Validity and validation of safety-related quantitative risk analysis: A review

Goerlandt, Floris; Khakzad Rostami, Nima; Reniers, Genserik

DOI
10.1016/j.ssci.2016.08.023

Publication date
2016

Document Version
Proof

Published in
Safety Science

ARTICLE IN PRESS

Safety Science xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Validity and validation of safety-related quantitative risk analysis: A review

Floris Goerlandt [a,b,*], Nima Khakzad [c], Genserik Reniers [b,c,d]

[a] Aalto University, School of Engineering, Marine Technology, Research Group on Maritime Risk and Safety, P.O. Box 14300, 00076 Aalto, Finland
[b] KU Leuven, Campus Brussels, Faculty of Economics and Organizational Sciences, Center for Economics and Corporate Sustainability, Warmoesberg 26, 1000 Brussels, Belgium
[c] TU Delft, Safety and Security Science Group, Jaffalaan 5, 2628 BX Delft, The Netherlands
[d] University of Antwerp, City Campus, Faculty of Applied Economic Sciences, Prinsstraat 13, 2000 Antwerp, Belgium

ARTICLE INFO

Article history:
Received 6 July 2016
Received in revised form 12 August 2016
Accepted 21 August 2016
Available online xxxx

Keywords:
Risk analysis
Validation
Foundational issues
Quantitative risk analysis
QRA

ABSTRACT

Quantitative risk analysis (QRA) is widely applied in several industries as a tool to improve safety, as part of design, licensing or operational processes. Nevertheless, there is much less academic research on the validity and validation of QRA, despite their importance both for the science of risk analysis and with respect to its practical implication for decision-making and improving system safety. In light of this, this paper presents a review focusing on the validity and validation of QRA in a safety context. Theoretical, methodological and empirical contributions in the scientific literature are reviewed, focusing on three questions. Which theoretical views on validity and validation of QRA can be found? Which features of QRA are useful to validate a particular QRA, and which frameworks are proposed to this effect? What kinds of claims are made about QRA, and what evidence is available for QRA being valid for the stated purposes? A discussion follows the review, focusing on the available evidence for the validity of QRA and the effectiveness of validation methods.

1. Introduction

Risk analyses are used in many application areas, and many frameworks, models and specific applications have been presented in the scientific literature. Quantitative risk analysis (QRA)[1] is included in many textbooks (Aven, 2008; Bedford and Cooke, 2001; Meyer and Reniers, 2013), is considered an important topic to teach engineering students and health and safety professionals (Saleh and Pendley, 2012; Wybo and Van Wassenhove, 2016), and is used in many application areas (Marhavilas et al., 2011). Reviews of risk analysis methods in specific application areas indicate that QRA is applied, inter alia, for nuclear installations (Garrick and Christie, 2002), offshore oil and gas platforms (Vinnem, 1998), maritime transportation in waterways (Li et al., 2012), chemical installations (Khan et al., 2015) and related land use planning (Pasman and Reniers, 2014), in the construction industry (Taroun, 2014) and for cyber security (Cherdantseva et al., 2016).

On a more fundamental level, some authors have raised the issue about the general lack of attention to validation in risk research. In an early commentary, Cumming (1981) points out that quality control procedures for risk analysis methods are not well developed, while this is an important problem which must be dealt with. Aven and Heide (2009) and Rosqvist (2010) also note the limited attention to validity and validation in risk analysis, whereas Pasman et al. (2009) find that quality criteria for QRA shall be internationally established. More recently, Rae et al. (2014) formulate the issue as follows "[...] the combination of little empirical study with little natural feedback [...] leaves us in almost total darkness as to the validity and efficacy of QRA". To the best of the authors' knowledge, the only comprehensive review made regarding risk analysis validation dates from almost three decades ago, by Suokas and Rouhiainen (1989).

If, as argued by Hansson and Aven (2014), risk analysis is a discipline in its own right rather than a trans-scientific, interdisciplinary practice as interpreted by Weinberg (1981) or a "scientistic" approach[2] as suggested by Reid (2009), there should be warrants about the scientific validity of QRA. In the understanding

* Corresponding author at: Aalto University, School of Engineering, Marine Technology, Research Group on Maritime Risk and Safety, P.O. Box 14300, 00076 Aalto, Finland.
E-mail address: floris.goerlandt@aalto.fi (F. Goerlandt).

[1] In different industries, other terminology is applied, including PRA (probabilistic risk assessment), PSA (probabilistic safety assessment) and FSA (formal safety assessment). All these are referred to as QRA in this paper.

[2] According to Reid (2009), a scientistic approach is "an approach that makes use of scientific notions, images and methods, not only for the purposes of scientific enquiry, but also for the purpose of invoking the credibility, prestige and authority of 'scientific' knowledge to support an argument or to promote a point of view".

http://dx.doi.org/10.1016/j.ssci.2016.08.023
0925-7535/© 2016 The Authors. Published by Elsevier Ltd.

of foundational issues in risk analysis by Aven and Zio (2014), concepts and principles for establishing validity, and frameworks and methods for validating risk analysis methods and their results are important elements for strengthening the scientific foundations of the discipline. Irrespective of whether or not risk analysis is a science, considering QRA as an engineering method, the validity of the method with respect to its purpose, as well as procedures for establishing this, are important issues, both for system designers (Rae et al., 2014) and for regulators (Kirchsteiger, 1999). Nonetheless, e.g. in important regulatory documents concerning risk analysis, like the Seveso Directive (Seveso III, 2012), no specific requirements are stated concerning risk analysis quality control procedures.

In light of the above, the purpose of this paper is to make a review of the state of the art concerning the validity and validation of QRA. Focus is on QRA in a safety context, i.e. where QRA is used in a context of major accidents, engineering design for safety, or personal safety. More specifically, focus is on contributions in the scientific literature where validity or validation is considered as a research topic in itself, or is dealt with as part of a proposed framework for performing quantitative risk analysis. Hence, validation exercises of specific applications are outside the current scope. This scope limitation is made for practical reasons because risk analysis is a very wide research area, and for methodological reasons as it is important to delineate the scope of work, as found also e.g. by Li and Hale (2015). Nonetheless, it is noted that work has recently also been dedicated to validation of security risk models (Zhuang et al., 2016) and occupational safety risk assessment (Pinto et al., 2013), showing the relevance of validation in risk analysis as a topic for research and discussion.

The main questions addressed in this review are as follows:

- Which theoretical views on validity and validation of QRA can be found?
- Which features of QRA are necessary to distinguish a "good" from a "bad" QRA?
- What frameworks or methods have been proposed to validate a particular QRA?
- What kinds of claims are made about QRA, and what evidence is available for QRA being valid for the stated purposes?

A note on terminology is in place, distinguishing two uses of the term "validity" and the related term "validation". Conceptual validity is understood here as the condition where an operationalisation of a concept measures what it intends to measure. This understanding is in line with validity e.g. in social sciences, see Trochim and Donnely (2008), and is applied e.g. by Aven and Heide (2009) in questioning whether QRA fulfils the scientific criteria of reliability and validity. Pragmatic validity is the condition where a method meets its intended requirements in terms of the results obtained, as understood e.g. by Rae et al. (2014) in questioning the efficacy of QRA. Validation is understood here as the process by which validity is established, noting that different authors apply different terminology for this process, e.g. verification (Graham, 1995), quality control/qualification (Rosqvist and Tuominen, 2004; Suokas and Rouhiainen, 1989), credibility assessment (Busby and Hughes, 2006) and evaluation (Goerlandt and Montewka, 2015).

The review method is first described in Section 2, listing the materials considered relevant for the purposes of the paper. Subsequently, theoretical contributions are outlined in Section 3, and methodological and empirical contributions addressing different approaches to validation in Section 4. Section 5 describes a number of other contributions related to validation of QRA. In Section 6, a discussion is made, focusing on the effectiveness of the different approaches to validation and on the evidence for the claims made about QRA. Section 7 concludes.

2. Review method

In traditional indexing systems such as Scopus and Web of Science, risk analysis is not considered as a separate category in the scientific research areas. Instead, contributions related to risk are typically listed under "mathematics", "social sciences" or "engineering". Hence, general searches in those systems on terms like "risk analysis", "validation" and "QRA" results in very many hits, with low relevance to the above stated aims. Therefore, another review method has been applied, focusing on specific journals publishing papers on risk analysis methods, quantitative risk analysis and the foundations of risk analysis.

To identify these journals, a comparable method was used as in Reniers and Anthone (2012): an internet search was performed for journals based on the keywords "risk", "risk analysis", "risk assessment", "risk management" and "safety". In addition, an online journal ranking tool (SJR, 2016) was used to identify possibly relevant journals, based on a search for the keywords "risk" and "safety" in the journal title. In addition, a list of the top 35 safety-related journals in a world-wide ranking reported by Reniers and Anthone (2012), was considered to contain possibly relevant journals. Together, these searches resulted in a draft list of journals, of which, based on a description of the aims and scope on the journal websites, only journals covering safety-related risk analysis were retained. A double-check was performed using an analysis by Li and Hale (2016), ensuring that safety-related journals containing risk assessment clusters, were included. Table 1 shows the final list of considered journals.

Subsequently, acknowledging the different terminologies used in risk research for key concepts, articles were searched in these

Table 1
Journals considered in the literature review.

Journal title	Abbr.
Accident Analysis and Prevention	AAP
ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering	AJRUA
ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering	AJRUB
Disaster Prevention and Management	DPM
Fire Safety Journal	FSJ
Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards	GR
Human and Ecological Risk Assessment: An International Journal	HERA
Injury Prevention	IJ
International Journal of Business Continuity and Risk Management	IJBCRM
International Journal of Disaster Risk Reduction	IJDDR
International Journal of Disaster Resilience in the Built Environment	IJDRBE
International Journal of Reliability, Quality and Safety Engineering	IJRQSE
International Journal of Risk Assessment and Management	IJRAM
Japanese Journal of Risk Analysis	JJRA
Journal of Contingencies and Crisis Management	JCCM
Journal of Flood Risk Management	JFRM
Journal of Hazardous Materials	JHM
Journal of Loss Prevention in the Process Industries	JLPPI
Journal of Occupational Safety and Ergonomics	JOSE
Journal of Risk & Reliability: Proceedings of the Institution of Mechanical Engineers Part O	PIMEPO
Journal of Risk Analysis and Crisis Response	JRARC
Journal of Risk Research	JRR
Journal of Safety Research	JSR
Open Occupational Health and Safety Journal	OOHSJ
Process Safety and Environmental Protection	PSEP
Process Safety Progress	PSP
Reliability Engineering and System Safety	RESS
Reliability: Theory & Applications	RTA
Risk Analysis	RA
Risk and Decision Analysis	RDA
Safety Science	SS
Stochastic Environmental Research and Risk Assessment	SERRA
Structural Safety	STS

Please cite this article in press as: Goerlandt, F., et al. Validity and validation of safety-related quantitative risk analysis: A review. Safety Sci. (2016), http://dx.doi.org/10.1016/j.ssci.2016.08.023

journals based on a combination of two keywords, one related to "risk" and one related to "validity/validation". The first set of keywords is {"risk analysis", "risk assessment"}, the second set {"validity", "validation", "verification", "credibility", "evaluation", "qualification"}. The full publication history of these journals was covered. The resulting sets of papers were manually screened in light of the review scope and defined research questions, see Section 1. This was done based on the title, abstract and full text, if the latter was readily available and the information in the abstract provided too little information. Finally, the reference lists of relevant articles were scrutinized, relevant articles identified and added to the list. Similarly, using Scopus and Web of Science indexing systems, it was traced which articles had cited the identified relevant articles, and further relevant articles added to the list. All searches described in this paragraph were executed during April 2016.

Subsequently, the materials were grouped in three clusters. A first cluster concerns theoretical contributions, i.e. papers which focus on issues such as how validation and validity is understood from a conceptual or foundational point of view. A second cluster concerns methodological and empirical contributions, addressing work where specific approaches to validation are proposed and/or applied. A third cluster contains work where validation is addressed, without specifically proposing or applying a method.

3. Theoretical contributions

The theoretical contributions related to the validity of QRA are divided into three categories: conceptual, foundational and pragmatic. Section 3.1 addresses the conceptual validity, i.e. in how far QRA succeeds in measuring the concept it intends to describe. Section 3.2 outlines different views on validation in relation to different scientific foundations adopted in a risk analysis. Section 3.3 focuses on pragmatic validity, outlining different uses suggested for QRA and claims about QRA corresponding to these uses. In Section 3.4, an overview is given of generic methods to establish pragmatic validity of a given QRA.

3.1. Conceptual validity

Aven and Heide (2009) question in how far QRA meets the scientific requirements of reliability and validity, adopting an implicit conceptualization of risk as "uncertainty about and severity of the events and consequences of an activity with respect to something that humans value", as elaborated in Aven and Renn (2009). They discuss four approaches to QRA. Traditional statistical analysis, the probability of frequency approach by Kaplan and Garrick (1981) and Bayesian approaches estimating non-observable parameters, e.g. Szwed et al. (2006), focus on an underlying true risk.[3] In contrast, Bayesian approaches predicting observables (Aven, 2004) do not focus on an underlying true risk. To this purpose, they define four validity criteria:

V1. The degree to which the produced 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 (fictional quantities or observable events).

Table 2
Conceptual validity of QRA under different approaches; Aven and Heide (2009).

	Approach to QRA	Criterion			
		V1	V2	V3	V4
1	Traditional statistical analysis, large amount of relevant data available	Y	–	–	Y/N
2	Traditional statistical analysis in other cases	N	–	–	Y/N
3	Probability of frequency and Bayesian approaches estimating non-observable parameters	N	Y/N	Y/N	Y/N
4	Bayesian approaches predicting observables	–	Y/N	Y/N	Y

The results of the discussion by Aven and Heide (2009) is shown in Table 2, where it is concluded that under none of the approaches, QRA in general fulfils the scientific validity requirements.[4] Under certain conditions, they can be, which to some degree depends on the adopted risk analysis approach.

3.2. Validity in relation to different scientific foundations

In risk research, different views exists regarding the nature of the risk concept, the appropriate types of evidence to characterise risk, and more generally regarding the appropriate scientific foundations for risk analysis. This is known as the realist-constructivist schism, see e.g. Bradbury (1989), Shrader-Frechette (1991) and Thompson and Dean (1996). Broadly speaking, risk realists work under the presumption that risk is a physical property, which can be characterized by objective facts. They understand the quantities resulting from technical analyses as representations of this physical property, and see QRA as a tool for estimating the 'true' risk. In contrast, risk constructivists reject the idea of risk existing as a physical property independent of the people assessing it, but start from the premise that risk is a construct shared by a certain social group. QRA is understood here as a tool to formalize judgments about risk. Several authors have, sometimes implicitly, addressed the validation issue from these different viewpoints, as exemplified below.

Watson (1994) discusses the meaning of the probability in a QRA, distinguishing multiple interpretations. Two views are particularly relevant in the context of validation of QRA: the realist interpretation and the subjective interpretation. In the former, a probability is a measurement of a property of the way the world is, i.e. a physical property. He reports that nuclear regulators have a strong intuition of the desirability of a realist interpretation, as found also by Brown (1993). He also asserts that regulations requiring demonstration of a probability of a given event not exceeding a limit value, require a realist interpretation to be meaningful. In the latter, a probability is a measure of an individual person's degree of belief, in line with e.g. de Finetti (1974). Consequently, arguing for the subjectivist view of probability, Watson (1994) understands QRA as an expression of argument, rather than a tool to reveal a 'truth' about risk. More important than the probabilities per se, is the different types of evidence underlying the subjective measure of uncertainty, and the arguments given to assign a certain probability. For the problem of validation of QRA, this implies that the focus is on providing

[3] The concept of an 'underlying true risk' understands risk as a kind of physical property which can be accurately described and predicted, in line with realist approaches as described in Section 3.2.

[4] In Aven and Heide (2009), the four validity criteria are listed as degrees, whereas in their analysis and discussion, binary answers (yes/no) are applied. It is not entirely clear how exactly the validity criteria are measured, or what kind of threshold is applied to translate the degrees from the definitions into binary yes/no categories. In empirical work where the validity criteria have been applied (Goerlandt and Kujala, 2014), an ordinal categorization is applied derived from statistical measures.

substance to the strength of the argument, rather than a proof that the calculations accurately represent reality.

In a critique on Gori (1995), who proposes that risk regulation is restricted to risks which can be objectively defined and verified by scientific facts, Graham (1995) expresses his support to risk analysis practices based, if necessary, on the basis of intuition, experience, hunches, guesses and plausible assumptions, as long as the basis for decision making is clear. Consequently, he states: "any determination that a risk has been 'verified' is itself a judgment that is made on the basis of standards of proof that are to some extent arbitrary, disputable, and subjective". Also e.g. Apostolakis (2004) rebuts common criticisms on QRA, such as over-reliance on the accuracy on the produced numbers, from the perspective that a strict realist interpretation is untenable.

Rosqvist (2010) addresses validation of QRA in low-probability high-consequence systems. He asserts: "the credibility [...] of a risk model is determined by the decision context: if the stakeholders in a decision situation have no objections against the modelling assumptions, data, expert judgments and inferences made, then the [...] risk analysis can be considered good enough." The responsibilities and views on validation of typical stakeholders in a risk analysis project are shown in Table 3. Central to his understanding is the mental model of each stakeholder, i.e. the "representation of the surrounding world, the relationships between its various parts and person's intuitive perception about their own acts and their consequences". Taking a constructivist foundation, he rejects the idea of a true underlying risk. He relates validation of a QRA to a reinterpretation of the conceptual validity criteria V2, V3 and V4 of Aven and Heide (2009). If the expert is able to transform the perceived uncertainties related to a quantity into a probability measure (V2), if all unknown quantities taken into account in the risk model (V3), and if the right quantities are addressed in order to apply the adopted risk criteria (V4), then the QRA is valid as a tool to support decision making.

Goerlandt (2015) discusses the implications of adopting realist or constructivist scientific foundations for the validation of risk analysis. For risk realists, validation amounts to attempts to confirm that the estimates are close to the 'true' value. The focus then becomes a confrontation of the model output with empirical 'facts'. For risk constructivists, validation focuses on justifying the choices made in producing statements about risk, which is a semi-formal, conversational and argumentative process.

Finally, it is worth noting that the analysis of conceptual validity of QRA by Aven and Heide (2009) implicitly also makes use of the realist-constructivist distinction. In approaches 1–3 of Table 2, the focus is on an underlying true risk as in realist scientific foundations, whereas in approach 4 of Table 2, there is no focus on a 'true' risk as in constructivist scientific foundations.

Table 3
Stakeholders in a risk analysis project: responsibilities and views on validation; Rosqvist (2010).

Stakeholder	Responsibility	Validation issues of particular interest
Decision maker	Liability related to the risk decision	Are the inferences relevant for the used risk criteria related to the system under study?
Risk analyst	Risk model development and interpretation of the results under the assumptions and limitations coupled to the risk model	Are proper data, experts, methods and tools used such that uncertainties have been addressed and resolved as far as possible within the project?
Domain expert	Provides knowledge related to the phenomena under study, especially quantitative expert judgments on system model parameters	Is all background knowledge available in order for me to given an informed assessment?

3.3. Pragmatic validity: claims made about QRA

A central concern in regarding the pragmatic validity of QRA is intended use of the analysis. A regular criticism to QRA is an uncritical, mechanistic application of its results in a decision context and over-reliance on the produced numbers (Apostolakis, 2004; Crawford, 2001; Hagmann, 2012). However, several other uses have been proposed for QRA models and their results. As the requirements for pragmatic validity of a QRA depend on the intended use (Suokas and Rouhiainen, 1989), methods for establishing validity need to be clear which use(s) the analysis is being validated for. It is instrumental to define generic claims made about QRA as a basis of evaluating the evidence for or against QRA being fit for the stated purpose. Rae et al. (2014, 2012) differentiate three such claims.

The *accuracy claim* implies that the aggregate estimate of total system risk is sufficiently accurate and precise to allow correct decision making, which is needed e.g. for classifying risk, selecting between competing designs, or comparing risk with pre-defined risk targets (acceptability criteria) (Rae et al., 2014). In such uses, QRA models are artefacts producing accurate and encompassing answers to the classical questions: "What can happen? How likely is it? What are the consequences?" (Haimes, 2009; Kaplan, 1997).

The *cost-effective usefulness claim* means that performing QRA provides a safety benefit, which is measurably better than provided by methods not relying on quantification. Such usefulness is required e.g. for identifying ways to improve a design, trading-off risk against other concerns, or for tracking changes in risk over time (Rae et al., 2014). Here, QRA models are used to summarize evidence from different sources, facilitate communication between stakeholders, provide a platform for reflection and discussion, highlight areas of uncertainty where more information or research is necessary, or to complement operational experience. Such uses for QRA are suggested e.g. by Suokas (1988), Apostolakis (2004) and Marks (2007).

The *usefulness claim* is a fall-back claim, where QRA provides a benefit, but no more than other methods not relying on quantification. In other words, this claim means that the application of QRA helps designers to make safer systems and/or decision makers to make better decisions, but QRA cannot be shown to be more cost-effective than other methods (e.g. qualitative methods, safety engineering methods or behavioural safety approaches). As it is known that QRA studies are typically rather resource-intensive, the mere usefulness claim is a quite poor justification to the practice of using QRA: usefulness without cost-effectiveness may reduce the overall effectiveness of a safety programme, as scarce resources could potentially be allocated to more effective safety building strategies (Rae et al., 2014).

Suokas and Rouhiainen (1989) present three basic questions for the pragmatic validity of a QRA:

- How well has the analysis identified hazards and their contributors?
- How well are the results achieved relative to the resources used?
- How accurately has the analysis estimated the risks of the system?

These questions are well in line with the three claims about QRA suggested by Rae et al. (2014), matching respectively the *usefulness claim*, the *cost-effective usefulness claim* and the *accuracy claim*. This confirms that these claims are appropriate as a basis to scientifically establish how effective QRA is in relation to its possible uses.

3.4. Classification of approaches for establishing pragmatic validity

Suokas (1985) suggests following approaches for establishing pragmatic validity of a risk analysis:

1. **Complete Benchmark exercise:** comparison with a complete parallel analysis of the same system or activity
2. **Partial benchmark exercise:** comparison with a parallel analysis on some parts of the same system or activity
3. **Reality check:** comparison with operating experience of corresponding systems
4. **Independent peer review:** examination of the output of the risk analysis by a (range of) technical expert(s)
5. **Quality assurance:** examination of the process behind the analysis.

These methods can be used irrespective of the intended use of the QRA and the corresponding claims of accuracy, cost-effective usefulness or usefulness as suggested by Rae et al. (2014) and Suokas and Rouhiainen (1989), see Section 3.3.

The first two approaches, full and partial benchmark exercises, are primarily intended for evaluating the coverage of an analysis method, the reliability of the results in terms of analysis content and outcome, for highlighting uncertainties in the model and for identifying practices for dealing with the uncertainties. They are usually a part of the development and general acceptance process of a method, but are too expensive as an approach to assess the pragmatic validity of a particular QRA.

The third approach, a reality check, concerns the validity of a generic analysis method (e.g. how well it identifies hazards and their contributors), and can be applied to validate the results of a specific QRA. Comparisons with accidents and disturbances can provide the necessary insights. In several countries, accident databases have been established for this or other purposes, but the quality of accident descriptions has been (Suokas and Rouhiainen, 1989), and remains (Grabowski et al., 2009; Psaraftis, 2012) a problem. Reflections on occurred incidents in comparison with the risk analysis can provide insights whether they belong to accepted risks, whether they represent unwarranted assumptions made in the analysis, whether they belong to issues left outside the scope or whether they should have been covered but have remained unidentified.

The fourth approach, an independent peer review, builds on personal experience of individuals having technical expertise on the considered phenomena, practitioners, or experts in QRA. This can be applied to a specific QRA, facilitated by summary documents and presentations.

The fifth approach, quality assurance of QRA, is rooted in the similarity between a risk analysis and a production system, having its own information and management systems. It is similar to independent peer review, but focus is more on the adequacy of the processes behind the production of a QRA. A key assumption is that there is a strong relation between the quality of the underlying processes and the quality of the QRA. Hence, the pragmatic validity of a QRA is established by assessing the quality of planning, organization and execution of the analysis, as well as the quality of the analysis itself.

4. Methodological and empirical contributions

In this Section, contributions proposing methods to validate a QRA and empirical studies on the validity of QRA are reviewed. These are grouped according to the validation approaches as distinguished by Suokas (1985), see Section 3.4.

4.1. Benchmark exercises

Amendola et al. (1992) report on a benchmark exercise, where a reference ammonia storage facility was analysed by 11 teams representing control authorities, research organizations, engineering companies and industries. A complete risk analysis was performed and the results were compared with respect to methodologies, data and models employed. The results varied widely, both from one another, and when compared to available historic data. The numerical output showed a typical spread of 1–2 orders of magnitude, with deviations of up to 4 orders of magnitude not uncommon. Comparison of the results proved difficult, due in part to the lacking harmonization of content and presentation of results. Nonetheless, several sources of uncertainty and variability between the methods were identified, which were regarded as a starting point for establishing a consensus procedure for chemical QRA.

Lauridsen et al. (2002, 2001) report a QRA benchmark exercise concerning an ammonia storage facility, with seven different teams executing an independent analysis. The results were compared in terms of outcome, methodologies, data and models. Large discrepancies were found between the teams for the total risk, of two to three orders of magnitude. The ranking of the scenarios also differed substantially between analysis teams, especially for those scenarios characterized by less severe consequences. The main significance of the work was the identification of uncertainties concerning assumptions, methods and data in the different methods.

Ham et al. (2011) report on an extensive benchmark exercise on a virtual hydrogen refuelling station. Nine independent teams executed a QRA on a defined test case, each according to their own approach and practice, including identification of scenarios, probability analysis, consequence analysis and risk estimation. A set of requested output was defined, and some input values were suggested for the test case. As large differences were expected from earlier benchmark exercises, no attempt was made to be exhaustive in the considered scenarios or to obtain an 'accurate' total risk number. Rather, the aim was to identify differences in methodologies, assumptions made and knowledge gaps.

The results of this exercise indicate significant differences between the modelling approaches, the underlying assumptions and the outcomes. However, a comprehensive comparison between the approaches was not achieved in this study. According to the authors, this is due to not all partners being able to provide the results in the requested format, but also due to the multitude of different underlying models, assumptions and parameters. Comparison of data pairs could be useful to understand the reasons for the differences, but this has not been extensively pursued. In a second phase of the benchmark exercise, focus was on the consequences as calculated for selected scenarios, which allowed more meaningful insights in the reasons behind the differences.

Goerlandt and Kujala (2014) performed a comparative analysis of three models for assessing the risk of ship-ship collision, focusing on the probability of the accident. Two quantitative models and one qualitative method were applied to a case study of vessel traffic in a selected area of the Gulf of Finland. The study was framed in a context of three scientific reliability criteria proposed by Aven and Heide (2009):

R1. The degree to which the risk analysis methods produce the same results at reruns of these methods
R2. The degree to which the risk analysis produces identical results when conducted by different analysis teams, but using the same methods and data

R3. The degree to which the risk analysis produces identical results when conducted by different analysis teams with the same scope and objective, but no restrictions on methods and data.

Criterion R1 was met to a high degree for all considered methods. However, based on a systematic analysis of key parameters in the models, criterion R2 showed variations over an order of magnitude in the results, leading to a low to medium reliability. R3 reliability is low, with variations up to two orders of magnitude not uncommon. Significantly, the methods did not consistently rank the sea areas with highest risk, implying that different methods point to very different areas of highest accident risk.

4.2. Reality checks

Suokas (1988) performed a comparison of the results of four hazard identification methods, the backbone of QRA, with accident descriptions. The methods included hazard and operability study (HAZOP), action error analysis (AEA), failure mode and effect analysis (FMEA) and management oversight and risk tree (MORT). The aim was to evaluate which of the contributors would have been identified by each of the methods. Information on accidents was collected from seven process plants by interviewing process operators and their foremen. The evaluation of the methods revealed that together, the methods identified 55% of the contributors found in accident descriptions, with HAZOP most successful (36%) and MORT least successful (6%). Especially contributors from the technical subsystem were identified, with the methods less able to address the human and information and management subsystems. This showed the inherent limitations of QRA for large systems. It was concluded inter alia that QRA should complement and make best use of operational experience, that several methods should be applied in hazard identification to complement each other, that better tools were needed to address human and organizational factors. This is in line with recent research on knowledge management to identify atypical scenarios as a means to improve risk analysis practices by making better use of past experience and available information, see e.g. Paltrinieri et al. (2012).

Rouhiainen (1993) evaluated several risk analyses from the perspective of operational experience, in particular critical incidents, collected from the systems analysed. While not all analyses involved QRA, the completeness of the hazard identification techniques in comparison to the occurred incidents also is meaningful in a QRA context, to assess the completeness of scenarios considered. Two studies are particularly relevant. In a first study, three methods were applied to a pulp manufacturing process: the potential problem analysis (PPA), hazard and operability study (HAZOP) and action error analysis (AEA). A comparison with 397 incidents collected from production reports about the system revealed that the possibility of 73 of these were not identified by the analysis methods. In a second study, 51 incidents were collected from a liquid chlorine handling subsystem of a chlorine plant. Comparison with identified hazards based on PPA and HAZOP revealed that 12 were not identified based on the analysis. In both studies, several serious incidents were not identified. It was also found that for the considered studies, operational experience is of limited use for numerical comparisons with the analysis results.

Sornette et al. (2013) perform an evaluation of the risk as calculated in a QRA for a nuclear plant, compared with accidents and incidents in a (non-standard) database compiled by Sovacool (2008). They find that QRA is not adequate to reflect the true risks involved in the nuclear industry, in particular in regards the probability of large catastrophes.

4.3. Independent peer review

Garrick (1982) proposed a method for peer review in the nuclear industry. The review is facilitated by summary documents and presentations, with an emphasis on key results, major contributors to risk and the basic thought processes involved. The people conducting the review should be recognized technical experts, with experience in assessing complex system risk. In some cases, the high-level review may need to be supplemented with detailed technical reviews on issues raised in the high-level review. The levels of review are outlined in Table 4.

The United States National Research Council (NRC, 1998) described a set of criteria that are useful for peer review of any risk analysis study, see also Ahearne (1999). These are summarized in Table 5.

Rosqvist and Tuominen (2004) outline an exploratory peer review process for use in context of a Formal Safety Assessment (FSA), a quantitative risk-cost benefit procedure to support international maritime regulatory decision making (IMO, 2002). Several methodological validation criteria are proposed. Table 6 outlines those relevant for the risk analysis phase.

Psaraftis (2012) reports on the review criteria applied by the FSA Expert Group, based on MSC 86/26. These are:

1. consider whether the methodology was applied in accordance with the FSA guidelines and the guidance on the use of Human Element Analysing Process (HEAP) and FSA;
2. check the reasonableness of the assumptions and whether the scenarios adequately addressed the issues involved;
3. check the validity of the input data and its transparency (e.g. historical data, comprehensiveness, availability of data, etc.);
4. check whether risk control options and their interdependence were properly evaluated and supported by the assessment;
5. check whether uncertainty and sensitivity issues have been properly addressed;
6. check whether the scope of the assessment was met;
7. check whether expertise of participants was sufficient for the range of subjects under consideration.

Table 4
Levels of independent review of QRA, Garrick (1982).

QRA topic	Type of review	Comment
Objective and statement of purpose	High level	Must put the whole study and scope of review in perspective
Project plan and scope of work	High level	Must put the whole study and scope of review in perspective
Figures of merit	High level	Important distinctions between core melt, source term, and health effects must be clear
Methodology	High level, technical specialists	Attention should be given to areas such as treatment of uncertainty, plant recovery, dependent phenomena, and human error
Data base	High level, technical specialists	Special attention should be given to the different types of data involved and the process of going from generic to plant-specific use
Results	High level, technical specialists	Emphasis should be on major contributors, state of the art, and surprises and differences from other QRAs
Implementation and application	Technical specialists practitioners	This review should consist of detailed verification of key analyses leading to dominant contributors under the direction of the high-level review
Verification of selected results	Practitioners	This should be done under the direction and guidance of the high-level review

Table 5
Criteria for risk analysis validation, NRC (1998).

Criterion	Explanation
Constraints	Is there a clear statement of the constraints placed on the contractor and a clear statement of the impacts of these constraints?
Data collection	Are the data collection procedures clearly explained and are they based on established methods?
Key factors	Are all key factors included in the analyses? Is a credible explanation given for any that are not?
Assumptions	Are all assumptions identified and explained? Can the effects of these assumptions be traced through the analyses?
Methodologies	Are the analytic tools based on established procedures, or, if not, are they clearly explained and supported?
Transparency	Can the logic be followed readily? Is the influence of specific inputs and approaches, such as simulations, identified? Are the data and analyses handled competently?
Sensitivity	Were sensitivity analyses done regarding key assumptions and uncertainties?
Results	Do the results follow from the methods, the data presented, and the assumptions?
Conclusions and recommendations	Are the conclusions and recommendations consistent with the study results? Are the conclusions consistent with the results of the sensitivity analyses? Are the conclusions and recommendations adequately supported by evidence, analyses, and arguments?

Montewka et al. (2014) propose a Bayesian QRA modelling framework, including three computational methods to support model validation in an expert review. A one-way sensitivity analysis identifies the essential variables having the highest impact on the model outcome. A value-of-information analysis identifies the most informative variables with respect to the output variable, in terms of the concentration of the output's probability mass, detected using the entropy concept. An influence analysis quantifies the effect of changes in assumptions on the model outcome.

The above expert review methods focus on the resulting probabilities from the risk models, typically in a context of comparison with pre-defined risk acceptance criteria (Psaraftis, 2012; Rosqvist and Tuominen, 2004). Goerlandt (2015), extending work by Goerlandt and Montewka (2015), proposes a two-stage Bayesian QRA framework, where the risk model is used primarily as a tool for argumentation and as a basis for reflection, as suggested by NRC (1996) and taking a contemplative attitude to the evidence base as suggested by Marks (2007). Building on ideas about conceptual validity from social sciences (Trochim and Donnely, 2008) and risk research (Aven and Heide, 2009) and applying methods for model validation (Forrester and Senge, 1980; Pitchforth and Mengersen, 2013), a framework for assisting expert reviewers in the risk model validation is proposed, see Fig. 1. The focus of the framework is the practical use of the model, which is argumentatively approached through assessing the appropriateness of the model structure, content, discretization, parameterization and behaviour. For each of these model aspects, a series of questions is proposed, guiding the experts in their assessment of the model's usefulness. Value-related validity is approached by highlighting uncertainties and biases in the model construction, and procedural validity focuses on issues such as the appropriateness of expert elicitation procedures and transparency. The method is applied formatively, i.e. it does not lead to a gradation of the QRA with regards to its quality, but is only used to facilitate a discussion among reviewers, who make an informed decision about the QRA's fitness for use.

4.4. Quality assurance

Suokas and Rouhiainen (1989) introduced the concept of quality control in the validation of risk analysis. They also briefly outlined a checklist-based method for assessing the quality of a risk analysis, which progresses in accordance with the idealised phases of a risk analysis, see Fig. 2. This method is more elaborately described by Rouhiainen (1992, 1990) and Tuominen and Rouhiainen (1996). The assessment is based on the documentation of the analysis, which should therefore be detailed enough. As the expectations for the risk analysis depend from one activity to another, the method is intended to be flexible in order to deal with the range of requirements and expectations. The decision whether all aspects of the analysis have been taken into account adequately, and whether these have been executed appropriately, is left for the assessor to make. For instance, no specific methods which can or should be applied are suggested; it is assumed the assessor has the required knowledge of applicable methods and criteria for making a selection. Another example concerns the significance of the observed deficiencies. Similarly, the effect of missing or defective factors on the estimation of accident probability and consequence, should be evaluated. However, the method does not include any ranking on the importance of the aspects, and it is left to the assessor to make an informed judgment.

Pitblado (1994) describes the quality assurance approach adopted in a given consultancy organization for QRA of onshore and offshore facilities. Referring to generic national and international standards, including the ISO 9000, the approach taken to ensure quality is described. A central issue concerns feedback from clients concerning the services rendered, i.e. feedback of the QRAs as experienced by the users. One important issue here is that QRAs are often perceived to focus excessively on the technical issues of the analysis, and too little on identifying the user needs, framing and communication. A more balanced approach is required, as illustrated in Fig. 3.

A balance in the activities related to a QRA study are achieved through quality control processes similar as in ISO 9000. The technical quality system is built up as a hierarchy of documents, which is a basis for ensuring consistency between QRA analyses in different company business units and groups. The consultancy policy is a short document addressing issues as the company philosophy,

Table 6
Criteria for risk analysis validation, Rosqvist and Tuominen (2004).

FSA step	Criteria	Quality characteristics
Scope and objectives	Stakeholders are informed about adopted decision rules and criteria, verified by the peer group	Transparency
Hazard identification	Stakeholders' and domain experts' feedback on the completeness of hazard identification process is recorded and verified by the peer group	Completeness
Risk assessment	Model uncertainty and bias of risk model is addressed by domain experts and verified by the peer group Completeness uncertainty of risk model is addressed by domain experts and verified by the peer group	Credibility Completeness
Risk control options	Stakeholders' and domain experts' feedback on the completeness of risk control option are recorded and verified by the peer group	Completeness

Please cite this article in press as: Goerlandt, F., et al. Validity and validation of safety-related quantitative risk analysis: A review. Safety Sci. (2016), http://dx.doi.org/10.1016/j.ssci.2016.08.023

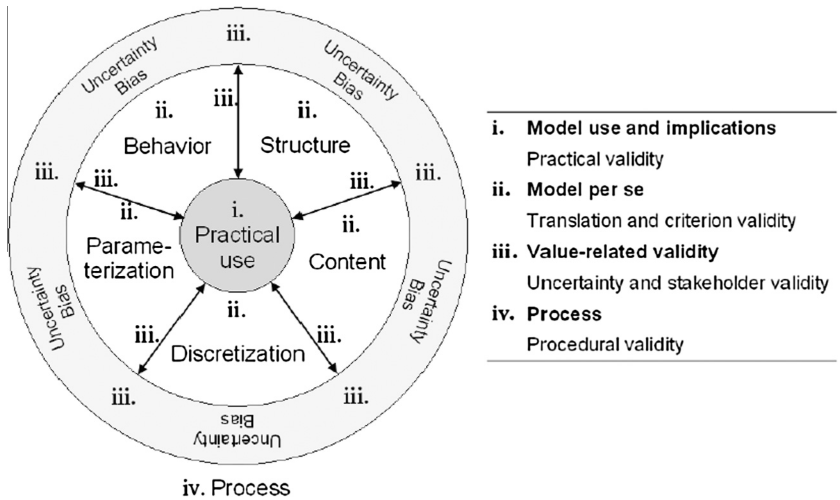

Fig. 1. Outline of QRA validation framework, Goerlandt (2015).

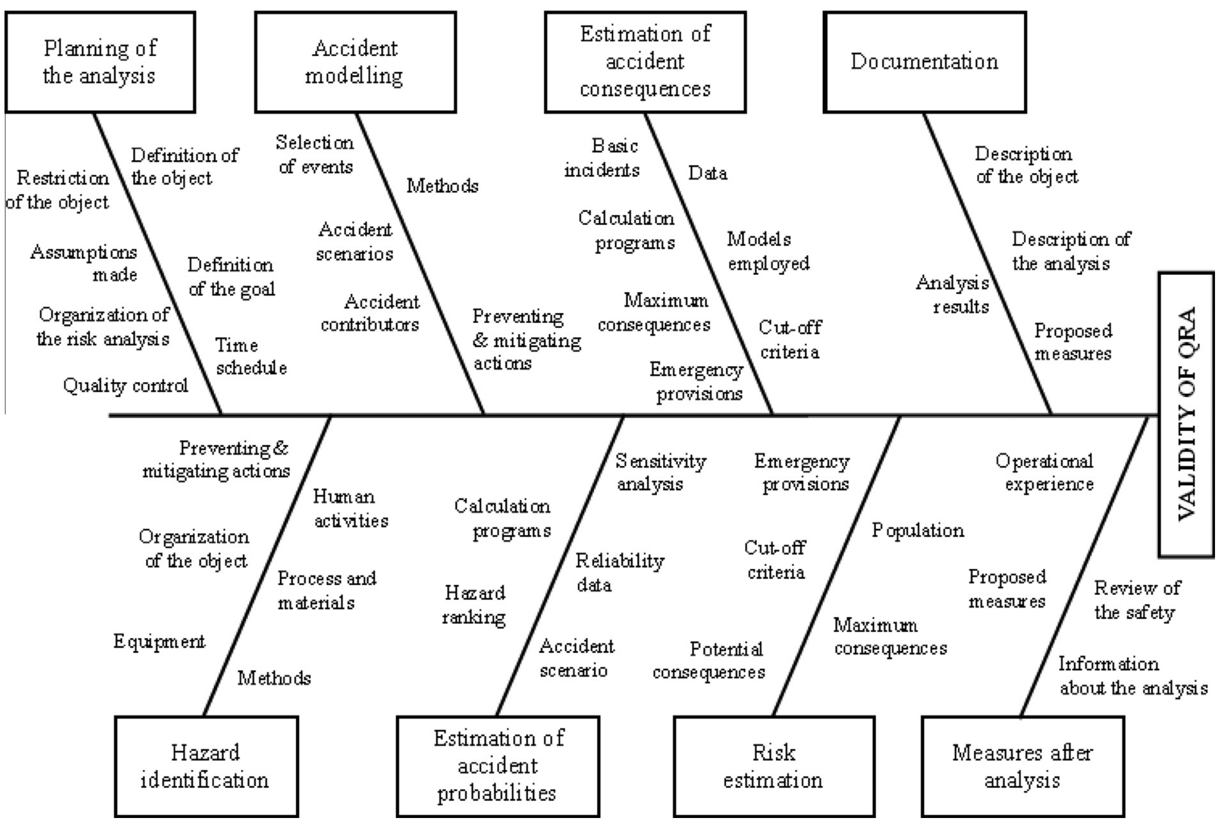

Fig. 2. Main aspects of a risk analysis covered by the quality assessment method, based on Tuominen and Rouhiainen (1996).

client needs, work methods and company culture. Guidelines address methods and strategies for carrying out different types of studies, whereas technical notes define detailed instructions for using individual QRA tools (models, data, and parameters). Finally, a procedure for deviations establishes how to proceed in cases a particular project wishes to deviate from the guidelines and technical notes, e.g. to account for new technologies or particular sources of uncertainty.

Vergison (1996) presented a quality assurance based technique for evaluating the validity of models for calculating the consequences of major hazards in QRA studies. A questionnaire was developed to cover a range of issues relevant to model validation, particularly aiming to bridge the gap between scientific advances and industrial application. The questionnaire is based on five concepts: scientific quality assurance, algorithmic quality assurance, computerization quality assurance, man-machine interface quality assurance and model validation and sensitivity analysis.

Rae et al. (2014) approach quality of QRA through a comprehensive classification of possible flaws in an analysis, from which a maturity model is derived to assess the quality of a QRA. The maturity model is based on a process model for a typical risk analysis, established based on a set of standards and manuals for risk analysis. These processes include: (i) provision of resources, (ii) establishment of context, (iii) identification of dangerous outcomes,

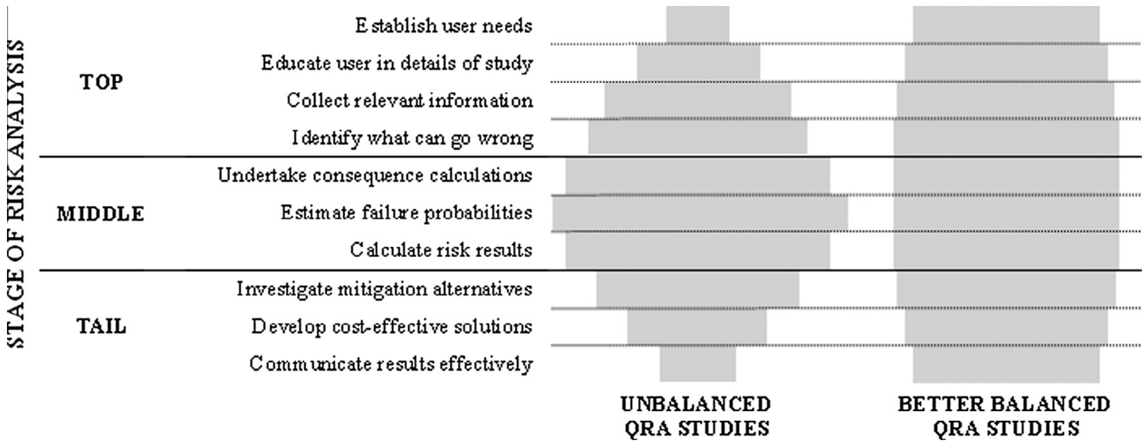

Fig. 3. Quality through a balanced approach to QRA, based on Pitblado (1994).

(iv) building models, (v) risk estimation, (vi) risk evaluation, (vii) planning actions, and (viii) communication of results. For each step of the process, an analysis was made of the possible flaws, based on concerns raised in the scientific literature, published peer reviews of risk analyses, and the developers' personal experience. These flaws were then assigned a level of the maturity model based on an assessment of their potential for undermining a risk analysis. Four practically achievable levels are defined, with flaw categories as defined in Table 7. The maturity model is a method for reviewing a QRA, where a higher level of maturity corresponds to a better risk analysis. At Level One, the QRA is not fit to use, mainly due to problems with documentation and presentation: as it is impossible to reconstruct how the analysis was performed, possible other flaws cannot be uncovered. At Level Two, the QRA is invalid in the sense that the flaws in the analysis greatly outweighs the underlying uncertainty being investigated. This concerns primarily substantive process failures. At Level Three, the QRA is valid but inaccurate: the analysis has several flaws which may distort the results, but likely do not dominate the underlying uncertainty being investigated. At Level Four, the QRA is accurate but challengeable: there can be disagreement about certain features of the analysis, because of limitations of the existing data and/or scientific knowledge.

The model can also act as a guide for research, e.g. by prioritizing empirical studies first on lower maturity levels, regarding which processes/techniques/social structures reduce errors, increase quality, or make practitioners more critically aware of the position of their work on the maturity scale.

Table 7
Maturity model for QRA quality assurance and review, Rae et al. (2014).

Maturity level	Flaw category
Level One: Unrepeatable	Failure to describe source material Failure to describe scope and objectives Failure to report methods Failure to report results
Level Two: Invalid	Major problems with source data Major omissions in the analysis which were not explicitly excluded by the scope Mismatch between the risk analysis and reality Major inaccuracies in the analysis Internal inconsistency Incorrect evaluation Misleading conclusions Failure to report limitations Systematic problems with the conduct of the analysis
Level Three: Valid but inaccurate	Insufficient rigour in selecting source data Incorrect processing of data Insufficient characterisation of uncertainty Shortfalls in conclusions and recommendations
Level Four: Accurate but challengeable	Disputed data sources Insufficient scientific knowledge

5. Other contributions related to QRA validity

A few other authors have addressed the validation of QRA, while not as such proposing a method for establishing pragmatic validity or performing empirical analyses according to one of the categories proposed by Suokas (1985), see Section 3.4. These are briefly outlined next.

Vasseur and Llory (1999) describe a synthesis of the results of an international survey to get advice from a number of international QRA experts and practitioners regarding, among other, the scientific and technical monitoring program in their organization, which includes the validation process to be implemented. As many issues are covered, from the viewpoint of individual experts, for a coherent summary of the findings, the reader is referred to the original publication. Overall, following issues are addressed: (i) QRAs and scientific validation, (ii) limit, restrictions and difficulties with QRA models, (iii) quality criteria for QRA models, (iv) indirect validation of QRA, (v) validation of the results obtained with QRA models, including assessment of assumptions, sensitivities and uncertainties, and (vi) the validation of the uses made of the QRA results.

Busby and Hughes (2006) discussed the validation of QRA from a normative perspective. Through interviews with experts from industry, regulators and academia, views were gathered on what the limits of QRA are in general, of the methodology, of the people using it, and of the context in which it is used. These limitations were subsequently developed into norms, i.e. requirements or conventions about how to best perform and apply QRA given its limitations. Some examples of such expert-defined norms are shown in Table 8. It was hypothesised that some of the norm categories could be further developed into a taxonomy, but this was not further developed.

Instead, a general discussion was given onto how build a framework in which to apply the norms. A central issue is the need for contextual sensitivity into the process of applying the norms: norms should be selected which are generally important, important in the considered context, likely to be infringed and representing issues where progress can be made. Furthermore, the norms

Table 8
Example categories of norms for validating QRA, abridged from Busby and Hughes (2006).

Norm category	Explanation	Example norm
Complete	Incorporating all important elements and concerns	Ensure violations are not omitted because they are controversial or unquantifiable
Knowledgeable	Drawing full on the knowledge available	Do not rely on risk assessors who work in supplier organizations that are insulated from failure knowledge
Grounded	Grounded in appropriate data and judgment	Avoid using statistical data that is not differentiated by conditions that are major causes of failure
Integrated	Pulling together the parts into an effective whole	Ensure risk assessment integrates knowledge that is fragmented across many parties
Systemic	Dealing with interactions and systems as a whole	Analyse systemic qualities like complexity and coupling rather than predict detailed event sequences
Guiding	Helping people using the assessment to use it effectively	Show how risk assessment processes raise awareness and test assumptions rather than providing a definite outcome
Open	Being open about problems and humble about achievements	Acknowledge the likely absence of data on causes that have only recently emerged as causes
Consultative	Involving stakeholders and taking account of concerns	Draw widely on working-level, supposedly non-authoritative sources of knowledge
Timely	Producing results soon enough to be acted upon	Conduct assessments early enough to influence design decisions that are costly to reverse
Accessible	Understandable to people with a stake	Provide easy ways to navigate from conclusions to analysis and assumptions
Facilitative	Providing a helpful basis for dealing with risks	Test for a culture of willingness in the organization in question to act on risk analyses

should be applied formatively, not summatively. That is, it should help people reason about a risk analysis, and perhaps improve it, rather than simply to declare it of a certain, quantitative standard. Possible uses for such a framework are demonstration, testing, dialogue and specification. Attention was also given to possible negative effects of a validation framework: if used inappropriately, it can be turned into a device to create a sense of credibility for the analysis, without strong foundations. However, if used as part of a conversational and transparent process, it can assist users of risk analyses to highlight the issues relevant to them, leading to a sense of trust in the analysis.

van Xanten et al. (2013) discusses the QRA practices in the chemical industry in the Netherlands, where a legally prescribed calculation method should be used to limit the variability of risk analyses. A case study is performed, to assess the appropriateness of this calculation model and to act as a basis for discussion regarding regulatory reform in the Dutch QRA practice. The evaluation focuses on two aspects of chemical QRA: (1) the calculation of individual and societal risk for land-use planning, and (2) the provision of insight into measures which may increase safety in the specific situation. No reference is made to a specific review method. Instead, five ad-hoc criteria are applied: transparency, verifiability, robustness, correctness and safety relevance. Based on the analysis, several recommendations are made to improve the Dutch use of QRA in decision making, including increased dialogue with stakeholders, increased attention to uncertainty, and improvements in presentation of the analysis results. These recommendations may also be of use e.g. for possible revisions of the Seveso Directive (Seveso III, 2012), as these currently include, as mentioned earlier, no guidance or requirements regarding quality assurance or validation.

6. Discussion

6.1. Evidence for the effectiveness of validation methods

An important aspect of the validation methods is their own validity. In other words: what evidence is available that these methods perform well for their purpose of validating a QRA? This is a significant question, as e.g. Busby and Hughes (2006) argue that there is a danger that with increased layers of formalisation we become less, rather than more, able to develop our judgment about what is good and bad in a risk analysis. Procedural approaches may be helpful, but inadequate or irrelevant procedures may exacerbate the problem rather than alleviate it. Given the relatively little attention to validity and validation, it is unsurprising that even less research on the validity of validation methods has been performed. No work was found addressing the effectiveness of benchmark exercises or reality checks.

Fabbri and Contini (2009) report on a benchmark exercise to evaluate how independent peer reviews of the same risk study could differ from one another. The focus is thus on the process of reviewing a risk analysis, questioning also how the differences may affect the finally adopted risk estimates. As a reference installation, as a tank wagon unloading station handling sulphur dioxide and a storage facility for ethanol was defined. A single tool for area risk analysis was applied, and a set of documents containing the relevant information for the plant, the surrounding area, and the results of the various steps of the risk analysis was produced. Nine independent teams performed an independent review of the QRA, using their own personal experience, judgment, data, and tools. No instructions concerning how to perform the peer review process were given. The results indicate that the participants held widely varying opinions about all phases of the risk analysis process. Such issues concerned the acceptability of the method, the quality of the reporting format, the clarity of assumptions, the adequacy of the conservativeness in the calculations, and the need for detailed analysis of catastrophic scenarios. Based on re-analyses and modifications found necessary by the different teams, deviations of up to 4 orders of magnitude in the different scenarios were obtained. Conclusions include the need for transparency and an open dialogue between the regulator and the operator about the QRA, the need for common formats for reporting QRA content and results, and the need for detailed guidance on how to evaluate a risk analysis from the regulator's standpoint.

This study indicates that, while independent peer review is seen as a necessary aspect of the application of QRA in decision making, see e.g. Apostolakis (2004), peer review suffers from similar problems as QRA itself, in particular a low reliability between different teams. Specific review criteria and guidelines, such as the ones proposed in Section 4.3, may alleviate some of the problems and reduce variability, but no research has been found to confirm this hypothesis. A central issue seems to be the way QRA is intended to be used: if a QRA is applied such that heavy reliance is placed on the results per se, as e.g. in the Dutch chemical industry (van Xanten et al., 2013), peer review may not be very effective. If however QRA is used as part of a dialogue between operators and regulators, as suggested e.g. by Marks (2007), Fabbri and Contini (2009) and Goerlandt and Montewka (2015), reviews may have more relevance as part of a process substantiating this dialogue.

The method for quality assurance proposed by Rouhiainen (1992) has been tested repeatedly, as reported in Rouhiainen (1993, 1992) and Tuominen and Rouhiainen (1996). Tests included the reliability of application of the method by different persons, and the validity in terms of the method's ability to detect deficiencies already identified earlier through comparison with incident and accident records. The results revealed that the assessors were able to identify the most significant deficiencies in the QRAs presented to them. While there was some variation in the deficiencies detected by test persons with different backgrounds and knowledge, the most significant ones were all identified. Typical deficiencies identified by the methods included: the lack of consideration of the importance of assumptions and of limitations of the hazard identification methods, the unsystematic and insufficiently detailed documentation, and the lack of examination of hazards caused by common cause failures and human error. This suggests that quality assurance can improve the quality of QRA. However, several limitations of the quality assurance method were identified as well: the method is not able to confirm that all relevant hazards and accident scenarios have been identified and appropriately addressed, it cannot reveal omissions and mistakes in calculations, and its application is subjective, both in terms of the aspects of concern and the importance of the identified deficiencies.

The maturity model by Rae et al. (2014) has not been extensively tested, but some initial confirmation has been achieved. Completeness, i.e. whether the model covers all types of errors made in risk analyses, has been tested based on the academic literature, published peer reviews of QRA and the authors' experience. While full completeness is not claimed, the evaluation indicates a good coverage of the model. Realism, i.e. whether the errors described in the maturity model are actually found in real QRA applications, is tested based on an application of the maturity model to a set of published peer reviews. Good agreement is found, indicating that the model is realistic. Finally, the appropriateness of the model, i.e. whether the lower levels of QRA maturity indeed correspond to more serious flaws, is argued for by the authors. While many QRA reviews detect flaws on various levels of the maturity model, the authors argue that such a categorization helps to focus efforts in the review and when making improvements in the analysis.

In sum, quality assurance has been found effective to reduce the number of deficiencies in QRA studies, while it has some inherent limitations with respect to how effective it can be. Importantly, the quality assurance rests on the hypothesis that a better process to produce a QRA leads to a better QRA. While based on the available evidence, this seems justified, it does not follow that this also necessarily improves decision making based on QRA, or that its application actually leads to safer designs. This leads to the question of the appropriateness of the different claims made about QRA:

6.2. Evidence for claims made about QRA

The empirical contributions of Section 4, while overall rather limited in number, are useful as a basis for discussing the claims made about QRA. The discussion made below builds on findings by Rae et al. (2014), complementing and extending these on some key points.

6.2.1. Accuracy claim

As noted by Rae et al. (2014), a key question in addressing the accuracy claim is how accurate the numbers need to be to support decision making. Referring also to Aven and Heide (2009), to several examples of case studies applying risk acceptance criteria, and to risk characterisations applying a mean probability value and confidence intervals, they argue that the required confidence interval is context-dependent. The way QRA is used defines this context: in case a QRA is used to justify that the system risk is lower than a predefined target level, the total error must be less than the margin between the estimated risk and the risk limit. A similar reasoning can be applied e.g. when the results of a QRA are expressed in an expected dose format, when maximum dose levels are predetermined.

Empirically determining the accuracy of a QRA is impossible: "one of the most powerful methods of science – experimental observations – is inapplicable to the estimation of overall risk" (Weinberg, 1981). A prerequisite for accuracy is reliability: the condition that upon repeated measurement, the results are 'similar'. Reliability is a matter of degree, as clear for instance from the reliability criteria proposed by Aven and Heide (2009), see also Section 4.1. Unreliable methods place an upper bound to the measurement accuracy. However, the reverse is not true: even perfectly reliable measurement may be invalid, if consistently the wrong results are obtained (Trochim and Donnely, 2008). Thus, benchmark and reliability studies can rebut, but not confirm the accuracy claim (Rae et al., 2014).

The benchmark exercises described in Section 4.1 all show considerable variation in the outcomes of the QRA studies: numerical differences of two to three orders of magnitude are not uncommon, and also rankings between different scenarios (a considerably lower requirement in measurement-theoretic terms), often showed large differences. The inherent variability found also in hazard identification techniques, an essential building block for QRA, adds to the unreliable nature of QRA. These benchmark exercises thus confirm the theoretical discussion on the unreliability of QRA, by Aven and Heide (2009), and provide strong evidence that the accuracy claim is false.

In practical terms, it is of course possible to limit the variability of outcomes by mandating the use of a particular QRA tool, as e.g. done in the Dutch chemical industry (van Xanten et al., 2013), but this does not mean the resulting numbers are more accurate in principle. It may lead to more comparable results between installations, but the uncertainty about the produced numbers, the dominant scenarios, and the relative importance of contributing factors remains. Based on the lack of proof for the accuracy claim, practices based on strict comparison of the QRA output with predefined acceptance criteria, such as described by van Xanten et al. (2013), seem untenable.

6.2.2. Cost-effective usefulness and usefulness claims

The cost-effective usefulness claim requires that using QRA has benefits in terms of system safety, or in terms of improved consensus building in decision making, which outweigh its costs (Rae et al., 2014). It is plausible that awareness of the ways in which a system can fail, knowledge of the dominant scenarios to defend the system against, the factors contributing to failures, the relative importance of different accident types, and more generally the safety information produced by performing a QRA, does provide a benefit to the management of risks. Quantification could assist this process, even if the numbers per se are not accurate. Alternative views of QRA, as processes to broadly describe uncertainties, including a wide assessment of the strength of evidence for producing risk statements and broad intersubjective assessments of 'black swans' (Aven, 2008; Paté-Cornell, 2012), may be plausible methods to support decision making and improving the system safety. Also e.g. other more modern approaches such as dynamic risk analysis, where the risk picture is dynamically updated in light of new information, can provide a path to more cost-effective uses of risk analysis as a tool supporting decision-making. We refer to Villa et al. (2016) for a recent review of such approaches.

Scientifically proving this however is not straightforward, especially when absolute measurements of safety are absent, and because controlled experiments on real-world applications are practically infeasible. The authors concur with the view of Rae et al. (2014) that "it is the responsibility of those making claims of usefulness to provide mechanisms for measuring those claims".

Given the wide variety of uses suggested for QRA, which do not rely on the accuracy claim per se, answering this question would require such mechanisms to confirm this for each stated purpose. To the best of the authors' knowledge, no such systematic mechanisms have been proposed to date, and neither have cost-effectiveness usefulness tests been made. The claim is thus neither proven nor refuted, while it likely is the highest achievable claim to be made about QRA, given the proof against the accuracy claim. For strengthening the foundations of risk analysis, as suggested by Aven and Zio (2014), developing, testing and applying such mechanisms should be considered important themes of fundamental risk research.

The authors at this point find it necessary to state our beliefs about QRA, as the above discussion may lead a reader to believe we oppose the technique. Again, we concur with Rae et al. (2014), in believing that QRA, adequately performed, probably does help people to design safer systems. We also believe QRA can help making good decisions. We however share the need for a sceptical attitude towards the validity of particular instances of QRA, and share the view that scientific confirmation of these claims would help to overcome doubts and criticisms about its efficacy.

Some credibility concerning the usefulness claim can be derived from occasional reports of successful application of QRA. For instance, Merrick et al. (2002) describe the Prince William Sound Risk Assessment, a quantitative risk analysis to assess the risk of oil spill in the waterway, in the wake of the Exxon Valdez grounding and oil spill, which had significant economic, ecological and socio-cultural effects (Herbst et al., 1996; Miraglia, 2002; Peterson, 2001). The analysis was performed as a collaborative research effort, where the analysis was the result of an analytic-deliberative process as described in NRC (1996). While the analysis did result in probabilistic statements about the relative importance of accident types in different locations, and included an analysis of the effect of risk interventions on the system safety, the main perceived value lays in the process of the analysis. Whereas the different stakeholder groups did not trust one another at the outset, the collaborative nature of the analysis fostered a cooperative risk management atmosphere among all stakeholders. Finally, all stakeholders accepted the results of the analysis as a basis for action, and a new ocean rescue tug vessel was deployed as a result. While this success story does not 'prove' that the system safety is indeed improved as resulting from the analysis, nor that the efforts were cost-effective, it does show that QRA can be useful, if only to facilitate consensus building for decision making. It is interesting to note also that the analysis results were accepted and acted upon, despite an independent review by NRC (1998) showing some deficiencies according to the criteria given in Section 4.3. The reasons for this are however unclear to the authors.

It is also worth noting that there are other types of criticism to QRA, which potentially put limitations on its usefulness to increase system safety effectively. This especially relates conflicting theories about the genesis of accidents in complex systems, with several accident theories challenging the appropriateness of linear accident causation models and related accident scenarios to understand failure processes and design countermeasures to improve safety. As QRA models typically apply linear or complex linear methods, such as fault trees, event trees, or Bayesian networks, the appropriateness of QRA for complex systems has been questioned by several authors (Hagmann, 2012; Hänninen, 2014; Leveson, 2012). Another challenge to applying QRA to complex system risk is the risk homeostasis theory, which suggests that additional safety measures can be attenuated by humans adapting their behaviour to keep a perceived level of risk constant (Wilde, 1998). Other theoretical difficulties with risk in complex systems are that risks have long incubation periods and migrate, i.e. addressing one risk may create another elsewhere in the system (Grabowski et al., 2000). It should however be noted that these theories are not conclusively proven themselves, adding to the complexity of the discussion (Pless, 2003; Rae, 2015). Therefore, establishing the validity of QRA also involves strengthening the evidence for or against the existing accident theories; another foundational issue in risk research (Aven and Zio, 2014).

7. Conclusion

In this paper, a review was made concerning the validity and validation of QRA, based on the scientific literature in key journals related to risk and safety. First, theoretical views on validity and validation were reviewed, showing the importance of a realist or constructivist foundation when performing and validating a risk analysis. The pragmatic validity of QRA was addressed in relation to different methods for establishing validity: benchmark exercises, reality checks, procedures for independent peer review and quality assurance. Each of these approaches were found to bring valuable insights into the validity of a particular QRA. However, more research is needed for instance with respect to the effectiveness of the proposed methods for the different identified uses of QRA, and which processes, techniques and social structures support their successful application. Finally, the review was used to discuss the claims made about QRA in relation to the available evidence. Rejecting the claim that accurate risk estimation is possible, the cost-effective usefulness claim seems plausible, but very little evidence for this claim has been found. This is therefore seen as another important area of future research.

References

Ahearne, J.F., 1999. The responsibilities of a probabilistic safety analyst. J. Risk Res. 2, 295–306.

Amendola, A., Contini, S., Ziomas, I., 1992. Uncertainties in chemical risk assessment: results of a European benchmark exercise. J. Hazard. Mater. 29, 347–363.

Apostolakis, G.E., 2004. How useful is quantitative risk assessment? Risk Anal. 24, 515–520.

Aven, T., 2008. Risk Analysis: Assessing Uncertainties beyond Expected Values and Probabilities. Wiley, Chichester, UK.

Aven, T., 2004. Risk analysis and science. Int. J. Reliab. Qual. Saf. Eng. 11, 1–15.

Aven, T., Heide, B., 2009. Reliability and validity of risk analysis. Reliab. Eng. Syst. Saf. 94, 1862–1868.

Aven, T., Renn, O., 2009. On risk defined as an event where the outcome is uncertain. J. Risk Res. 12, 1–11.

Aven, T., Zio, E., 2014. Foundational issues in risk assessment and risk management. Risk Anal. 34, 1164–1172.

Bedford, T., Cooke, R., 2001. Probabilistic Risk Analysis. Cambridge University Press, Cambridge.

Bradbury, J.A., 1989. The policy implications of differing concepts of risk. Sci. Technol. Hum. Values 14, 380–399.

Brown, R.V., 1993. Impersonal probability as an ideal assessment based on accessible evidence: a viable construct? J. Risk Uncertain. 7, 215–236.

Busby, J.S., Hughes, E.J., 2006. Credibility in risk assessment: a normative approach. Int. J. Risk Assess. Manag. 6, 508–527.

Cherdantseva, Y., Burnap, P., Blyth, A., Eden, P., Jones, K., Soulsby, H., Stoddart, K., 2016. A review of cyber security risk assessment methods for SCADA systems. Comput. Secur. 56, 1–27.

Crawford, J., 2001. What's wrong with the numbers? A questioning look at probabilistic risk assessment. J. Syst. Saf. 37, 1–11.

Cumming, R.B., 1981. Is risk assessment a science? Risk Anal. 1, 1–3.

De Finetti, B., 1974. The Theory of Probability. Wiley, London.

Fabbri, L., Contini, S., 2009. Benchmarking on the evaluation of major accident-related risk assessments. J. Hazard. Mater. 162, 1465–1476.

Forrester, J.W., Senge, P.M., 1980. Tests for building confidence in system dynamics models. TIMS Stud. Manage. Sci. 14, 209–228.

Garrick, B.J., 1982. On PRA Quality and Use (No. ALO–1017). California University, School of Engineering and Applied Science, Los Angeles, CA.

Garrick, J.B., Christie, R.F., 2002. Probabilistic risk assessment practices in the USA for nuclear power plants. Saf. Sci. 40, 177–201.

Goerlandt, F., 2015. Risk analysis in maritime transportation: principles, frameworks and evaluation (Doctoral Thesis). Aalto University, Helsinki, Finland.

Goerlandt, F., Kujala, P., 2014. On the reliability and validity of ship–ship collision risk analysis in light of different perspectives on risk. Saf. Sci. 62, 348–365.

Please cite this article in press as: Goerlandt, F., et al. Validity and validation of safety-related quantitative risk analysis: A review. Safety Sci. (2016), http://dx.doi.org/10.1016/j.ssci.2016.08.023

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.
Gori, G.B., 1995. Risk assessment reform: is it for real? Risk Anal. 15, 105–106.
Grabowski, M., Merrick, J.R.W., Harrald, J.R., Mazzuchi, T.A., van Dorp, J.R., 2000. Risk modeling in distributed, large-scale systems. IEEE Trans. Syst. Man Cybern. – Part Syst. Hum. 30, 651–660.
Grabowski, M., You, Z., Zhou, Z., Song, H., Steward, M., Steward, B., 2009. Human and organizational error data challenges in complex, large-scale systems. Saf. Sci. 47, 1185–1194.
Graham, J.D., 1995. Verifiability isn't everything. Risk Anal. 15, 109.
Hagmann, J., 2012. Fukushima: probing the analytical and epistemological limits of risk analysis. J. Risk Res. 15, 801–815.
Haimes, Y.Y., 2009. On the complex definition of risk: a systems-based approach. Risk Anal. 29, 1647–1654.
Ham, K., Marangon, A., Middha, P., Versloot, N., Rosmuller, N., Carcassi, M., Hansen, O.R., Schiavetti, M., Papanikolaou, E., Venetsanos, A., Angebø, A., Saw, J.L., Saffers, J.-B., Flores, A., Serbanescu, D., 2011. Benchmark exercise on risk assessment methods applied to a virtual hydrogen refuelling station. Int. J. Hydrog. Energy 36, 2666–2677.
Hänninen, M., 2014. Bayesian networks for maritime traffic accident prevention: benefits and challenges. Accid. Anal. Prev. 73, 305–312.
Hansson, S.O., Aven, T., 2014. Is risk analysis scientific? Risk Anal. 34, 1173–1183.
Herbst, A.F., Marshall, J.F., Wingender, J., 1996. An analysis of the stock market's response to Exxon Valdez disaster. Glob. Finance J. 7, 101–114.
IMO, 2002. Guidelines for Formal Safety Assessment (FSA) for Use in the IMO Rule-Making Process. International Maritime Organization, London, UK (MSC/Circ.1023-MEPC/Circ.392).
Kaplan, S., 1997. The words of risk analysis. Risk Anal. 17, 407–417.
Kaplan, S., Garrick, J.B., 1981. On the quantitative definition of risk. Risk Anal. 1, 11–27.
Khan, F., Rathnayaka, S., Ahmed, S., 2015. Methods and models in process safety and risk management: past, present and future. Process Saf. Environ. Protect. 98, 116–147.
Kirchsteiger, C., 1999. On the use of probabilistic and deterministic methods in risk analysis. J. Loss Prev. Process Ind. 12, 399–419.
Lauridsen, K., Christou, M., Amendola, A., Markert, F., Kozine, I., Fiori, M., 2001. Assessing the uncertainties in the process of risk analysis of chemical establishments: Part I & II. In: Zio, E., Demichela, M., Piccinini, N. (Eds.), Safety and Reliability towards a Safer World. Presented at the Proceedings of the European Conference on Safety and Reliability ESREL, Torino, Italy, pp. 592–606.
Lauridsen, K., Kozine, I., Markert, F., Amendola, A., Christou, M., Fiori, M., 2002. Assessment of uncertainties in risk analysis of chemical establishments: the ASSURANCE project Final summary report (No. Risoe-R, No. 1344(EN)). Forskningscenter Risoe, Denmark.
Leveson, N.G., 2012. Engineering a Safer World: Systems Thinking Applied to Safety, Engineering Systems. The MIT Press, Cambridge, Massachusetts.
Li, J., Hale, A., 2016. Output distributions and topic maps of safety related journals. Saf. Sci. 82, 236–244.
Li, J., Hale, A., 2015. Identification of, and knowledge communication among core safety science journals. Saf. Sci. 74, 70–78.
Li, S., Meng, Q., Qu, X., 2012. An overview of maritime waterway quantitative risk assessment models. Risk Anal. 32, 496–512.
Marhavilas, P.K., Koulouriotis, D., Gemeni, V., 2011. Risk analysis and assessment methodologies in the work sites: on a review, classification and comparative study of the scientific literature of the period 2000–2009. J. Loss Prev. Process Ind., 477–523
Marks, H.M., 2007. Contemplating risk assessment: a critique of NRC (1983, 1996). Hum. Ecol. Risk Assess. Int. J. 13, 7–19.
Merrick, J.R.W., Van Dorp, J.R., Mazzuchi, T.A., Harrald, J.R., Spahn, J.E., Grabowski, M., 2002. The Prince William Sound risk assessment. Interfaces 32, 25–40.
Meyer, T., Reniers, G., 2013. Engineering Risk Management. De Gruyter Graduate.
Miraglia, R.A., 2002. The cultural and behavioral impact of the Exxon Valdez oil spill on the native peoples of the Prince William Sound, Alaska. Spill Sci. Technol. Bull. 7, 75–87.
Montewka, J., Ehlers, S., Goerlandt, F., Hinz, T., Tabri, K., Kujala, P., 2014. A framework for risk assessment for maritime transportation systems—a case study for open sea collisions involving RoPax vessels. Reliab. Eng. Syst. Saf. 124, 142–157.
MSC 86/26, 2009. Report of the maritime safety committee on its eighty-sixth session.
NRC, 1998. Review of the Prince William Sound, Alaska. Risk Assessment Study. Committee on Risk Assessment and Management of Marine Systems, Marine Board, National Research Council.
NRC, 1996. Understanding Risk: Informing Risk in a Democratic Society. National Academy Press, Washington, D.C.
Paltrinieri, N., Dechy, N., Salzano, E., Wardman, M., Cozzani, V., 2012. Lessons learned from Toulouse and Buncefield disasters: from risk analysis failures to the identification of atypical scenarios through a better knowledge management. Risk Anal. 32, 1404–1419.
Pasman, H., Reniers, G., 2014. Past, present and future of Quantitative Risk Assessment (QRA) and the incentive it obtained from Land-Use Planning (LUP). J. Loss Prev. Process Ind. 28, 2–9.
Pasman, H.J., Jung, S., Prem, K., Rogers, W.J., Yang, X., 2009. Is risk analysis a useful tool for improving process safety? J. Loss Prev. Process Ind. 22, 769–777.
Paté-Cornell, E.M., 2012. On "Black Swans" and "Perfect Storms": risk analysis and management when statistics are not enough. Risk Anal. 32, 1823–1833.
Peterson, C.H., 2001. The "Exxon Valdez" oil spill in Alaska: acute, indirect and chronic effects on the ecosystem. Adv. Mar. Biol. 39, 1–103.
Pinto, A., Ribeiro, R.A., Nunes, I.L., 2013. Ensuring the quality of occupational safety risk assessment. Risk Anal. 33, 409–419.
Pitblado, R., 1994. Quality and offshore quantitative risk assessment. J. Loss Prev. Process Ind. 7, 360–369.
Pitchforth, J., Mengersen, K., 2013. A proposed validation framework for expert elicited Bayesian Networks. Expert Syst. Appl. 40, 162–167.
Pless, I.B., 2003. Is this journal really needed? Inj. Prev. 9, 289–291.
Psaraftis, H.N., 2012. Formal safety assessment: an updated review. J. Mar. Sci. Technol. 17, 390–402.
Rae, A., 2015. Open questions and closed minds: mapping the gaps and divisions in the safety body of knowledge. In: Presented at the Australian Safety Critical Systems Conference, Brisbane, Australia, pp. 1–6.
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.
Rae, A., Alexander, R., McDermid, J., 2012. The science and superstition of quantitative risk assessment. In: Presented at the Proceedings of PSAM 11 & ESREL 2012. International Association of Probabilistic Safety Assessment and Management, IAPSAM, Helsinki, Finland.
Reid, S.G., 2009. Confidence and risk. Struct. Saf. 31, 98–104.
Reniers, G., Anthone, Y., 2012. A ranking of safety journals using different measurement methods. Saf. Sci. 50, 1445–1451.
Rosqvist, T., 2010. On the validation of risk analysis – a commentary. Reliab. Eng. Syst. Saf. 95, 1261–1265.
Rosqvist, T., Tuominen, R., 2004. Qualification of formal safety assessment: an exploratory study. Saf. Sci. 42, 99–120.
Rouhiainen, V., 1993. Importance of the quality management of safety analysis. Reliab. Eng. Syst. Saf. 40, 5–16.
Rouhiainen, V., 1992. QUASA: a method for assessing the quality of safety analysis. Saf. Sci. 15, 155–172.
Rouhiainen, V., 1990. The quality assessment of safety analysis (Doctoral Thesis). Helsinki University of Technology, Espoo, Finland.
Saleh, J.H., Pendley, C.C., 2012. From learning from accidents to teaching about accident causation and prevention: multidisciplinary education and safety literacy for all engineering students. Reliab. Eng. Syst. Saf. 99, 105–113.
Seveso III, 2012. Council Directive 2012/18/EU (Seveso III) on the control of major-accident hazards involving dangerous substances. Off. J. Eur. Union L 197, 2012.
Shrader-Frechette, K.S., 1991. Risk and Rationality: Philosophical Foundations for Populist Reforms. University of California Press, Berkely.
SJR, 2016. Scimago Journal and Country Rank [WWW Document]. <www.scimagojr.com/> (accessed 4.28.16).
Sornette, D., Maillart, T., Kröger, W., 2013. Exploring the limits of safety analysis of complex technological systems. Int. J. Disaster Risk Reduct. 6, 59–66.
Sovacool, B., 2008. The costs of failure: a preliminary assessment of major energy accidents, 1907–2007. Energy Policy 36, 1802–1820.
Suokas, J., 1988. The limitations of safety and risk analysis. IChemE Symp. Ser. 110, 493–505.
Suokas, J., 1985. On the reliability and validity of safety analysis (Doctoral Thesis). Tampere University of Technology, Tampere, Finland.
Suokas, J., Rouhiainen, V., 1989. Quality control in safety and risk analysis. J. Loss Prev. Process Ind. 2, 67–77.
Szwed, P., van Dorp, J.R., Merrick, J.R.W., Mazzuchi, T.A., Singh, A., 2006. A Bayesian paired comparison approach for relative accident probability assessment with covariate information. Eur. J. Oper. Res. 169, 157–177.
Taroun, A., 2014. Towards a better modelling and assessment of construction risk: insights from a literature review. Int. J. Project Manage. 32, 101–115.
Thompson, P.B., Dean, W.R., 1996. Competing conceptions of risk. Risk Health Saf. Environ. 7, 361–384.
Trochim, W., Donnely, J.P., 2008. The Research Methods Knowledge Base, third ed. Atomic Dog Publishing.
Tuominen, R., Rouhiainen, V., 1996. Quality assurance of safety and risk analysis in space projects. Am. Astronaut. Soc. Sci. Technol. Ser. 87, 19–30.
van Xanten, N.H.W., Pietersen, C.M., Pasman, Hans.J., Vrijling, H.K., Kerstens, J.G.M., 2013. Rituals in risk evaluation for land-use planning. Chem. Eng. Trans. 31, 85–90.
Vasseur, D., Llory, M., 1999. International survey on PSA figures of merit. Reliab. Eng. Syst. Saf. 66, 261–274.
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, 281–297.
Villa, V., Paltrinieri, N., Khan, F., Cozzani, V., 2016. Towards dynamic risk analysis: a review of the risk assessment approach and its limitations in the chemical process industry. Saf. Sci. 89, 77–93.
Vinnem, J.E., 1998. Evaluation of methodology for QRA in offshore operations. Reliab. Eng. Syst. Saf. 61, 39–52.
Watson, S.R., 1994. On the meaning of probability in probabilistic safety analysis. Reliab. Eng. Syst. Saf. 45, 261–269.
Weinberg, A.M., 1981. Reflections on risk assessment. Risk Anal. 1, 5–7.
Wilde, G.J.S., 1998. Risk homeostasis theory: an overview. Inj. Prev. 4, 89–91.
Wybo, J.-L., Van Wassenhove, W., 2016. Preparing graduate students to be HSE professionals. Saf. Sci. 81, 25–34.
Zhuang, J., Bier, V., Guikema, S., 2016. Introductions to adversary behavior: validating the models. Risk Anal. 36, 650–652.