Article

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

Reyhaneh Sadeghi *[ID] 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

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.
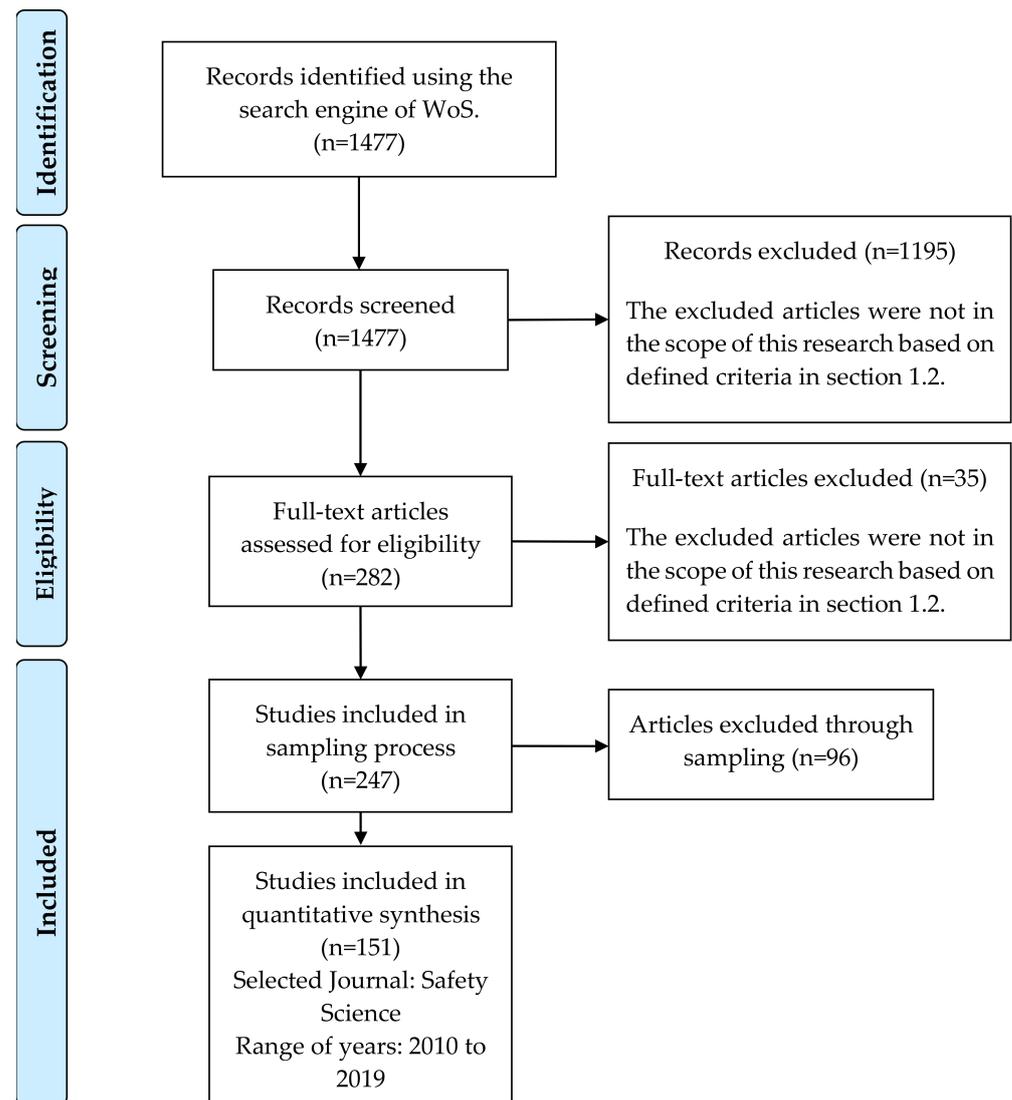

Figure 1. The process of constructing the dataset based on PRISMA Flow Diagram.

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 = (("Safety" OR "Risk" OR "Reliability" OR "Resilience") AND ("Model" OR "Method" OR "Approach" OR "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].
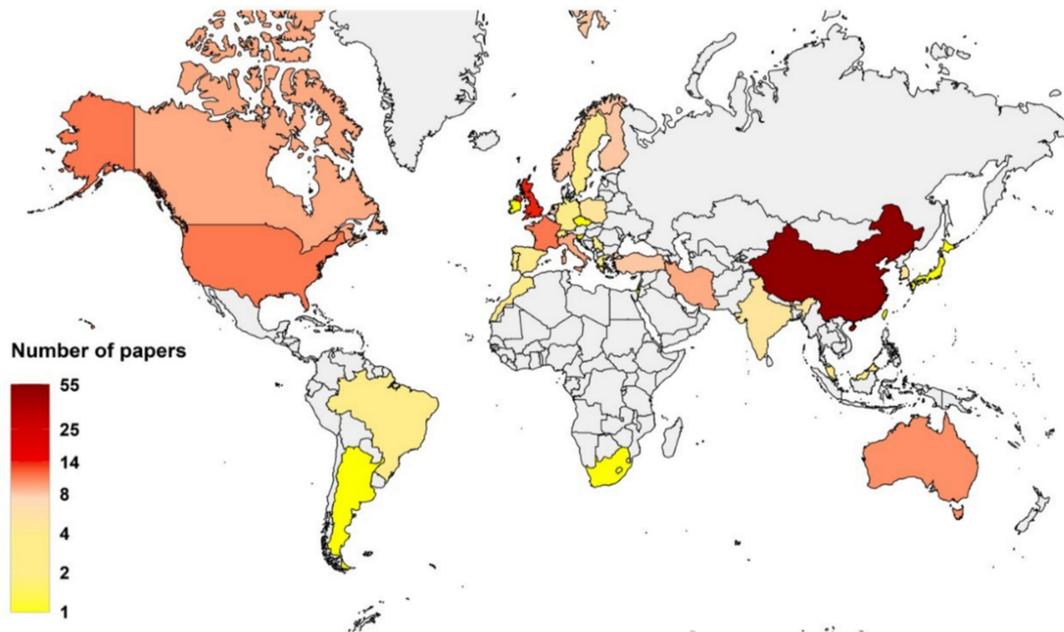

Figure 2. Geographical distribution of the articles in the data sample, period 2010–2019.

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].
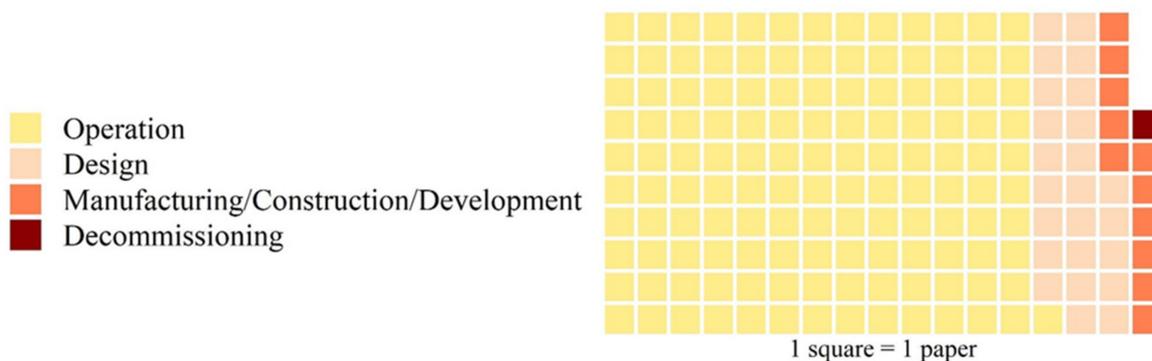

Figure 3. Waffle chart, distribution of the papers in terms of stage of the system life cycle.

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.
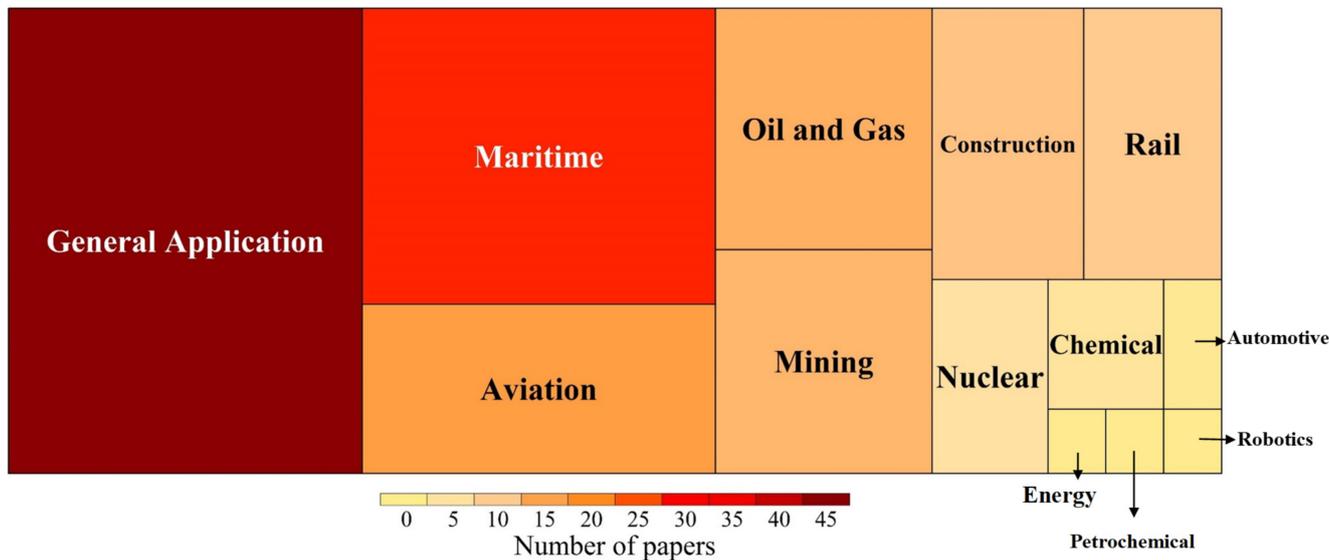

Figure 4. Distribution of papers considering industrial application domain.

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 Wienen 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 Wienen 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*.
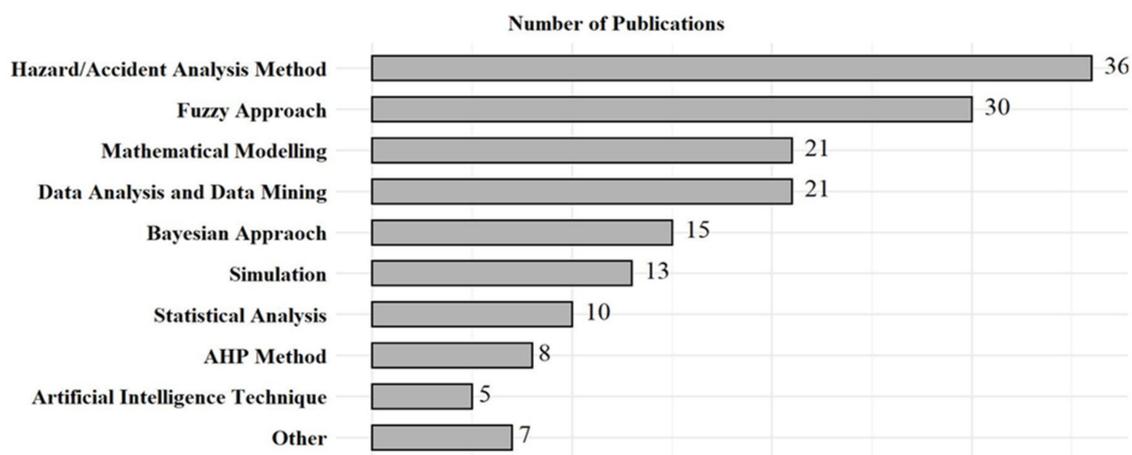

Figure 5. Number of publications in each model type.

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.
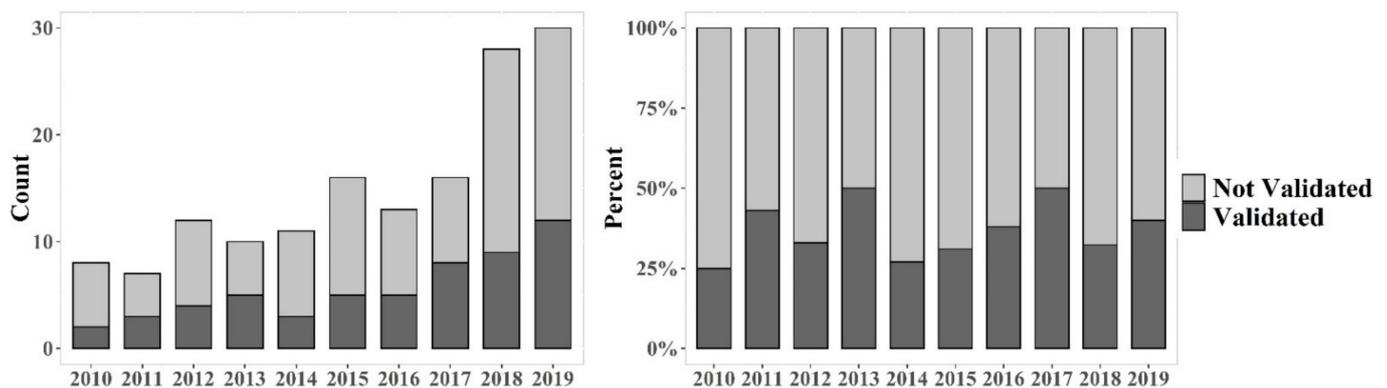

Figure 6. Count and the percentage of the papers in which the models are validated and the papers in which the models are not validated over the 10 years period.

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.
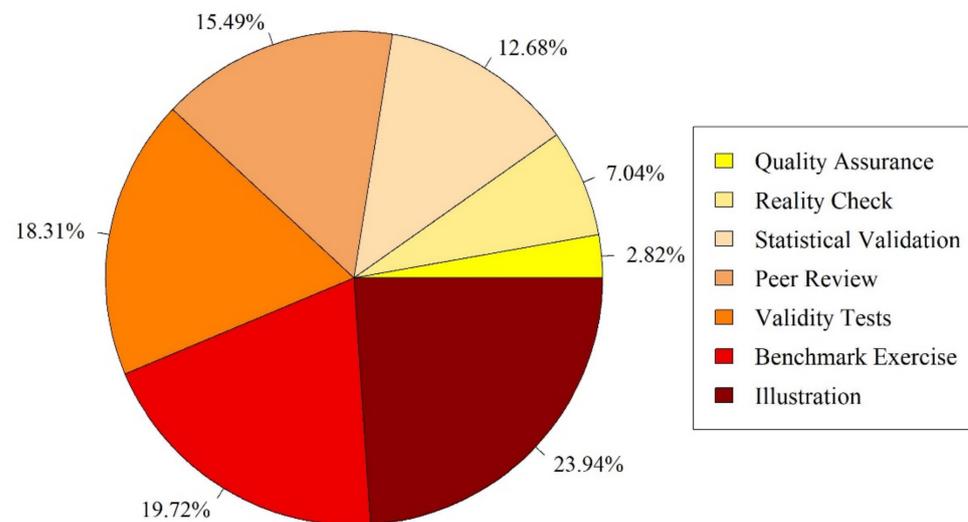

Figure 7. Pie chart-showing the distribution of articles in terms of the adopted validation approach, for cases where validation is performed.

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	*p*-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.
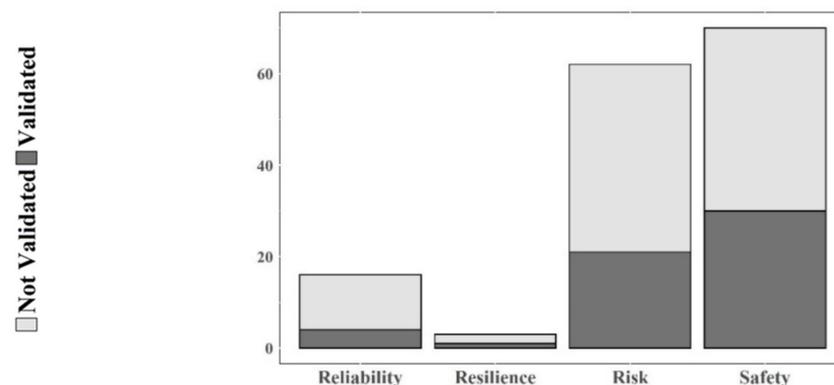
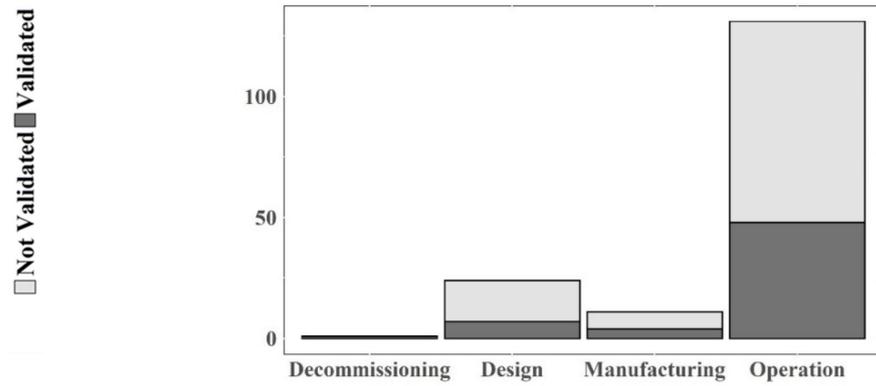**Figure 8.** Stacked bar plot showing validation versus safety concept.

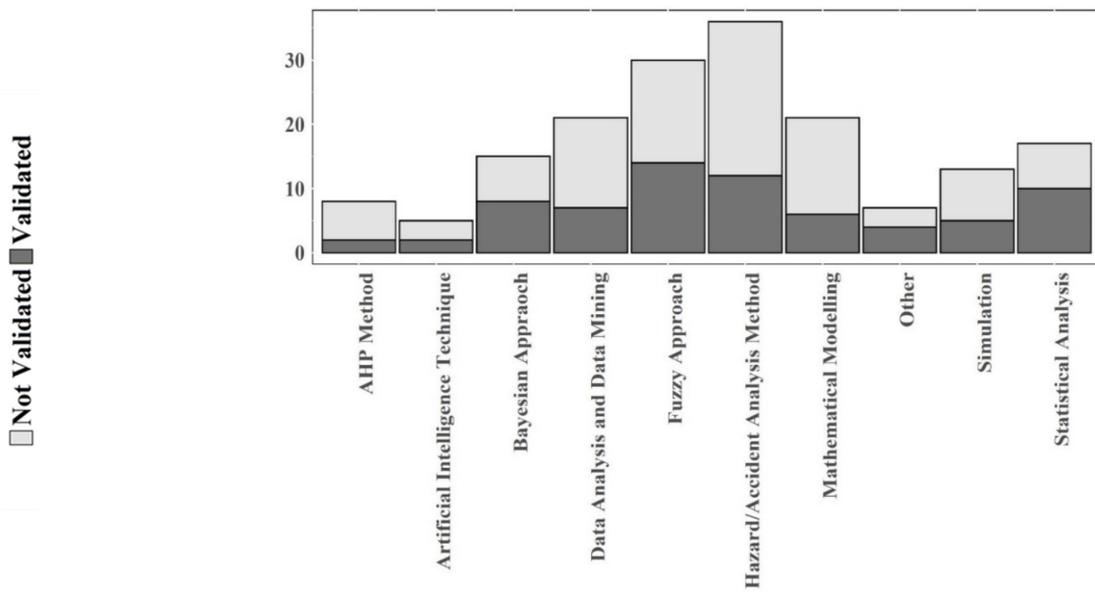**Figure 9.** Stacked bar plot for validation versus stage of the system lifecycle.

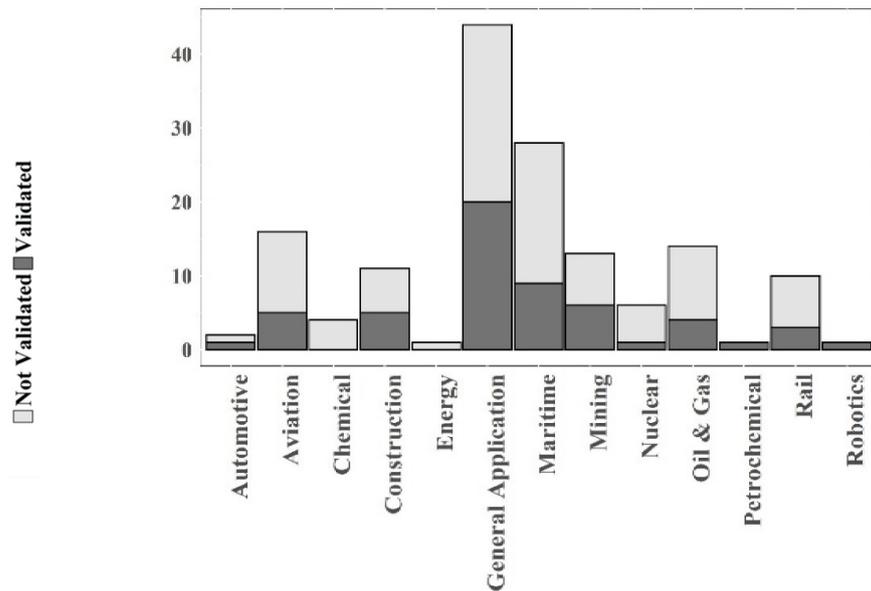**Figure 10.** Stacked bar plot for validation and model type/approach.

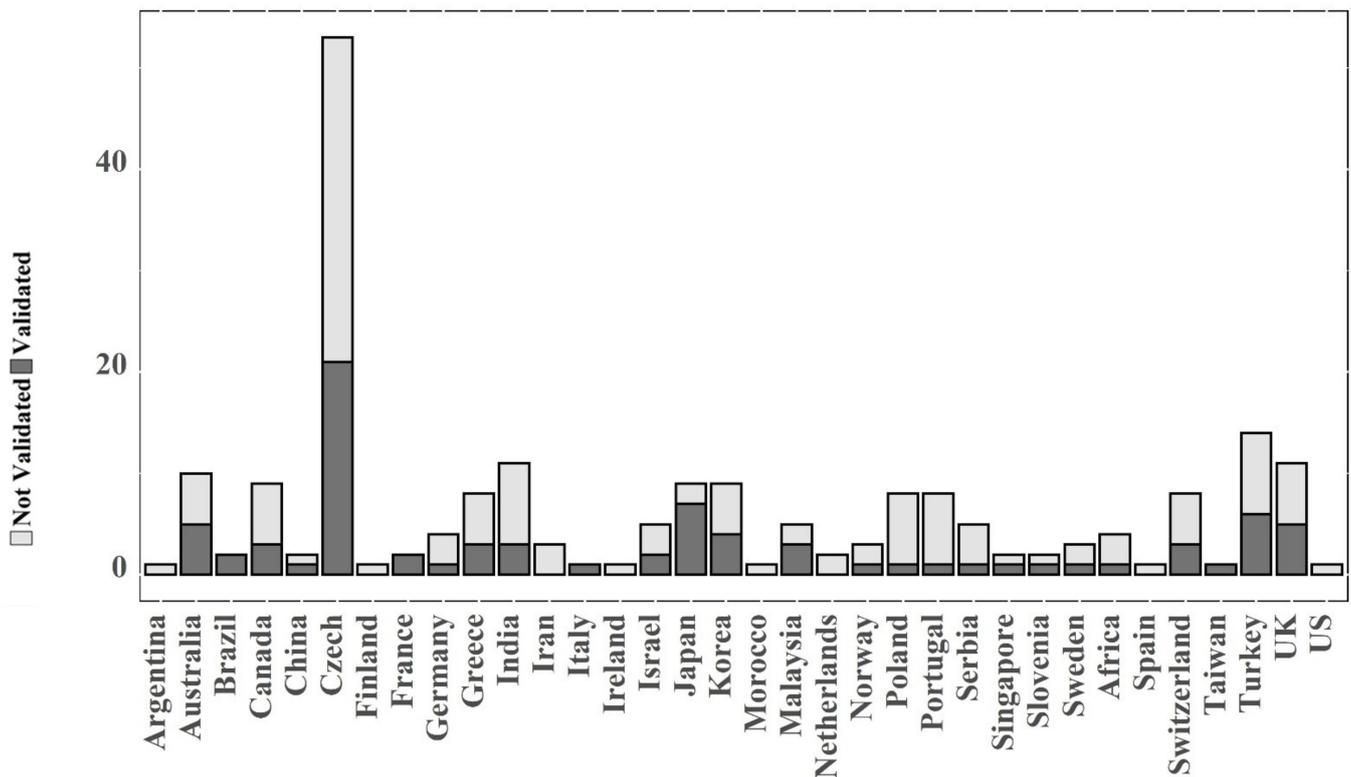**Figure 11.** Stacked bar plot for validation and industry.

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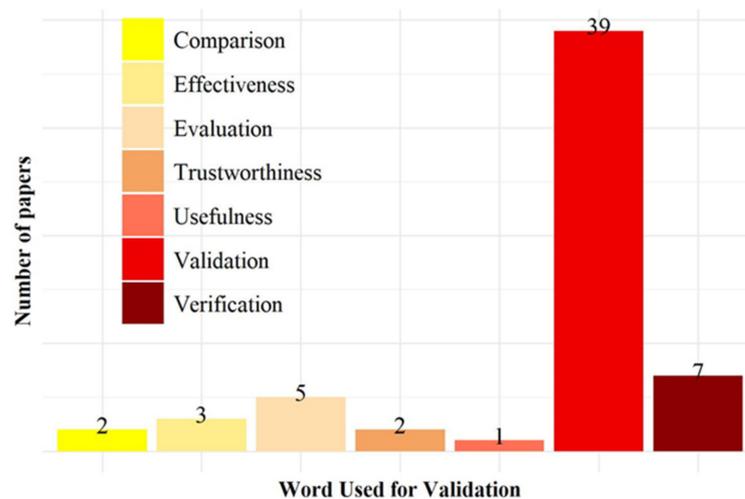
Figure 13. Distribution of papers in terms of the terminology used for validation.

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.

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	• Safety • Risk	• Reliability • Resilience
Year of publication	This ranges from 2010 to 2019.	
Country of origin	• Argentina • Australia • Brazil • Canada • China • Czech Republic • Finland • France • Germany • Greece • India • Iran • Ireland • Israel • Italy • Japan • Malaysia	• Morocco • The Netherlands • Norway • Poland • Portugal • Serbia • Singapore • Slovenia • South Africa • South Korea • Spain • Sweden • Switzerland • Taiwan • Turkey • United Kingdom • United States
Stage of the system life cycle	• Design Phase • Manufacturing/Construction/Development Phase • Operation Phase • Decommissioning Phase	
Industrial application domain	• Automotive Industry • Aviation Industry • Chemical Industry • Construction Industry • Energy Industry • General Application • Maritime Industry	• Mining Industry • Nuclear Industry • Oil and gas Industry • Petrochemical Industry • Rail Industry • Robotics Industry
Model type/approach	• Artificial Intelligence Technique • Accident Analysis Method • AHP Approach • Bayesian Approach • Data Analysis and Data Mining • Fuzzy Approach	• Mathematical Modeling • Simulation • Statistical Analysis • Other

Table A1. *Cont.*

Variables	Categories	
Validation approach	• Benchmark Exercise • Reality Check • Peer Review • Quality Assurance	• Validity Tests • Statistical Validation • Illustration • No Validation
Word used for validation	• Validation • Evaluation • Verification • Comparison	• Usefulness • Trustworthiness • Effectiveness

References

1. 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).
2. Rae, A.; Nicholson, M.; Alexander, R. The State of Practice in System Safety Research Evaluation. In Proceedings of the 5th IET Internatioanl 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).
3. 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.
4. 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.
5. Aven, T. Foundational Issues in Risk Assessment and Risk Management. *Risk Anal.* **2012**, *32*, 1647–1656. [CrossRef]
6. Goerlandt, F.; Khakzad, N.; Reniers, G. Special Issue: Risk Analysis Validation and Trust in Risk management. *Saf. Sci.* **2017**, *99*, 123–126. [CrossRef]
7. Hale, A. Foundations of safety science: A postscript. *Saf. Sci.* **2014**, *67*, 64–69. [CrossRef]
8. 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]
9. Goncalves Filho, A.P.; Waterson, P. Maturity models and safety culture: A critical review. *Saf. Sci.* **2018**, *105*, 192–211. [CrossRef]
10. 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]
11. 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]
12. Suokas, J.; Kakko, R. On the problems and future of safety and risk analysis. *J. Hazard. Mater.* **1989**, *21*, 105–124. [CrossRef]
13. 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]
14. Laheij, G.M.H.; Ale, B.; Post, J.G. Benchmark risk analysis models used in The Netherlands. *Saf. Reliab.* **2003**, 993–999.
15. Barlas, Y. Multiple tests for validation of system dynamics type of simulation models. *Eur. J. Oper. Res.* **1989**, *42*, 59–87. [CrossRef]
16. 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.
17. Eker, S.; Rovenskaya, E.; Langan, S.; Obersteiner, M. Model validation: A bibliometric analysis of the literature. *Environ. Model. Softw.* **2019**, *117*, 43–54. [CrossRef]
18. Le Coze, J.-C.; Pettersen, K.; Reiman, T. The foundations of safety science. *Saf. Sci.* **2014**, *67*, 1–5. [CrossRef]
19. 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]
20. 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]
21. 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]
22. Wybo, J.-L. Track circuit reliability assessment for preventing railway accidents. *Saf. Sci.* **2018**, *110*, 268–275. [CrossRef]
23. 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.
24. 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.; Anthone, 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. Bhattacherjee, A. *Social Science Research: Principles, Methods, and Practices*, 2nd ed.; Anol Bhattacherjee: 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.; Vriezekolk, 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 'evaludation': 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]