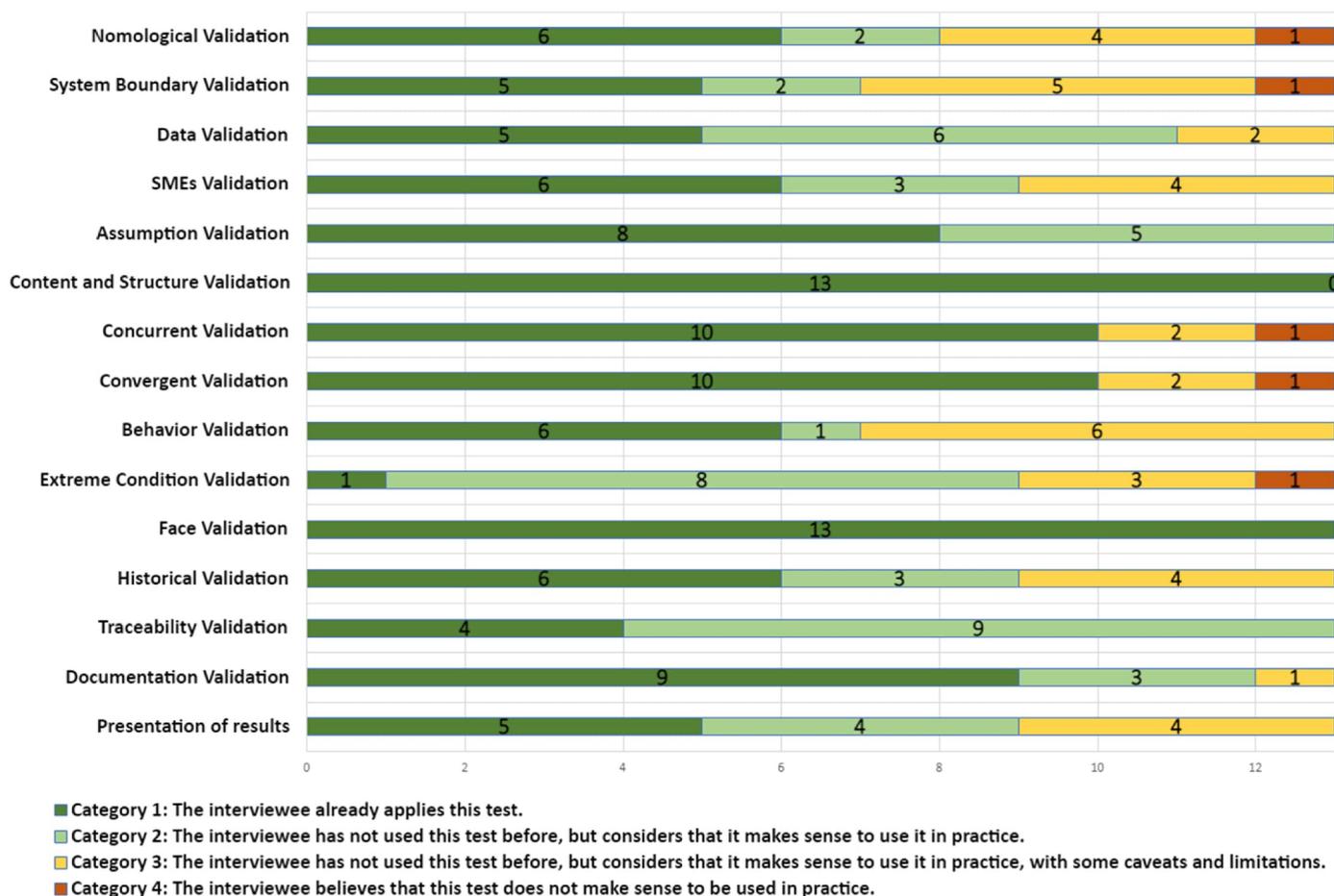


Fig. 3. Interviewees' experiences and opinions on each test.

throughout the analysis. For more information about this test, the reader is referred to [Sadeghi and Goerlandt \(2023b\)](#).

As can be seen in Fig. 3, six interviewees highlighted using this test already occasionally (Category 1) although they have labeled it differently and do not have the same name for it. As worded by one interviewee: “we use a similar test, but we do not call it ‘Nomological Validation’.” Through this test, STPA experts search for other hazard analyses for closely related systems to see if there exist other analyses to be used as a basis for comparison throughout the STPA analysis of the system of interest. Two interviewees highlighted that this test makes sense to them to be used in practice (Category 2), while four experts pointed out that although this test may have some value, they foresee some limitations which make them cautious to start using it (Category 3).

The first limitation concerns the challenges in identifying similar and identical systems, given that each system is unique, and thus not 100% comparable. It was explained that even if the technical aspects of the systems are the same, other contextual aspects, such as the behaviors of the system users, or the national rules in countries where the systems operate, may be different. The definition of the system boundary, for example, if it only includes the technical aspects of a system, or if both social aspects, as well as technical aspects are within the system's boundary, can affect this test's applicability and effectiveness. As soon as humans are included in the boundary, experts considered that it would be very difficult to define identical and similar systems to the system of interest. Furthermore, differentiating identical and similar systems with respect to the system of interest is also considered a challenge. That is, experts highlighted that how a system can be categorized as a similar system or an identical system may need some clearly defined criteria to guide the validation team, but experts were not aware

of a systematic approach to do this.

Another limitation is that other analyses, possibly relying on other hazard analysis techniques, which are used as a basis for comparison may have certain limitations that are not openly discussed. For instance, some losses, system level hazards, and constraints may have been left undiscovered in those studies due to a lack of expertise in their analysis team. If they are identified in the analysis of the system of interest, the validation team can simply question the findings by solely relying on the results of the comparison studies. As reported by one interviewed expert: “the experts should be mindful of the newly identified items that are not identified in the other studies. Since we identified new things that cannot be observed anywhere else, it does not mean that they are not important and can be neglected.”

One participant's concern with applying this test is the occurrence of anchoring bias. Anchoring bias refers to the tendency to rely heavily upon the first piece of information that people come across which can lead to skewed judgments ([Tversky and Kahneman, 1974](#)). The interviewee expressed the concern as follows “the findings of the identified reports or similar studies would determine the boundaries of how I look at things. To avoid this bias, I would probably first build the basis of the analysis so that the validation team can brainstorm and work on it without any mental constraints that come from reading somebody else's work. Then, I would look at those identified studies to compare and to get additional ideas.”

Only one participant believed that this test does not make sense to be used in practice (Category 4). According to this expert, although this test would not even add much value to academic projects, it is frequently used in academia, and it is a disconnect from what should be happening in practice. They explained “you cannot do a literature

search to see how somebody else designs a system. This is not how the analysis should be performed for designing a product that is going to be used in the real world." It was also pointed out that some STPA experts just copy and paste ideas from analyses of other systems, which can impose a significant risk on the system design process.

4.1.2. System boundary validation

A system boundary is defined in relation to the purpose and context of the analysis. Validation of the system boundary is concerned with, first, the explicit definition and documentation of the system boundary, and second, the conceptualization of the effects of changes in the system boundary (Sadeghi and Goerlandt, 2023b). Five participants stated that they already have applied a similar test (Category 1 in Fig. 3) through which the defined system boundary is reviewed. It was emphasized that a focus on the system boundary cannot be overlooked and that if it is improperly defined, this will affect the entire analysis. Making the boundary explicit provides an opportunity for the analysis team to clearly delineate the whole system (which is often done through visual representation), from which experts can explicitly consider the impact that the definition of the system boundary would have on the whole analysis. More importantly, the validation team can ensure that anyone involved in the analysis knows the boundary and that their mental models align.

Two interviewees expressed that this test makes sense (Category 2), and five interviewees highlighted that although it theoretically makes sense, they have some concerns about how it can be carried out in practice (Category 3). They all reported having difficulty with identifying the changes as well as conceptualization of the effects of those changes in the system boundary. One participant explained this concern as follows: "STPA, in general, is textual and qualitative. Identifying how things change could be hard. If you compare two sentences in a natural language, it can be hard to say whether they refer to the same thing or not." Therefore, they all believed that the guide questions in the theory-based validation framework of Section 2 would not suffice for helping the validation team to perform this test, highlighting that clearer guidance or a formalized technique is required.

Only one interviewee believed that this test would not add value (Category 4). The interviewee's reasoning for this was that the analysis team identifies the system boundary with respect to the analysis's purpose. If, for instance, a regulatory authority validates the boundaries, they may have a definition of what they consider the system to be and their perception of the system's definition would be different from the analysis team. Thus, they would look for what may not be aligned with the analysis team's definition of the system. From this participant's point of view, different opinions on the system boundary would solely cause controversy and the conversation around it would continue for a long time without reaching an agreement.

4.1.3. Data validation

The 'Data Validation' test emphasizes the importance of the accuracy and comprehensiveness of input data. According to the interviewed STPA experts, the availability of input data depends on two factors: (i) the level of novelty of the system from the perspective of the company in terms of their experience with the system, and (ii) the stages of a system lifecycle for which the analysis is performed. In terms of the former, it was pointed out that if a completely new system was developed, no data would be available. Hence, only data from other (similar) systems can be used. With regard to the latter, one explained "you have some data sources at the beginning, and as you progress, you find more data. If STPA is applied for the conceptual phase, you start with some basic information. As the system matures and moves into the design phase, other factors come into play, such as legislation that is not important in

the beginning. But if the system is in the operation phase, usually all the details, such as diagrams, experts, designers, legislation, are known."

Five interviewees commonly use this test and emphasized its importance (Category 1). One interviewee explained: "Data validation has to happen because confidence in the analysis should be built bottom-up. I am using some data for hazard analysis, then I use the hazard analysis to design and build a system. Your analysis is as strong as its weakest part, and sometimes a system would collapse just because of using non-validated data." Six participants stated that although they have not used this test before, they consider it a reasonable test to be used in practice (Category 2). One of them suggested including the issue of data interpretation in this test, explaining: "Data is one point, data interpretation is another point, which perhaps is related to the capability of the analysis team. How the data is interpreted is an important point which needs investigation."

Two interviewees raised limitations of this test, although they believed it to be an important test (Category 3). The first limitation is its time-consuming nature. According to one expert, if the system of interest is a relatively simple system, this would not be a problem. However, if a complex system, such as a nuclear power plant, is studied, the input data volume is huge. Then, the time and resource availability for performing this test can become a major project planning and resourcing issue. Another highlighted limitation is the lack of sufficient data. When there is no input data available, data validation cannot happen. One interviewee explained that in one of their STPA projects, the only data source they had available concerned experts' input. They organized project meetings every few weeks to get the technical information they needed from the experts. This was the only way they could make sure they have access to updated data. The interviewee stressed that this requires an effective communication plan between different teams involved, explaining: "the communication between different teams should take place correctly. For instance, in one of our projects, we were informed about the changes in the design too late, which was a huge issue."

4.1.4. Subject Matter Experts (SMEs) validation

As STPA analyses heavily rely on experts, the process of (Subject Matter Experts) SMEs selection, elicitation, and combination could greatly affect the results of the analysis. 'SMEs validation' test highlights the importance of carrying out these processes methodically and systematically. Six interviewees raised using this test occasionally as part of their validation process (Category 1). According to them, one of the influential factors in having such a process is the maturity of the organization that performs the analysis: "When you are dealing with highly mature organizations, because of the level of rigor and how process-minded they are, you will find a documented process that has been followed and also reported for each project." Another factor is the presence of regulatory requirements. Highly regulated industries require companies to provide documentation around the SMEs validation process for their system or product which is going to go through a certification process. For instance, one expert stated that if you are doing a formal risk process, you have to capture everybody's background, to show evidence that they are trained, why they are suitable, what role they have in the analysis, and what their qualifications are. This is a very formal process.

Three interviewees stated that this test makes sense to be used in practice (Category 2). For instance, one mentioned: "SME selection is important because you might have people who are biased toward a certain opinion, so you want a balanced team of SMEs, as well as a facilitator who manages possible conflicts within the team." In addition, although no one highlighted this test as being unreasonable (Category 4), four interviewees pointed out that it is not always possible to use this test in practice (Category 3). The primary issue making this test challenging is the lack of available resources. The selection of SMEs in practice is commonly constrained by the

availability of resources and by who is assigned to the project, which usually is a top-down decision. One interviewee pointed out: "Are they suitably qualified for the program? That is usually not an option to choose." Even in smaller projects or academic studies, sometimes, it is quite difficult to find any expert who is willing to participate in the analysis. To some extent, it depends on the social network of the person who performs the analysis.

4.1.5. Assumption validation

The 'Assumption validation' test, which is another test proposed to be performed in the first step of STPA (Fig. 1), stresses the importance of identifying and documenting assumptions and agreeing on them with the stakeholders and decision makers. A plethora of assumption validation tests exists in the literature. In the theory-based STPA validation framework by Sadeghi and Goerlandt (2023b), a method proposed by Landry et al. (1983) is adopted, rooted in ideas by Mason and Mitroff (1981). Eight interviewees highlighted commonly performing assumption validation in their STPA analyses (Category 1), using a technique that they consider suiting the project best, i.e. not necessarily applying the proposed technique in the framework.

One of the most frequently used assumption validation techniques among participants is an expert review process through which the assumptions are presented to domain experts to investigate if they are reasonable. One participant explained this through an example in a railway application: "if there are some assumptions about how train drivers drive, they need to be confirmed with either a train driver or someone who teaches train drivers or someone who is representative that knows how they actually do it". In case domain experts are not available, it was reported that the assumptions can be confirmed with the relevant literature.

One interviewee explained their internal process for assumption validation which they called the "Assumption Challenge." Through this process, every single documented assumption is reviewed to make sure they are correct, clear, complete, and what is the consequence of the violation of an assumption. "The assumptions are stated one by one, they are discussed until the whole team reaches an agreement on them." Furthermore, throughout the analysis, the assumptions are periodically reviewed to ensure they are still valid. If they are not valid anymore, the analysis should be (partially) redone.

Five interviewees indicated that they have not performed assumption validation before, but, based on their experience, they recognize it to be an important test (Category 2). One explained an example about the challenges that came up during an STPA analysis only because the assumptions were not agreed upon: "I noticed that people were working with a lot of assumptions in their head, and I did not know that they have those assumptions until later stages of the analysis." This resulted in some conflicts between team members while doing the analysis. "This is something that could come up in the assumption validation test, which would have saved us some time" the interviewee recognized.

In terms of ensuring the completeness of the assumption set, it was stressed that this cannot be decided methodically. It is challenging as the assumptions can be completed to the extent that the analysts want. As formulated by one STPA expert: "you can go as deep as you want, but at some point, you need to define what you consider to be enough for your analysis in terms of assumption completeness." Therefore, experts consider that completeness is essentially achieved when the analysis team feels satisfied with the scope covered in the analysis.

4.1.6. Content and Structure Validation

As seen in Fig. 1, the 'Content and Structure Validation' test is the first proposed test for the second STPA step, where the control structure is developed. As its name suggests, this test targets the validation of both

the content and structure of the control structure for the system under analysis. Thus, it aims to investigate the accuracy and comprehensiveness of the elements included in the control structure, as well as their functional relationships. All interviewees reported using a similar test for the control structure validation (Category 1), and they all emphasized the importance of this test because the rest of the analysis rests upon the definition of the control structure.

A formal review process is one of the repeatedly mentioned techniques through which the whole team, including SMEs, reviews the control structure to ensure its accuracy and completeness. The review process sometimes is implemented through workshops where different procedures within the system are reviewed with experts. Then, the places on the control structure where the mentioned procedures happen are identified. Through this process, system aspects that are not reflected in the control structure and thus are missing can be identified.

Some of the STPA experts, who are mainly working on small industry or academic projects, benefit from the available technical documentation, including flow diagrams, hardware and software diagrams, and documented procedures, to validate the control structure. This is because, as explained in Section 4.1.4, sometimes they do not have access to experts for validation work, so they can validate the control structure against the system's technical documentation. "I know the control structure is different from the physical model, but the available technical documentations form a good basis for validation" one interviewee explained.

One challenge with this test which is highlighted by several interviewed experts is when those involved in the validation process are not familiar with STPA. Since STPA uses terminologies and concepts that are relatively new and constitutes a type of technical jargon, not everyone is familiar with these. Based on the interviewees' experience, this problem is more evident especially with the control structure, as it is a specific feature for STPA, with little to no comparable representations used in other hazard analyses. Hence, not having training about it, complicates comprehending the STPA analysis and thus poses a challenge to validation work. So, in most cases, a significant amount of time should be spent training the SMEs about how to do STPA, to enable a fruitful validation process.

4.1.7. Concurrent and Convergent Validation

'Concurrent and Convergent Validation' tests refer to other STPA analyses for identical or similar systems (or subsystems) to use the developed control structure in those studies as a basis for comparison. According to the framework by Sadeghi and Goerlandt (2023b), in the 'Concurrent Validation' test the control structure of identical systems, and in the 'Convergent Validation' test, similar systems are used. These are popular tests among STPA experts, as ten interviewees have been already commonly using this test (Category 1). One of the interviewees explained that they found this test particularly helpful, especially when circumstances do not enable access to SMEs.

It was also mentioned that a comparison analysis for the whole system typically does not exist. However, experts found that analyses for specific parts of the system or subsystems often can be found, which can be used for validation of part of the control structure. A possibly problematic issue with this approach is that validating part of the control structure does not mean that the whole control structure is a suitable basis for further STPA steps. If done this way, the whole control structure also needs to be validated using other techniques, such as an expert review.

Two interviewees considered these two tests as making sense and believed that these indeed can be helpful, but noted that these also have a limitation (Category 3). They highlighted that on the surface level, systems can look similar or identical, whereas there can be all kinds of hidden feedback loops and dependencies that make the operation of the systems different. For example, identified differences in their control structures can be due to the hidden differences in the systems that the validation team is not aware of them. Thus, such comparisons can

generate misleading results. This is a similar issue as with ‘Nomological Validation’, where experts highlighted that how a system can be categorized as a similar system or an identical system may need clearly defined criteria to guide the validation team, but experts were not aware of a systematic approach to do this.

One participant stated that these ‘Convergent and Concurrent Validation’ tests do not make sense (Category 4). The interviewee’s reasoning for this is that there is no way to guarantee that a control structure defined by someone else, for another system, is valid. This expert highlighted that, for example in academic publications, control structures may have been defined by students who may have never done an actual engineering design, so the fact that it is published does not mean that it is correct. The interviewee explained: “I think in general people, instead of using the results of other studies for validation purposes, they just copy and paste other analysts’ results and use it as a shortcut.”

4.1.8. Face validation

For the general case of hazard analysis validation, ‘Face Validation’ is one of the most frequently used tests in both academic (Sadeghi and Goerlandt, 2021) and industry contexts (Sadeghi and Goerlandt, 2023a). In the STPA validation framework of Section 2, this test is proposed for validating both the Unsafe Control Actions (UCAs) and loss scenarios, see Fig. 1. Through this test, the validation team, who are knowledgeable in the domain of the studied system, reviews the identified UCAs and loss scenarios to judge whether they appear comprehensive and accurate. All interviewed STPA experts stated that they perform face validation (Category 1). They, however, do not limit its use to just these two steps and apply it to almost all steps. The interviewees also reported that the knowledge in the minds of people (i.e. tacit knowledge), often cannot be found elsewhere. So, the judgments of SMEs often play a vital role in an STPA analysis validation.

Each interviewed expert reported a somewhat different process for face validation, ranging from informal review processes to more elaborate workshops. For instance, one interviewee mentioned that instead of having a group discussion through which experts’ opinions can be biased by each other’s opinion, they ask the team of experts to think about it separately and write down their comments and then they all share their ideas to discuss, through a type of Delphi exercise. Some participants mentioned that they do not have a structured way for this validation test, but that they request experts to join meetings on an ad-hoc basis to present their results and to gather their comments. The main issue raised about this test is that it greatly relies on the expert’s knowledge and experience, so the results can be unreliable, especially if the experts are chosen without careful consideration.

4.1.9. Behavior validation

The ‘Behavior Validation’ test, which was first suggested by Har-kleroad et al. (2013), compares the behavior of the system with and without enforcement of the identified constraints on each controller. Through this test, the validation team confirms whether the identified constraints change the behavior of the system the way it is expected (Sadeghi and Goerlandt, 2023b). Six STPA experts mentioned that they already use this test (Category 1). According to them, performing a behavior validation is a must when the system of interest is safety critical and complex.

As seen in Fig. 3, one interviewee believed that this test makes sense (Category 2), and six interviewees highlighted some limitations in it (Category 3). As stated by them, this test would make the most sense if the STPA analysis is performed for a system in the operation phase, where the relevant aspects of the actual system’s behavior can be monitored. If this test is performed for a system in the design phase, the test can be performed through a simulation model, which is not ideal. Their rationale for this assertion is that there may be many challenges with and uncertainties associated with a simulation model. One

interviewee quoted a well-known aphorism in this context: “all models are wrong, some are useful.” For instance, if there is a discrepancy between the results of the simulation model and the behavior of the system the analysis team expects after enforcing the constraints, it would not be clear which one is reliable. However, if a simulation model is used for performing this test that should be considered to be an initial task, with validation testing being continued to the operation phase, where the actual results can be observed.

In addition, some interviewees highlighted that behavior validation is typically performed as part of the Validation and Verification activities through the systems engineering process, which focuses on the design validation, rather than on validating the safety analysis as a process in itself. Considering the required available resources for doing this test, interviewees were concerned whether it is reasonable to have such a test twice, one for the STPA analysis, and one when the designed system is operationally tested. Furthermore, a few interviewees highlighted that the purpose of the analysis for which STPA is performed could be another factor in deciding whether to apply this test. For example, if STPA is used for preliminary hazard analysis, it does not make sense to be used as this is the first step of hazard analysis and other analyses would be performed in later stages of the safety analysis process.

4.1.10. Extreme Condition Validation

The ‘Extreme Condition Validation’ test can be used to evaluate the plausibility of identified constraints for extreme and unlikely interactions of components within the system, as well as between the system and its environment. Only one interviewee mentioned that they have used a similar test in practice for one of their projects (Category 1). This expert explained that in their analysis they considered some extreme conditions both within the system and outside of the system to investigate how the system would react and whether the identified constraint would be effective. This test was performed through several discussions with the design team, and with other SMEs who were involved in the analysis. However, there was no systematic approach to select what extreme conditions to focus on.

Eight interviewees highlighted that the test makes sense (Category 2). One mentioned: “I think it is an interesting concept, and it can likely be very useful. Because the idea is that you create controls that are strong enough or designed well enough to be able to put constraints on risk factors. This can help in determining if the constraints are held up in different scenarios.” Three participants mentioned that although it makes sense, it has some limitations (Category 3). The main challenge they see with this test is how to define the extreme conditions, stating that the effectiveness of the test probably depends significantly on how well the test can be designed. One expert explained that they would refer to historical data or expert’s experience for defining the extreme scenarios, which however would work only in limited cases when the system of interest is not a significantly novel system. In case a novel system or a highly complex system is designed, this expert found that this test cannot be easily performed. They concluded that some confidence can be gained through this test, but that it cannot be claimed that the results are fully validated until the system is in operation.

From one interviewee’s perspective, this test does not make sense (Category 4). The reasoning for this was that STPA already is a worst-case scenario analysis as is conceived so extreme conditions should have been considered already within the analysis. The expert explained: “it is reasonable to ask the validation team to check if all the worst cases have been identified, but that would be more related to the context validation, where you ensure that the list is complete, and nothing was overseen.” This interviewee also highlighted that the analysis team does not have influence over anything outside of the boundary and that therefore, the boundaries should be defined carefully and validated which would be done using the ‘Boundary Validation’ test. The expert further explained this by giving an example: “for example, an earthquake is possible to occur and damage the system. Thus, this should be

analyzed as part of the analysis.”

4.1.11. Historical Validation

As can be seen in Fig. 1, the ‘Historical Validation’ test is proposed for validation of the fourth step of an STPA analysis, which is identifying loss scenarios. Through this test, any available historical data (e.g., incident data) of the system of interest, or of identical or similar systems identified through the ‘Nomological Validation’ test is reviewed. This test investigates whether the incident/accident contributing factors are relevant to the system of interest, and if so, whether they are covered in the identified loss scenarios (Sadeghi and Goerlandt, 2023b). Six interviewees highlighted having already used this test, finding it helpful (Category 1). One mentioned: “Thankfully, the company I am working for does not have accident data of its own, but there are so many published data that we refer to.” They also explained that human error has always been a risk when relying solely on SMEs. Thus, this test is helpful for cross-checking the accuracy and comprehensiveness of the SMEs input.

Three interviewees highlighted that although they have not used this test, they believe it makes sense (Category 2). Four interviewees see some limitations in this test (Category 3). As reported, the first issue with referring to the accident investigation reports is that these may be biased. If the report was prepared by an internal team, the experts asserted that they might have been under pressure not to disclose everything. If the investigation was reported by an external team, they might not have had access to all details of what actually happened. One expert labeled this as survivorship bias, where the focus would be on entities that passed a selection process while overlooking those that did not (Elston, 2021). The report can be also affected by hindsight bias, which refers to the tendency to look back at an accident that could not be foreseen at the time, thinking that the outcome was easily predictable (Dekker, 2014).

Furthermore, some experts asserted that if one can find historical data, it would barely say anything about past events and it is certainly very hard to say anything conclusive about the future. For example, a high risk component may exist in the system but may not have triggered an adverse event for various reasons, such as the operating conditions or the environment in which the system operates. However, for systems operating in a dynamic environment, not everything can be predicted as conditions can change. Thus, variations in the environment can trigger an accident, which may not be evident from historical events.

Another issue with this test is that when a new system is built, historical data would not be available. Even when a change is made to an existing system, it is hard to tell whether all assumptions still hold, and to what extent old data is still representative. One expert explained: “the change can propagate throughout the system and then you would have a totally different system and a lot of your assumptions are no longer valid.” Thus, while historical data can be used as an inspiration as a kind of reality check, relying too much on historical data when it is inappropriate would cause an even bigger challenge.

4.1.12. Traceability Validation

According to the STPA handbook (Leveson and Thomas, 2018), traceability between different items of an STPA analysis should be maintained throughout the analysis. Once all steps of STPA are completed, the validation team reviews the traceability to ensure that all information is logically consistent and appropriately documented. Four interviewees reported using this test already (Category 1). From their perspective, the reason why it is important is that it shows that nothing has been missed and different elements of STPA are properly traceable. Furthermore, the analysis may change over time, for instance, due to a change in the system design. It is helpful to have traceability, so that when something changes, it is relatively easy to inspect what other aspects of the STPA analysis are affected by this change, and what parts of

the analysis should be updated.

The other nine STPA experts reported that they have not checked the tractability in STPA analyses they have been involved in, but all experts believe that this test makes a lot of sense and indeed seems reasonable to perform (Category 2). One interviewee explained: “it kind of provides you an opportunity for a last-minute check.” They also mentioned that if you are not using software that generates the tractability automatically, there can be different errors, such as typos. This test could help investigating and mitigating such errors.

4.1.13. Documentation checking and review of presentation of results

‘Documentation checking’ and ‘Review of Presentation of Results’ tests are concerned with the accuracy and comprehensiveness as well as the credibility of the results of an STPA analysis. The former is conducted to evaluate the quality of documentation and to ensure it is in formats understandable to stakeholders. It is noted here that the documentation enables a review of the entire process behind the analysis, rather than focusing on just one step. The latter concentrates on the communication of the results to stakeholders and decision-makers. Through this test, the validation team ensures that the sources of uncertainty and the limitations of the analysis are included and will be brought up in the presentation of the final results.

Nine interviewees highlighted that they already perform documentation checking (Category 1). Some of them emphasized that the documentation is produced as they go through the STPA process as it is time-consuming and should not be left for the end of the analysis. Thus, checking the documentation’s correctness and completeness is performed after each documentation step. Several interviewees also pointed out that they commit to using clearly defined concepts and terminology from the STPA analysis and the underlying STAMP accident model, to facilitate a clear understanding by stakeholders and decision-makers. As expressed by one interviewee: “STPA uses terminologies that have different meanings than those conventionally used. Not everyone would understand the documentation of the analysis without clarifying terminologies.” So, they pointed out the importance of documentation in formats understandable to those involved in, and those relying on, the analysis.

Three interviewees mentioned that they have not performed documentation checking before, but that they believe it makes sense (Category 2). One participant see a limitation in this test (Category 3). According to this STPA expert, this documentation checking cannot be performed in projects where there is a time/resource limitation. It was mentioned that this test would be easier to implement in large-sized companies, as they have access to enough resources to do it. However, it cannot be necessarily conducted in small companies or projects, unless it is either mandatory or if there is a strong commitment to perform a high-quality analysis.

In terms of reviewing the presentation of STPA results, five interviewees mentioned that they perform a similar test (Category 1), with the results usually being provided in the form of a presentation before it is disclosed to the stakeholders and decision-makers. These interviewees pointed out that they make sure that this presentation encompasses the key findings and results, recommendations, limitations, uncertainties, and other things as requested. According to one expert: “the limitations and the sources of uncertainty are important for decision-makers to understand, as knowing them could affect the decisions.” It is also highlighted that if the results appropriately communicated, the stakeholders and decision-makers would easily understand the analysis. For instance, what is covered in the scope of the analysis, what the issues were in the analysis, how they were tackled, and so on. One interviewee mentioned that they always try to involve the experts and the stakeholders as they go through the documentation of the analysis. This way credibility can be obtained cumulatively.

Four interviewees mentioned that they have not performed this test while affirming that they believe it makes sense in principle and that it could add value in practice (Category 2). Four interviewees, however, see some limitations in this test (Category 3). The first limitation is that not anyone, even the validation team, knows the system design sufficiently well to detect the potential mistakes or limitations of the analysis in a report. They may find grammar mistakes, but the expert interviewees considered that a validation team would not, in general, find the mistakes that actually matter from a safety point of view. As one interviewee explained: "I think the premise is that an external person is going to find the mistakes that the analysis team cannot find. That is possible, but I am a little skeptical of that." They consider that a documentation check would happen more on a surface level, and would be more related to good communication practices than to content-checking.

The second issue raised is that if the analysis team has a presentation, it is going to be very abstract because the majority of people do not want to go into details. For instance, one interviewee stated: "we always communicate the results with the stakeholders and decision makers so we can get a feedback from them. However, we do not always receive feedback. The reason is they are not concerned with the details, especially the top management level."

4.2. Interviewees' views on processes and theoretical concepts included in the proposed STPA validation framework

In the previous section (Section 4.1), the interviewees' opinions on the proposed tests in the STPA validation framework were presented. This section first specifically focuses on the interviewees' thoughts on the concepts and propositions embedded in the STPA validation framework (Section 4.2.1 to 4.2.3). Then, the potential challenges the interviewees may face with using this framework are presented (Section 4.2.4).

4.2.1. Validation team: independence

The proposed STPA validation framework outlined in Section 2, and elaborated in Sadeghi and Goerlandt (2023b) suggested having two separate teams, one in charge of implementation and one for validation of STPA. The logic behind this suggestion is justified by referring to a psychological phenomenon called "the IKEA effect." This refers to a cognitive bias through which people overvalue their own creations (Norton et al., 2012). In addition, although the two teams work in tandem, they ideally are independent. Lack of independence between the two teams may result in not taking into account the potential errors or limitations in the study, which further leads to loss of information (Sadeghi and Goerlandt, 2023b).

To investigate if this assumption aligns with what happens in practice, STPA experts were asked whether, in their projects or organizations, a separate team from the analysis team performs validation. Although the responses to this question varied, all interviewees believed that having two separate and independent teams is important. However, some STPA experts highlighted that they face practical challenges to have an independent team in practice.

A few interviewees, mainly those working in or with large-sized organizations, highlighted that they perform validation independently as it is inherent in their organizational culture, while also being recommended by the industry or organizational guidelines they follow. In contrast, some experts pointed out that they have never had a separate validation team, even though they all believe that this was a limitation in their validation practices. For instance, one interviewee explained: "in our projects, the validation team has always been the same as the analysis team, which is problematic. Independence would result in better results."

Furthermore, as reported by some interviewees, primarily those working on a project-basis or providing consulting services, stated that

although having an independent validation team is really important, its feasibility depends on two factors: the project's available resources, and the internal processes that are laid out for them by the organization or the project. It was pointed out that sometimes, the number of analysis team members is considerably low so they cannot form two separate teams. Instead, they prefer to allocate all available experts to the STPA analysis and then perform the validation with the same team rather than having two small separate teams, which may reduce the expertise and evidence base for the actual analysis. They further explained that when the team is small, it would be challenging to have two diverse teams because, most likely, in the best-case scenario, only one expert would be available for each area of required expertise.

The STPA experts are also asked to follow the internal processes set out for the specific project that they work on. It is possible that the expert, for instance when providing services as a consultant, does not have authority over how the STPA implementation and validation teams are formed. One interviewee explained this as follows: "Sometimes, the client requests us to perform STPA analysis and its validation. In such cases, they usually are just looking for the validation report and that would be enough for them."

4.2.2. Different applications of the proposed STPA validation framework

As outlined in Section 2, the proposed validation tests can be either used in parallel with the STPA implementation (Fig. 1) or in a post-hoc manner (Fig. 2). In a parallel process, the validation tests of each STPA step are carried out as soon as the step is completed, while in the post-hoc manner application of the framework, all tests are performed once the whole STPA analysis is completed. In the proposed framework, the parallel application of the tests with STPA implementation is considered the best way of performing validation, i.e. it is hypothesized to lead to the best results. However, it would not be always possible due to practical limitations such as the need to keep to project schedules.

To test this assumption, in the first interview session, the STPA experts were asked whether they perform the validation in a parallel process or in a post-hoc manner. The responses showed that what has been proposed in the framework seems reasonable and is aligned, to some extent, with what actually happens in practice. The majority of the interviewed experts perform aspects of validation in parallel with an STPA implementation. They mentioned that they have a session to work on STPA analysis, followed by a session with experts to review it. Thus, they validate the results as they go through the analysis. They pointed out that the logic behind this process is that they want to find possible errors and correct them before continuing the analysis further. As formulated by one expert: "we did one step and then handed over the results to our validation team. They gave feedback to us, and there were possible modifications that we should make before moving to the next step."

In addition, some interviewees highlighted that the choice between having a parallel and a post-hoc process depends on the specific project they work on. As quoted by an interviewee in the automotive industry: "in the projects where the analysis is performed at the higher level, for example at the vehicle level description, functional description, or system level description, the validation can be done afterward. However, if the analysis is done at the lower level, for instance, the software level, the complexity would dramatically increase where you can never do one big shot of validation." In such cases, they decompose the analysis into smaller steps and perform the validation for each step in a series of parallel processes.

Furthermore, when the different application types of the framework (Figs. 1 and 2) were presented to the STPA experts in the second interview sessions, the majority of them expressed that both applications make sense, depending on the situation in which the framework is used. Some judged that only the parallel process makes sense, finding that the post-hoc use of the framework is not reasonable. One reason why they believe the parallel process would bring more value is that if the validation team gets involved right from the outset of the analysis, there

would be much more understanding and clarity about the system of interest and the subsequent STPA steps. Also, they found that based on the level of detail of the analysis and the complexity of the system, one validation step encompassing all would not even be practically possible. One expert explained: “If you are dealing with a complex, large-scale system, you can never perform one system-wide validation, and it should be decomposed into smaller elements.”

Another reason highlighted by the interviewees is that performing the validation of STPA afterward will cost a lot of man-hours. It was also expressed that, if the validation results are positive, then it could be concluded that a lot of money had been wasted. If the results are negative, this would be an even bigger issue, because the system may need to be redesigned which costs even more money. Thus, if the validation team takes small steps in a parallel process, such issues would not occur.

With respect to using the validation framework by regulatory authorities in a post-hoc manner, as identified as a possible application mode in [Sadeghi and Goerlandt \(2023b\)](#), one interviewee explained that viewing validation in this way would be too simplistic to be meaningful and useful in practice. The main issue raised is that an external expert would not know the specific design sufficiently in-depth, and would not have the required insights in the complete system, to perform a meaningful validation, even if a formal process were followed. The expert made the following statement: “the idea that two experts, for instance from a regulator, will know and have visibility into what they need to comment on is kind of superficial. That is actually inherently the reason why they just ensure that you did follow the general process and they are not saying whether your analysis is actually correct.”

However, some interviewees pointed out that although they perform STPA implementation and (aspects of) validation in parallel, this does not occur exactly as proposed in the framework. That is, they perform validation as they go through the implementation process, but it is not necessarily performed as soon as each step is completed. Having such a strict process may not be always possible in practice, mainly for reasons related to project scheduling.

4.2.3. Validation cessation

To understand how STPA experts in practice handle validation cessation compared with the theory-based notions as presented in the framework outlined in Section 2, practitioners were asked how they decide on ceasing the validation process in their first interview session. Two themes emerged from interviews, which align well with the proposed ideas in the framework.

First, most interviewees highlighted that validation cessation is decided subjectively based on a practitioner judgment. They explained that when the STPA analysis and validation teams think that the analysis is acceptable and there is a consensus about it, then they decide to stop. One interviewed expert explained this using an example: “when you are looking for similar analyses to perform a comparison, at some point, you notice saturation, meaning that you keep finding similar information. So, you do not have specific quantitative criteria but you feel like you can stop searching.”

Second, some interviewees mentioned that the validation is generally a time-bound and resource-bound activity and it is continued until time and/or resources run out. As formulated by one of the experts: “we define the scope clearly, and the activities based on our available time and resources. When we covered everything within the scope considering our timeline, we stop the process.”

In the second interview session, when the framework was presented, the idea behind the validation cessation was also explained to them. Almost all interviewees agreed with the related principles in the framework. However, one practitioner raised the point that there should

be some predefined criteria by the organization or industry on how to decide on the validation cessation, for example, something similar to the As Low As Reasonably Practicable (ALARP) principle ([Hurst et al., 2019](#)). He explained: “if I am the safety manager on a project, I am not responsible for the progress of the project. So, if I see an open risk, I should bring the project to a full stop no matter what the timeline of the project is. If it is going to be just based on the practitioners judgment, it would not culminate in satisfactory results.”

4.2.4. Potential challenges with using the framework

Most interviewees affirmed that, in general, the proposed framework appears to be a good foundation to better ground and formalize an STPA validation process in practice. However, several challenges were raised by interviewees. A recurring theme is the lack of clear guidance on how exactly to perform some of the tests makes the use of this framework challenging. Some validation tests, such as ‘Traceability Validation’, appear rather straightforward to most interviewees, and they know how to apply them. Other tests are conceptually more difficult to grasp, and interviewees expressed a desire for more comprehensive and clear guidance for those, for instance, ‘Assumption Validation’ and ‘System Boundary Validation’. To tackle this challenge, several interviewed experts suggested developing a formalized technique for each test in addition to the guide questions. From their view, these questions suffice to serve as a guide for focusing attention on certain aspects of an STPA analysis to inspect, and for providing ideas about what to look for and how. However, several interviewees found that these guide questions provide too little detail on how exactly to perform the tests.

The generic nature of the guide questions led some interviewees to opine that these are insufficiently specific to be used in practice, as put for instance by one interviewee: “while some of these tests are really important to be done, at the same time, the guide questions are generic and confusing.” They doubted whether the provided questions for each test are exhaustive. One asked: “this is obviously an important direction, and I think this is also a right path that you are trying to reach out to certain people who have used STPA and understand if the framework makes sense or not. My concern is whether there is a guarantee that I do not need to ask any other questions and whether the list of guide questions is exhaustive?”

In addition, some experts highlighted that the framework is quite generic and that in their experience, it is difficult to use generic frameworks for all projects in practice. They believed that it is helpful to have an explicit, documented STPA validation framework that can be used as a guideline. However, these experts stressed they believe a generic framework needs to be amended for each use case. One interviewee explained: “my challenge is how it can be used for different stages of a system lifecycle. It will be a huge value to define different detailed frameworks for each stage of a system lifecycle.” Another issue raised by experts related to the generic nature of the framework is that they believe the degree of usefulness depends on the level of maturity of the organization. According to one interviewee, whether most of the tests are used is a matter of organizational maturity in terms of the safety management system. They believed that the framework can be helpful for small organizations “as is”, while mature organizations may add more tests to this framework.

Two interviewees mentioned that, in their experience, the STPA technique is best fit at the beginning of a project, i.e. in the early concept design, where there does not necessarily exist a correct answer. One interviewee explained: “this framework gives food for thought, and I think what troubles me a bit is that STPA in practice should really be done at the beginning of design, and the purpose of it is actually to help making decisions on design. It is like saying you designed the kitchen, did you do it correctly? And I think in an academic setting, there is a strong emphasis that there is a correct answer while in reality there is not this sense of there is one correct answer.” This interviewee explained that this is more consistent with something like a design Failure Mode

and Effects Analysis (FMEA) (Ericson, 2005) where there exists a clear mechanical part in an engineering system and the failure modes are identified and it can be asked whether something can actually physically happen and what would be its consequences.

A few interviewees highlighted that they believe when experts are asked to validate an STPA analysis, they are generally motivated to find errors or limitations because they are incentivized and paid to. The challenge they see for validation is what people perceive as possible, and the risk of coming to a deadlock situation in which the validation team has unnecessary opposing views. In such cases, the analysis team has a burden of proof to show that the validation team's scenarios actually are not physically possible because the basics of the design have not been taken into account. As explained by one interviewee: "there is actually potentially a real concern of whether experts are voicing what is truly possible versus actually just creating a distraction of things that you have to prove is not valid."

The issue of probative blindness was perceived to be a barrier to effectively using this framework by some interviewed experts. This issue has been already raised by Sadeghi and Goerlandt (2023b) as one of the risks of using this framework. This phenomenon, which is described by Rae and Alexander (2017b), refers to establishing false assurance about safety where this is not in line with actual safety. As put by one interviewee: "in some cases, I would be more worried about passing a test rather than failing a test in this validation framework." The reason is that once one test is passed, it will build some confidence in the STPA analysis that may be false. Thus, it is possible that using a test is more perilous than it is advantageous.

One interviewee also made an important point, which was already raised by Sadeghi and Goerlandt (2023b), namely how it can be shown that doing validation would indeed lead to better results. He explained: "I think the proposed framework would culminate in better results, but it is probably not super straightforward to show that", further explaining that this is complex because it is hard to show whether a system is actually safe, and it is difficult to be 100% confident about the safety of a system. He explained that this is somehow connected to the issue of the validity of the validation framework and mentioned: "how do you validate the validation? This can go on and on."

A further challenge with using the validation framework is that it may be time-consuming and would likely extend an already lengthy and resource-intensive process of STPA implementation. One interviewee remarked: "I think that doing this is a good checkup or line of defense to make sure that things go right. But it takes resources, time, and expertise, and it is not easy. And you can do that only for systems that they want to invest in safety because it is within the business model."

Convincing stakeholders to consider validation as part of the project is another challenge raised by several interviewees. They explained that stakeholders or decision makers may be already satisfied with the STPA analysis obtained before validation. The concern then is how to convince those stakeholders that validation is required. One interviewee mentioned: "I am afraid that people do not have time and do not see the need, so you have to make them aware that this is important. This can start from educating the students."

Finally, some interviewees had some difficulty with the choice of the term "validation", believing that a fundamental ambiguity exists between this concept and "verification". The fact that both terms are used in engineering design projects in a somewhat different manner than in the STPA validation framework of Section 2, appeared to be a reason for this confusion.

5. Discussion

5.1. Discussion on the results

The results of the interviews with STPA experts showed that the theoretical ideas and validation tests in the STPA validation framework, which is briefly explained in Section 2, are to some extent in line with

what has been done in practice. All interviewees expressed that they see value in performing validation, and they already use, mainly, informal validation approaches for their STPA analyses, which are in line with the findings of the research by Sadeghi and Goerlandt (2023a). They also expressed that although they see some limitations in the framework, the proposed validation framework seems to be helpful.

The importance of having a comprehensive validation process for STPA analysis is emphasized by this study. Using all validation tests and processes proposed in this framework, rather than relying solely on one or two tests, is critical for ensuring the validity of different elements of the analysis. For example, a concern raised by an interviewee was the use of academic publications to validate control structures, as the fact that it is published does not necessarily mean that the author of the publication is the right SME to rely on their analysis (Section 4.1.7). This highlights the interdependence between the validation of the control structure and the validation of the subject matter expert. That is, if the SME is not "valid" then neither is the control structure developed by that SME. Thus, by using all the proposed tests together, confidence in the analysis can be strengthened.

The results showed that all interviewed STPA experts already perform 'Face Validation' and 'Content and Structure Validation' tests, mainly benefiting from and relying on SMEs knowledge and experience. Even, for instance, in case historical data is available and the 'Historical Validation' test can be performed, an expert can be consulted to determine if such data is still relevant and to what extent (Pasman and Rogers, 2020). This, even more, stresses the importance of the 'SMEs Validation' test as the results of validation hugely rely on SMEs. Baybutt (2018) discusses 28 cognitive biases that may affect the results of experts judgment which may invalidate the results of hazard and risk analysis studies. Rae and Alexander (2017a) emphasize the importance of SMEs validation, and being aware of the limitations of expert judgment, for instance finding that while their knowledge of causal mechanisms may have a special standing, relying on their quantitative estimates of parameter values often is more problematic. Likewise, experts need to carefully investigate the existing techniques and choose the proper technique which suits the project best (Pasman and Rogers, 2020).

According to Sadeghi and Goerlandt's (2023a) empirical research, validation can be a time-consuming and resource-intensive process, and practitioners may struggle to convince stakeholders of its necessity. These are also highlighted as some of the challenges the STPA experts may face if using this framework. The absence of evidence to show that the framework improves an STPA analysis and makes the system safer can make it difficult to convince stakeholders. Additionally, for Quantitative Risk Analysis (QRA), which has been used for a much longer time and the validity of which has been studied to a greater (although also quite limited) degree than of STPA, there is very little evidence that QRA leads to better decisions or safer systems (Goerlandt et al., 2017).

Sadeghi and Goerlandt (2023b) have previously identified the issue of probative blindness (as outlined in Section 4.2.2) as one of the limitations of the framework. This is not; however, specific to this framework. Probative blindness can be a challenge with any form of validation activity, even those informal processes used by experts. This could be an inherent issue with any form of validation if the experts take a binary approach. Thus, a shift in perspective on validation may be required. That is, instead of expecting the framework to culminate in a validated STPA with 100% complete and accurate results or a binary reject/accept decision, experts can aim for a subjective and formative framework. The former aims to increase the intersubjective agreement among the whole analysis team, and the latter helps the analysis team to find potential errors. This is in line with the issue in risk science also means a hazard analysis technique aims at developing a shared construct, not a "true" analysis (Aven and Guikema, 2011; Rosqvist, 2010). Thus, the tests should more be seen as guides for different stakeholders to agree on the results of the analysis.

A practitioner raised the concern that ideally there should be pre-defined criteria by the organization or industry, such as the As Low As

Reasonably Practicable (ALARP) principle (as discussed in Section 4.2.3), to determine when to stop validation. However, the assumption behind this framework is that validation cessation can be better decided through a discussion between the two teams, since having some pre-defined criteria is a rather simplistic approach to decision-making, and is often argued against in risk research (Aven and Kørte, 2003). In practice, validation may still be ceased even if criteria are not met, due to the limitations of the project. Relying solely on a criterion for ceasing the validation process can lead to poor decision-making, as it largely disregards the system in which the measure is implemented (Langdalen et al., 2020).

The issue of inconsistency in validation-related terminologies has been raised by various scholars in the academic field. Sadeghi and Goerlandt (2021) found through their empirical study that the terms used for validation in literature are not consistent and various authors employ different terms, such as validation, verification, usefulness, trustworthiness, and evaluation. Defining terms, as noted by Oberkamp and Trucano (2008), is a challenging task, and it is not possible to harmonize the terms used by experts. However, the risk of inconsistent terminology can be reduced to some extent by clearly defining the terms. For instance, validation as defined in the framework concerns the comprehensiveness, correctness, and credibility of the results of an STPA analysis. With this definition, there is a better chance to achieve a shared understanding of the term among the different analysts and stakeholders.

Furthermore, some degree of flexibility in the application of the framework is beneficial in addressing the “job perception gap,” which refers to the gap between the defined formal procedures and the informal procedures or local practices (Möller et al., 2018). For example, if the process requires the validation team to perform validation in parallel, while this is not carried out in practice. As such, the use of this framework can be amended and tailored taking into account the purpose of its use, and the available resources of the project.

5.2. Directions for future research

The limitations of this framework (as outlined in Section 4.2.4) would not disqualify the framework, as the experts pointed out that these limitations do not prevent them from using this framework. Indeed, some of the highlighted limitations and challenges could be related to any form of validation and not specifically to this framework. As long as the limitations are thought through, discussed, and communicated, the risk would be low because the uncertainties in the results are known. Moreover, these limitations open up new avenues for research to propose and test solutions for these known challenges. It is worth mentioning that this research does not aim to conclude that this framework can be used in practice without further research and considerations, but rather a step forward in establishing confidence in the proposed framework. More research needs to be done to provide empirical evidence of the usefulness of this framework. Thus, this section proposed some related research directions.

One challenge mentioned by some interviewees is that the proposed tests and the guide questions do not provide clear guidance on how each test can be used in a real case study. While the framework aims to highlight key areas of focus, additional research may be necessary to clarify the use of each test. Thus, proposing a formalized technique for each test to tackle this challenge could be a fruitful future research direction. For instance, to benefit from the ‘convergent and concurrent validation’ tests, techniques for comparing systems based on complexity could be developed, or criteria for distinguishing similar or identical systems could be established.

The proposed validation framework was seen as too general by some of the interview participants, who find a need to customize it for each use case. The framework offers a high-level guide, and the authors agree that one validation framework is unlikely to work for all cases in various industrial contexts or for different applications of STPA. Ideally, each

project team or company needs to tailor the framework to their specific needs. For instance, two interviewees highlighted that they use STPA only for the early concept design, and from their point of view, the proposed validation framework is not in line with this application of STPA. Thus, developing a modular framework with associated guidance, which can be tailored to a specific practical context, could be a promising future research direction. This way, the relevance and applicability of specific validation tests can be evaluated on a case-by-case basis to ensure that the framework is effective for the particular project context.

The validity of the proposed validation framework is also another concern with using this framework. According to Goerlandt et al. (2017), the validity of the validation methods is a crucial aspect that needs to be thoroughly considered. This research study is a step in the direction of evaluating the validity of the framework. Another future research direction, which was also highlighted by Sadeghi and Goerlandt (2023b), is to perform a comparative case study research. Through such a study, for instance through an exercise in a laboratory setting as in Hulme et al. (2022), it can be investigated whether the proposed framework can actually achieve its envisioned goals. This can provide evidence for the effectiveness of validation.

Even if a formalized validation framework increases efficacy of an analysis, it is important to consider its practicality in terms of resource and time allocation. After all, usefulness without time- and cost-effectiveness is limited, as a technique with poor return on investment may actually reduce the overall effectiveness of a safety program (Rae et al., 2014). This is especially relevant for industry practitioners who often struggle to convince stakeholders of the value and necessity of validation, as found in Sadeghi and Goerlandt (2023a). To address this issue, it may be useful to draw on knowledge of human performance in time-critical work, which suggests that errors are more likely to occur when available time is limited (Hall et al., 1982). This knowledge could inform an investigation into the optimal allocation of resources to validate risk and hazard analyses in relation to the level of improvement achieved.

In addition, some interviewees noted that the ‘Behavior Validation’ test is usually carried out as part of the Validation and Verification activities within the systems engineering process, which primarily focuses on validating the design, rather than validating the safety analysis process itself. Future research could explore the interplay between STPA validation and Verification and Validation (V&V) activities, as an integrated set of processes, instead of solely validating STPA independently.

5.3. Limitations of the study

As with any qualitative research, the generalizability of the findings from this qualitative study should be considered. Despite a relatively small sample size, data saturation occurred (Guest et al., 2006). However, the purpose of qualitative research is not to generalize findings derived from selected samples to a wider population (Polit and Beck, 2010). In addition, the study’s reliance on voluntary participation made it susceptible to sampling bias (Cheung et al., 2017). Although efforts were made to include a diverse range of STPA experts, including both researchers and industry practitioners, it is not possible to assert that the study participants constituted a representative sample of all STPA experts. Further research is necessary to validate this. Additionally, it is important to note that the level of expertise of the STPA experts was inferred from their years of self-reported experience with the method, which may not fully represent their actual knowledge and skill with the method.

It is important to note that the aim of this study was not to demonstrate the effectiveness of the proposed validation framework in terms of increasing the safety of a system. It was intended to identify the similarities and differences between the proposed framework and the practices of STPA experts. As discussed in Section 5.2, further empirical work is required to empirically establish the effectiveness of the framework in producing improved outcomes.

6. Conclusions

In this study, we sought to evaluate the reasonableness of the STPA validation framework proposed by Sadeghi and Goerlandt (2023b) by seeking insights from STPA experts using an interview research methodology. Our findings revealed that all theory-based proposed validation tests have already been used in practice by at least one STPA expert, albeit informally. All experts see value in the proposed formalized validation framework, although some limitations and challenges were raised. In general, the interviewees found the framework to be too generic for use in all projects, indicating that it may need to be tailored to the specific context it is used for.

Recognizing that some of the highlighted challenges are relevant to any form of validation, we suggest performing more research focusing on validation in risk analysis and safety engineering in general, and the STPA validation framework in particular. Future research could propose and test solutions for challenges raised by experts concerning the presented framework, as well as investigate the framework further by applying it in laboratory experiments and in real-world case studies, comparing the results of an STPA analysis before and after its application. This comparison should be done both in terms of the quality of the STPA analysis, the credibility to decision-makers and stakeholders, and the possible ensuing effects on safety.

Commitment to evidence-based safety would strengthen the scientific basis of selecting, applying, and validating safety analysis techniques, and provide support for designing and implementing validation processes in particular contexts. Ultimately, our hope is that this study will contribute to the ongoing efforts to improve the practicality of the proposed STPA validation framework for risk and hazard analyses in various application domains.

List of Acronyms

V&V	Validation and Verification
SME	Subject Matter Expert
STPA	System Theoretic Process Analysis
STAMP	Systems-Theoretic Accident Model and Processes
MIT	Massachusetts Institute of Technology
UCA	Unsafe Control Action
ALARP	As Low As Reasonably Practicable
FMEA	Failure Mode and Effects Analysis
REB	Research Ethics Board

Appendix A. Interview Questions

Part 1: Collecting general information about the interviewees

- 1.1 Which sector are you working in?
- 1.2 What is your level of education? And what is your field of study?
- 1.3 How long have you been using STPA?

Part 2: Gathering information about the STPA experts' experience in STPA validation

Through the questions in this section, we aim to investigate how the STPA experts make sure that their STPA analysis is done well, that it provides comprehensive, accurate, and credible results. If the STPA experts have not performed any types of validation before, we ask them to imagine someone requested them to validate an STPA analysis and answer the questions below.

- 2.1 Do you validate your STPA analysis? Do you have a structured and formalized process for performing validation?
- 2.2 Do you have a separate team for performing validation?
- 2.3 Whom would/do you involve in the validation process? How do you decide on who should be involved in STPA validation?
- 2.4 Would/Do you validate the analysis in parallel with the STPA implementation? Or would/do you validate the analysis once the implementation is completed?

In the "Purpose of the analysis" step.

Author Contributions

Reyhaneh Sadeghi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Resources, Visualization, Writing – original draft. **Floris Goerlandt:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

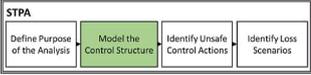
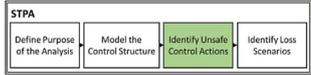
Data availability

The data that has been used is confidential.

Acknowledgments

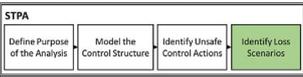
The work in this article has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The second author is furthermore supported by the Canada Research Chairs Program, through a grant by the Natural Sciences and Engineering Research Council (NSERC). We would like to thank the interviewed experts who contributed to this study by generously sharing their time, knowledge, and experiences. We are also indebted to two anonymous reviewers, whose insightful comments have been instrumental to improve an earlier version of this paper.

Table B1 (continued)

STPA Step	Main Activities	Validation Function	Validation Tests	Related Questions
			Subject Matter Experts (SMEs) Validation	<ol style="list-style-type: none"> 1. Is the process of SMEs selection, elicitation, and combination clearly and completely documented? 2. Is the SMEs selection systematically conducted? Are the SMEs selection criteria reasonable considering the purpose of the analysis? Is the entire system covered with appropriate knowledgeable experts? 3. Is the SMEs elicitation process systematically conducted? Is this process reasonable considering the purpose of the analysis? 4. If more than one SME is involved in the analysis, are the results of SME elicitation combined in a meaningful way?
			Assumption Validation	<ol style="list-style-type: none"> 1. Are the assumptions fully identified, accurately described, understood, documented, and agreed upon? 2. If the opposite of any particular assumption were true, would it have a substantial effect on the identified losses, system-level hazards, and system-levels constraint? 3. Are the degree of importance and certainty of each relevant assumption credibly determined?
2 <u>Model the Control Structure</u>	 <ol style="list-style-type: none"> 1. Controller a. Control Algorithm b. Process Model 2. Control Actions 3. Feedback 4. Controlled Processes 	Comprehensiveness and Accuracy	Content and structure Validation	<ol style="list-style-type: none"> 1. Does the created control structure include all relevant elements and system components? 2. Does the created control structure include all relevant functional relationships between these elements? 3. Is the control structure an accurate representation of the system? 4. Does the level of detail included in the control structure suffice for the purpose of the analysis?
			5. Concurrent Validation	<ol style="list-style-type: none"> 1. Has any STPA for an identical system been identified in the nomological validation? 2. If yes, is the developed control structure, including the controllers, controlled processes, control actions, and feedbacks, the same as the control structure in the identified STPA? If there are differences, why do these appear?
			1. Convergent Validation	<ol style="list-style-type: none"> 1. Has any STPA for a similar system been identified when conducting nomological validation? 2. If yes, how similar is the control structure to other control structures in the analysis of similar systems? If there are differences, why do these appear?
3 <u>Identify Unsafe Control Actions</u>	 <ol style="list-style-type: none"> 1. List of Unsafe Control Actions and Casual Factors 2. Controller Constraints A controller constraint specifies the controller behaviors that need to be satisfied to prevent UCAs. 	Comprehensiveness and Accuracy	Face Validation	<ol style="list-style-type: none"> 1. Are the identified UCAs logical and accurate from the validation team's perspective? 2. Are there any other possible UCAs that have not been identified by the STPA implementation team? 3. Are the identified UCAs accurately translated into constraints on the behavior of each controller?
			3. Extreme Condition Validation	<ol style="list-style-type: none"> 1. Are the identified constraints plausible for extreme and unlikely interactions of components within the system? 2. Are the identified constraints plausible for extreme and unlikely conditions in the system's environment?
			Behavior Validation	<ol style="list-style-type: none"> 1. How does the enforcement of the identified constraints change the behavior of the system? 2. Are the changes in the system's behavior as expected by the STPA implementation team?
4 <u>Identify Loss Scenarios</u>	<ol style="list-style-type: none"> 1. Identify scenarios lead to UCA 	Comprehensiveness and Accuracy	Face Validation	<ol style="list-style-type: none"> 1. Are the identified loss scenarios logical and accurate from the validation team's perspective?

(continued on next page)

Table B1 (continued)

STPA Step	Main Activities	Validation Function	Validation Tests	Related Questions
				<ol style="list-style-type: none"> Are all possible causal factors accounted for when identifying loss scenarios associated with the UCAs? Are the contributing factors of the previous incidents/accidents of the studied system covered in the identified scenarios? Are the contributing factors of the previous incidents/accidents of identical (sub-) systems identified in the nomological validation covered in the loss scenarios? Are the contributing factors of the previous incidents/accidents of the similar (sub-) systems identified in the nomological validation covered in the loss scenarios?
			<ol style="list-style-type: none"> Historical Validation Traceability Validation 	<ol style="list-style-type: none"> Can the identified loss scenarios be traced to all relevant UCAs, hazards, and losses? Can the identified losses in the first step of STPA be traced to all relevant hazards, UCAs, and scenarios? Are the traceabilities properly documented?
5 Final Results	<p>Once the analysis is finalized, the documentations should be put together and communicated with the stakeholders.</p> <ol style="list-style-type: none"> Solution Documentation Presentation of the result of STPA and validation to stakeholders 	Credibility	<ol style="list-style-type: none"> Documentation checking Review of Presentation of results 	<ol style="list-style-type: none"> Is the overall process behind the STPA implementation reasonably documented? Is the STPA documentation correct, clear, and complete? Is the documentation in formats understandable to users and stakeholders who may not be knowledgeable about STPA? Are the sources of uncertainty clearly documented? Are the limitations of the analysis clearly documented? Can the limitations be justified with regard to the purpose of the analysis? Does the presentation include the appropriate information regarding where and how stakeholders' interests (identified in the problem situation) are included in the analysis? Does the presentation clearly explain the sources of uncertainty? Does the presentation clearly explain the limitation of the analysis?

References

Alam, Md K., 2020. A systematic qualitative case study: questions, data collection, NVivo analysis and saturation. *Qual. Res. Org. Manag. Int. J.* 16 (1), 1–31. <https://doi.org/10.1108/QROM-09-2019-1825>.

Arnold, R., 2009. *A Qualitative Comparative Analysis of Soam and Stamp In Atm Occurrence Investigation*. Lund University.

Aven, T., Guikema, S., 2011. Whose uncertainty assessments (probability distributions) does a risk assessment report: the analysts' or the experts. *Reliab. Eng. Syst. Saf.* 96 (10), 1257–1262. <https://doi.org/10.1016/j.res.2011.05.001>.

Aven, T., Korte, J., 2003. On the use of risk and decision analysis to support decision-making. *Reliab. Eng. Syst. Saf.* 79 (3), 289–299. [https://doi.org/10.1016/S0951-8320\(02\)00203-X](https://doi.org/10.1016/S0951-8320(02)00203-X).

Baybutt, P., 2018. The validity of engineering judgment and expert opinion in hazard and risk analysis: the influence of cognitive biases. *Process Saf. Prog.* 37 (2), 205–210. <https://doi.org/10.1002/prs.11906>.

Baybutt, P., 2021. On the need for system-theoretic hazard analysis in the process industries. *J. Loss Prev. Process. Ind.* 69, 104356 <https://doi.org/10.1016/j.jlp.2020.104356>.

Bowen, G.A., 2008. Naturalistic inquiry and the saturation concept: a research note. *Qual. Res.* 8 (1), 137–152. <https://doi.org/10.1177/1468794107085301>.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. *Qual. Res. Psychol.* 3 (2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>.

Chaal, M., Bahootoroody, A., Basnet, S., Valdez Banda, O.A., Goerlandt, F., 2022. Towards system-theoretic risk assessment for future ships: a framework for selecting Risk Control Options. *Ocean Eng.* 259, 111797 <https://doi.org/10.1016/j.oceaneng.2022.111797>.

Cheung, K.L., ten Klooster, P.M., Smit, C., de Vries, H., Pieterse, M.E., 2017. The impact of non-response bias due to sampling in public health studies: a comparison of

voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Publ. Health* 17 (1), 276. <https://doi.org/10.1186/s12889-017-4189-8>.

Dallat, C., Salmon, P.M., Goode, N., 2019. Risky systems versus risky people: to what extent do risk assessment methods consider the systems approach to accident causation? A review of the literature. *Saf. Sci.* 119, 266–279. <https://doi.org/10.1016/j.ssci.2017.03.012>.

Dekker, S., 2014. *The Field Guide to Understanding "Human Error"*, third ed. CRC Press.

Elston, D.M., 2021. *Survivorship Bias*. June 18). <https://www.sciencedirect.com/science/article/pii/S0190962221019861>.

Ericson, C., 2005. *Hazard Analysis Techniques for System Safety (1. Aufl.)*. Wiley-Interscience.

Fleming, C.H., Leveson, N.G., 2014. Improving hazard analysis and certification of integrated modular avionics. *J. Aero. Inf. Syst.* 11 (6), 397–411. <https://doi.org/10.2514/1.1010164>.

Gaskell, G., 2000. Individual and group interviewing. In: *Qualitative Researching with Text, Image and Sound*. SAGE Publications Ltd, pp. 38–56. <https://doi.org/10.4135/9781849209731>.

Goerlandt, F., Khakzad, N., Reniers, G., 2017. Validity and validation of safety-related quantitative risk analysis: a review. *Saf. Sci.* 99, 127–139. <https://doi.org/10.1016/j.ssci.2016.08.023>.

Guest, G., Bunce, A., Johnson, L., 2006. How many interviews are enough?: an experiment with data saturation and variability. *Field Methods* 18 (1), 59–82. <https://doi.org/10.1177/1525822X05279903>.

Hall, R.E., Fragola, J., Wreathall, J., 1982. *Post-Event Human Decision Errors: Operator Action Tree/time Reliability Correlation*, p. 48.

Harkleroad, E., Vela, A., Kuchar, J., 2013. *Review of Systems-Theoretic Process Analysis (STPA) Method and Results to Support NextGen Concept Assessment and Validation (ATC-427)*.

Hulme, A., Stanton, N.A., Walker, G.H., Waterson, P., Salmon, P.M., 2022. Testing the reliability and validity of risk assessment methods in Human Factors and

- Ergonomics. *Ergonomics* 65 (3), 407–428. <https://doi.org/10.1080/00140139.2021.1962969>.
- Hurst, J., McIntyre, J., Tamauchi, Y., Kinuhata, H., Kodama, T., 2019. A summary of the 'ALARP' principle and associated thinking. *J. Nucl. Sci. Technol.* 56 (2), 241–253. <https://doi.org/10.1080/00223131.2018.1551814>.
- Kuzel, A.J., 1992. Sampling in qualitative inquiry. In: *Doing Qualitative Research*, vol. 3. Sage Publications, Inc, pp. 31–44.
- Landry, M., Malouin, J.-L., Oral, M., 1983. Model validation in operations research. *European Journal of Operational Research* 14 (3), 207–220. [https://doi.org/10.1016/0377-2217\(83\)90257-6](https://doi.org/10.1016/0377-2217(83)90257-6).
- Langdalen, H., Abrahamsen, E.B., Selvik, J.T., 2020. On the importance of systems thinking when using the ALARP principle for risk management. *Reliab. Eng. Syst. Saf.* 204, 107222 <https://doi.org/10.1016/j.res.2020.107222>.
- Leveson, 2012. *Engineering a Safer World: Systems Thinking Applied to Safety*. The MIT Press, Cambridge, Mass. https://web-s-ebcoshost-com.ezproxy.library.dal.ca/ehost/bookviewer/ebook/ZTAwMHhuYV9fNDIxODE4X19BTg2?sid=e9969089-f149-426b-bb6e-776f0eca0b81@redis&vid=0&format=EB&lpid=lp_1&rid=0.
- Leveson, N., 2015. A systems approach to risk management through leading safety indicators. *Reliab. Eng. Syst. Saf.* 136, 17–34. <https://doi.org/10.1016/j.res.2014.10.008>.
- Leveson, N., Thomas, J., 2018. STPA Handbook. https://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf.
- Martínez, R.S., 2015. *System Theoretic Process Analysis of Electric Power Steering for Automotive Applications*. Massachusetts Institute of Technology.
- Mason, R., Mitroff, I., 1981. *Challenging strategic planning assumptions: Theory, cases, and techniques*. Wiley.
- Möller, N., Ove Hansson, S., Holmberg, J.-E., Rollenhagen, C., 2018. *Handbook of Safety Principles*. John Wiley & Sons, Incorporated. <http://ebookcentral.proquest.com/lib/dal/detail.action?docID=5216287>.
- Norton, M.I., Mochon, D., Ariely, D., 2012. The IKEA effect: when labor leads to love. *J. Consum. Psychol.* 22 (3), 453–460. <https://doi.org/10.1016/j.jcps.2011.08.002>.
- Oberkampf, W.L., Trucano, T.G., 2008. Verification and validation benchmarks. *Nucl. Eng. Des.* 238 (3), 716–743. <https://doi.org/10.1016/j.nucengdes.2007.02.032>.
- Pasman, H.J., Rogers, W.J., 2020. How to treat expert judgment? With certainty it contains uncertainty. *J. Loss Prev. Process. Ind.* 66, 104200 <https://doi.org/10.1016/j.jlpp.2020.104200>.
- Patriarca, R., Chatzimichailidou, M., Karanikas, N., Di Gravio, G., 2022. The past and present of System-Theoretic Accident Model and Processes (STAMP) and its associated techniques: a scoping review. *Saf. Sci.* 146, 105566 <https://doi.org/10.1016/j.ssci.2021.105566>.
- Polit, D.F., Beck, C.T., 2010. Generalization in quantitative and qualitative research: myths and strategies. *Int. J. Nurs. Stud.* 47 (11), 1451–1458. <https://doi.org/10.1016/j.ijnurstu.2010.06.004>.
- QSR International Pty Ltd, 2020. NVivo (released in March 2020). <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>.
- Rae, A., Alexander, R., 2017a. Forecasts or fortune-telling: when are expert judgements of safety risk valid? *Saf. Sci.* 99, 156–165. <https://doi.org/10.1016/j.ssci.2017.02.018>.
- Rae, A., Alexander, R.D., 2017b. Probative blindness and false assurance about safety. *Saf. Sci.* 92, 190–204. <https://doi.org/10.1016/j.ssci.2016.10.005>.
- Rae, A., Alexander, R., McDermid, J., 2014. Fixing the cracks in the crystal ball: a maturity model for quantitative risk assessment. *Reliab. Eng. Syst. Saf.* 125, 67–81. <https://doi.org/10.1016/j.res.2013.09.008>.
- Rosqvist, T., 2010. On the validation of risk analysis—a commentary. *Reliab. Eng. Syst. Saf.* 95 (11), 1261–1265. <https://doi.org/10.1016/j.res.2010.06.002>.
- Sadeghi, R., Goerlandt, F., 2021. The state of the practice in validation of model-based safety analysis in socio-technical systems: an empirical study. *Saf. Now.* 7 (4) <https://doi.org/10.3390/safety7040072>. Article 4.
- Sadeghi, R., Goerlandt, F., 2023a. Validation of system safety hazard analysis in safety-critical industries: an interview study with industry practitioners. *Saf. Sci.* 161, 106084 <https://doi.org/10.1016/j.ssci.2023.106084>.
- Sadeghi, R., Goerlandt, F., 2023b. A proposed validation framework for the system theoretic process analysis (STPA) technique. *Saf. Sci.* 162, 106080 <https://doi.org/10.1016/j.ssci.2023.106080>.
- Sandelowski, M., 1995. Sample size in qualitative research. *Res. Nurs. Health* 18 (2), 179–183. <https://doi.org/10.1002/nur.4770180211>.
- Sandelowski, M., 2004. *Using Qualitative Research* 14 (10). <https://doi.org/10.1177/1049732304269672>.
- Sultana, S., Okoh, P., Haugen, S., Vinnem, J.E., 2019. Hazard analysis: application of STPA to ship-to-ship transfer of LNG. *J. Loss Prev. Process. Ind.* 60, 241–252. <https://doi.org/10.1016/j.jlpp.2019.04.005>.
- Thomas, J., de Lemos, F.L., Leveson, N., 2012. Evaluating the Safety of Digital Instrumentation and Control Systems in Nuclear Power Plants, p. 66. <http://sunnyday.mit.edu/papers/MIT-Research-Report-NRC-7-28.pdf>.
- Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. *Science* 185 (4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>.
- van der Helm, R., 2006. Towards a clarification of probability, possibility and plausibility: how semantics could help futures practice to improve. *Foresight* 8 (3), 17–27. <https://doi.org/10.1108/14636680610668045>.
- Ventikos, N.P., Chmurski, A., Louzis, K., 2020. A systems-based application for autonomous vessels safety: hazard identification as a function of increasing autonomy levels. *Saf. Sci.* 131, 104919 <https://doi.org/10.1016/j.ssci.2020.104919>.
- Wróbel, K., Montewka, J., Kujala, P., 2018. System-theoretic approach to safety of remotely-controlled merchant vessel. *Ocean Eng.* 152, 334–345. <https://doi.org/10.1016/j.oceaneng.2018.01.020>.