VTT PUBLICATIONS 507

On the use of expert judgement in the qualification of risk assessment

Tony Rosqvist

VTT Industrial Systems

Dissertation for the degree of Doctor of Technology to be presented with due permission for public examination and debate at Helsinki University of Technology, Espoo, Finland, in Auditorium E, Otakaari 1, Espoo, on the 11th of December, 2003, at 12 noon.

ISBN 951–38–6243–7 (soft back ed.)
ISSN 1235–0621 (soft back ed.)

ISBN 951–38–6244–5 (URL: http://www.vtt.fi/inf/pdf/)
ISSN 1455–0849 (URL: http://www.vtt.fi/inf/pdf/)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 5, PL 2000, 02044 VTT
puh. vaihde (09) 4561, faksi (09) 456 4374

VTT, Bergsmansvägen 5, PB 2000, 02044 VTT
tel. växel (09) 4561, fax (09) 456 4374

VTT Technical Research Centre of Finland, Vuorimiehentie 5, P.O.Box 2000, FIN–02044 VTT, Finland
phone internat. + 358 9 4561, fax + 358 9 456 4374

VTT Tuotteet ja tuotanto, Tekniikantie 12, PL 1301, 02044 VTT
puh. vaihde (09) 4561, faksi (09) 456 6752

VTT Industriella System, Teknikvägen 12, PB 1301, 02044 VTT
tel. växel (09) 4561, fax (09) 456 6752

VTT Industrial Systems, Tekniikantie 12, P.O.Box 1301, FIN–02044 VTT, Finland
phone internat. + 358 9 4561, fax + 358 9 456 6752

Technical editing Leena Ukskoski

Otamedia Oy, Espoo 2003

Rosqvist, Tony. On the use of expert judgement in the qualification of risk assessment. Espoo 2003. Technical Research Centre of Finland, VTT Publications 507. 48 p. + app. 82 p.

Keywords risk assessment, quality of risk assessment, qualification of risk assessment, expert judgement, expert model, probabilistic safety assessment, formal safety assessment, risk modelling

Abstract

The increased use of risk assesment in governmental and corporate decision-making has increased the role of expert judgement in providing information for safety related decision-making. Expert judgements are required in most steps of risk assessment: hazard identification, risk estimation, risk evaluation and analysis of options. The use and elicitation of expert judgement is therefore subject to on-going research. Furthermore, expert judgement is also required in the quality assurance or quality verification of risk assessment.

The research presented in the thesis addresses qualitative and probabilistic methods supporting the use of expert judgement in specific decision contexts; introduces a conceptual and procedural framework for quality verification of risk assessment; and presents techniques for the aggregation of probability distributions specified by experts' percentile information.

The methodological view to risk assessment adopted in the thesis is requisite modelling, where a decision and risk model is developed for a certain decision context to support decision-making under uncertainty, and refined until the decision-maker has confidence in the results and prescriptions obtained from the model.

The decision and risk modelling approaches presented in the thesis are related to specific decision contexts in maritime safety, maintenance management, and software reliability. The modelling approaches presented can, however, be utilised for risk-informed decision-making in other application areas as well.

Academic Dissertation

Helsinki University of Technology
Systems Analysis Laboratory

Supervising Professor:	Prof. Raimo P. Hämäläinen Helsinki University of Technology, Espoo
Supervisor:	Prof. Urho Pulkkinen VTT Industrial Systems, Espoo
Preliminary Examiners:	Prof. Louis Goossens Delft University of Technology
	Prof. Jørn Vatn Norwegian University of Science and Technology, Trondheim
Official Opponent:	Prof. Enrico Zio Politecnico di Milano

Preface

This thesis has been inspired by the work experience of several years (1994–) in the field of reliability and risk analysis as a researcher at VTT (Technical Research Centre of Finland). The work experience has ranged from statistical data analyses to process and quality issues related to risk assessment. The customers have been from various industry sectors such as mining, seafaring and energy as well as the Finnish Maritime Administration and the Radiation and Nuclear Safety Authority of Finland.

This thesis summarises and presents ideas and method developments related to problems that I have encountered in research and development projects. In particular, the use of expert judgement during the steps of risk assessment has provoked questions related to the confidence in the obtained results, and 'best practices' of utilising expert judgement in risk assessment. Hopefully, this thesis will provide new insight to these questions.

I want to thank several people, without whom this thesis would not have seen the daylight: Prof. Raimo P. Hämäläinen (Helsinki University of Technology, Systems Analysis Laboratory) for supervising the writing process; Prof. Urho Pulkkinen, Dr. Jan-Erik Holmberg, Mr. Risto Tuominen, Dr. Kari Laakso, Mr. Hannu Harju, Mr. Mika Koskela, Mr. Tapio Nyman and Mrs. Sanna Sonninen (all from VTT Industrial Systems) for giving their time to instruct, comment and co-write my papers; and Dr. Olli Ventä for giving me time and space for developing my ideas. I also want to thank my colleagues in the group Automation and Information Systems of VTT Industrial Systems for an inspiring atmosphere, and many others in the support personnel in VTT Industrial Systems and VTT Information Service, too many to mention.

Last, but not least, I want express my warmest thanks to my life companion Päivi for her belief in me.

Ingå, October 14th 2003 Tony Rosqvist

List of Publications

This thesis consists of the present summary and the following publications, referred to as Papers I–VI. The co-authors' contributions are indicated where relevant.

I Rosqvist, T., Tuominen, R. 2004. Qualification of Formal Safety Assessment – An Exploratory Study. *Safety Science*, 42(2), 99–120.

The paper presents a conceptual and procedural framework for the qualification of Formal Safety Assessment (FSA), i.e. how to establish confidence in the FSA results through quality verification. The framework can be viewed as a platform for risk communication to the decision-maker. The paper also presents a new way to structure model uncertainty, and discusses its role in the qualification of FSA. The risk assessment approach is called precautionary risk assessment. The co-author contributed to the text in all sections.

II Rosqvist, T., Nyman, T., Sonninen, S., Tuominen, R. 2002. The implementation of the VTMIS system for the gulf of Finland – a FSA study. *Proceedings of the RINA International Conference on Formal Safety Assessment*, 18–19 Sept 2002. London.

The paper presents a case study of the application of the Formal Safety Assessment method in assessing the risk impact of implementing a Vessel Traffic Management Information System for the Gulf of Finland. The risk impact was translated into monetary terms, and the risk assessment problem was reformulated as an investment problem. The co-authors were involved in the case as engineering experts.

III Rosqvist, T. 2002. Stopping time optimisation in condition monitoring. *Reliability Engineering and System Safety*, 76(3), 319–325.

The paper develops a decision model which utilises experts' percentile judgements on failure time as input after learning that an incipient fault has been detected, and computes the optimal stopping time for a risk averse decision-maker as output, using a genetic algorithm. The paper also develops a Bayesian expert model, where the log transformed expert judgements are modelled as multivariate normal random variables.

IV Rosqvist, T. 2000. Bayesian aggregation of experts' judgements on failure intensity. *Reliability Engineering and System Safety*, 70(3), 283–289.

The paper develops a hierarchic Bayesian expert model for specifying the failure intensity function of a repairable system. Expert judgement is elicited in the form of numbers of failures in the time windows chosen by the experts. The basic failure intensity functions studied are the two-parameter functions Power Law and Log-Linear. Specific measures of model fit are discussed.

V Rosqvist, T. 2004. Stakeholder compensation model based on decision analysis. *Journal of Multi-criteria Decision Analysis – Special Issue: Towards electronic democracy* (in press).

The paper develops a prescriptive decision model for a decision-maker faced with the problem of fairness in a multi-stakeholder decision-context, where the distribution of consequences and the risk attitude of the stakeholders differ. The paper is motivated by the need to develop fairness criteria for Formal Safety Assessment as recognised in Paper I.

VI Rosqvist, T., Koskela, M., Harju, H. 2003. Software quality evaluation based on expert judgement. *Software Quality Journal*, 11(1), 39–55.

The paper develops a new measurement approach for quality attributes of software where experts give their judgement in terms of a generic measure, the 'subjective achievement level'. The judgements are given in the form of triangular probability distribution functions expressing uncertainty with respect to the achievement levels. Acceptance rules for accepting the software are presented. The software measurement framework is developed to support group decision-making. The co-authors provided essential information on software quality related concepts.

Contents

1. Introduction

The only certainty in life is death; uncertainty lies in when and how death occurs, and whether it is final.

- W.D Rowe (1977)

1.1 Background, motivation and research approach

The role of risk assessment in structuring and quantifying information and uncertainty supporting decision-making in governmental agencies and companies has grown steadily during the last decades. In particular, decision-making related to the operation of hazardous socio-technological systems has adopted risk assessment as an approach to evaluate the acceptability of operating a system, and risk as a measure to compare the performance of systems.

Risk assessment has received a regulatory status in several industrial sectors which have given their own names for risk assessment, as practised in these sectors. For instance, in the nuclear energy and maritime transport areas, risk assessment is known as Probabilistic Safety Assessment (PSA) and Formal Safety Assessment (FSA), respectively. This does not mean that risk assessment is procedurally similar in the different application areas. In fact, risk assessment varies with respect to the use of experience data, expert judgement, risk modelling, decision rules and criteria. Furthermore, even within an industry sector, risk assessments differ, complicating risk comparison and risk communication of risk assessment results (Renn, 1992).

A challenge for risk analysts or risk assessment practitioners is to develop more harmonised practices for conducting risk assessment where issues related to establishing confidence in risk assessment results are systematically and generically addressed. This entails a conceptual framework for addressing determinants of confidence of the decision-makers with respect to risk assessment results. This has implications at the level of data acquisition, elicitation of expert judgement and risk modelling approaches applied during risk assessment.

In particular, since risk assessment is based on the principle of decomposition, many variables and model parameters relate to specific phenomena or events for which no empirical data is available. This motivates the use of expert judgement as a source of information in estimating the unknown variables and parameters.

One objective of the thesis is to develop a conceptual framework for quality verification of risk assessment. If quality verification is conducted as a peer review process, the process is called qualification in the thesis. The subject matter of qualification is the basic risk assessment tasks conducted by the risk assessment team: hazard identification, risk estimation, risk tolerability evaluation and analysis of options (IEC 60300-3-9, 1995). Qualification of risk assessment as applied in FSA, is addressed in Paper I. Need for qualification was identified in three FSA case studies (Rosqvist et al., 1998; Rosqvist et al., 2001; Nyman et al., 2002) in which the author participated. The concept ‘qualification’ was encountered in the project ‘PSA Qualification’, initiated by the Radiation and Nuclear Safety Authority of Finland (STUK) in 2001, with the aim at researching the conditions and procedures for quality assurance of Probabilistic Safety Assessments (PSAs) in specific decision contexts. This research is summarised in Simola (2002).

The other objective of the thesis is to summarise developments in modelling of expert judgement in risk assessment from the point of view of the decision contexts described and decision models presented in the Papers. In the Papers I–II expert judgement has been utilised to provide qualitative, as well as quantitative information. In Papers III, IV and VI, expert judgement is utilised more specifically as they relate to the quantification of decision or risk model parameters. In these papers, a common challenge has been to develop expert models that support the elicitation of quantities that experts are accustomed to observe and think about. Thus, the format of expert elicitation and the underlying expert model are crucial for the credibility of the expert’s judgements. In Papes III and IV the elicited expert judgements are observable variables, which are probabilistically related to model parametets. The elicitation is thus indirect. In Paper VI, the elicited expert judgements are related to the model parameters directly.

A technical note, Paper V, develops a stakeholder compensation model, supporting the decision-maker in ruling compensation between stakeholders in a decision context in a way that promotes fairness and also maximises social utility. The basic need for addressing fairness stems from the asymmetric distribution of risk and the costs of controlling these between stakeholders in relation to the operation of a hazardous system. This is noted in Paper I in relation to cost-benefit assessment in FSA.

The research approach in Papers II–VI is requisite modelling (Phillips, 1984), entailing conceptual development and technical solutions to support decision-making in specific decision contexts. The described decision and risk estimation models are generic at a conceptual level, and can thus be tailored for cognate decision-contexts.

The research approach in Paper I is exploratory. The conceptual framework related to the quality of risk assessment entails the definition of methodological quality characteristics and qualification criteria for quality assurance or verification. In the thesis, the conceptual framework is further developed and complemented with a process model of the qualification process, depicting the relationship between risk assessment and its quality verification, organised as a peer review.

All terms in the thesis follow the definitions in IEC 60300-3-9 Ed.1.0: Dependability Management Part 3: Application Guides Section 9: Risk analysis of technological systems (1995), unless otherwise noted.

1.2 Structure of the thesis

The papers of the thesis present formats for encoding expert judgement at various levels and stages of risk assessment and its qualification. The main stages of risk assessment are, according to IEC 60300-3-9, risk analysis, risk evaluation and risk reduction/control. Each stage comprises of specific tasks, as shown in Figure 1. The contents of the papers are mapped to the risk assessment activity diagram according to focus of content. In the thesis, qualification of risk assessment is viewed as separate from standard risk assessment and risk management, as depicted in Figure 1.

The rest of the thesis is organised as follows: section 2 summarises the background information, the motivation, and the contributions of Papers II–VI. Section 3 summarises the work in Paper I and provides further background information and developments: a process model of qualification of risk assessment is presented. The thesis ends with concluding remarks in section 4.

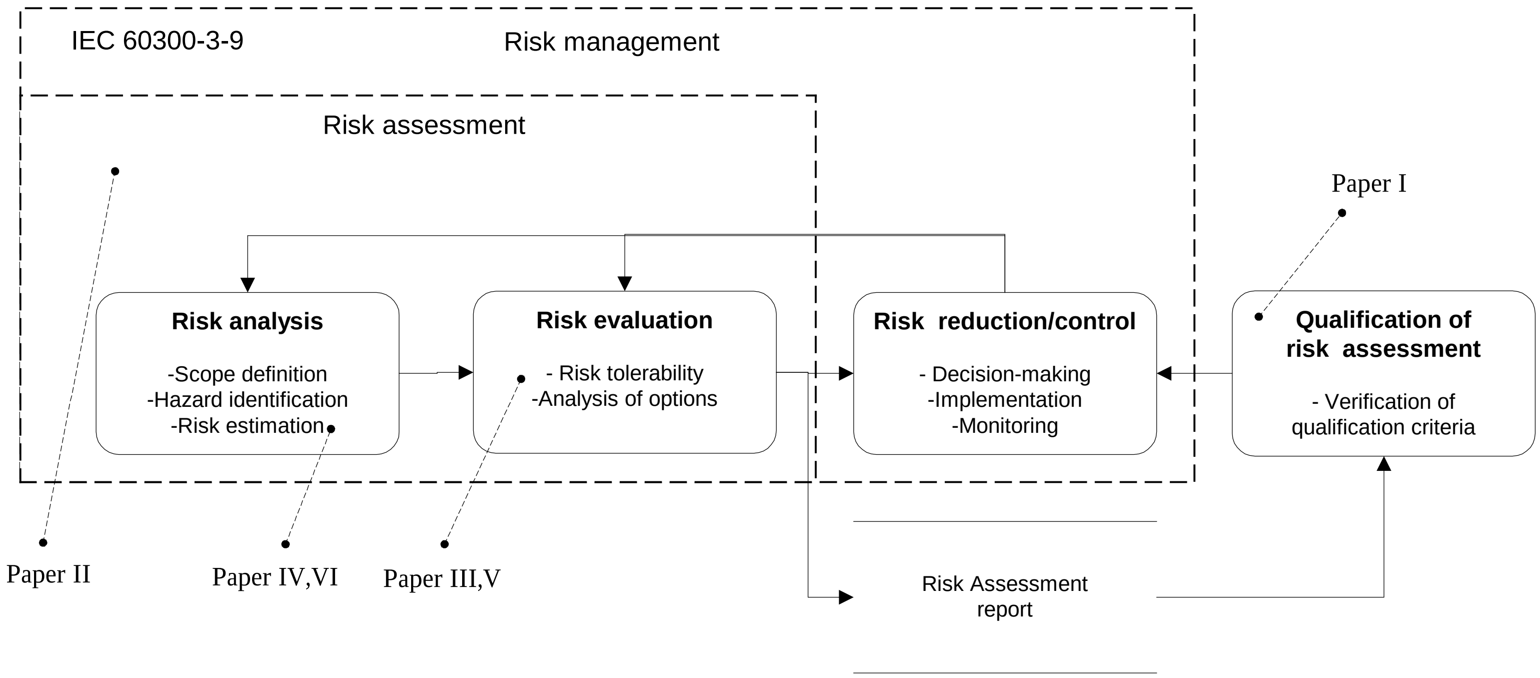

Figure 1. Positioning of the Papers I–VI on the IEC 60300-3-9 Risk Management Activity Diagram augmented with Qualification of risk assessment.

2. Expert judgement in risk assessment

2.1 Roles related to risk assessment

Roles representing different sources of information and accountability, relevant for the substance and the process of the risk assessment may be identified, as shown in Table 1. The descriptions are based on Pulkkinen and Simola (2000), and augmented by the author.

Table 1. ***.*** *Generic roles in risk assessment and risk assessment qualification.*

Role	Basic function
1. Decision-maker	• Presents the strategic view, presents the status of the decision-making process, and the objective of the outcome of the group decision • Responsible for the decisions based on the risk assessment • Identifies and selects the stakeholders • Defines the resources needed in the process • Provides the decision criteria
2. Referendary	• Selects the experts • Describes the case • Comments on the formats of the experts' judgements • Takes part in the discussions • Asks the opinion of the stakeholders on the quality of the results: has something been neglected? • Accepts the summary report • Explains the content and the conclusions of the risk assessment to the decision-maker
3. Normative expert	• Expert in expert judgement methods • Responsible for expert training, elicitation of judgements, and combination of judgements in the case of quantitative judgements • Responsible for the elicitation of the stakeholders' preferences and the development of utility functions • Responsible for reporting • Draws conclusions based on the decision criteria
4. Domain experts*	• Familiar with the issue • Responsible for the analysis of the issue and giving judgements (qualitative / quantitative) on it
5. Stakeholders	• Affected by the decision • Give feedback during the risk assessment process • Affirm the scope and completeness of issues

* refered to as 'expert' in the Thesis when there is no confusion with the normative expert

The assignment of roles in the steps of risk assessment and its qualification is an important task initiated by the decision-maker and complemented by the referendary, who is responsible for the overall risk assesment to the decision-maker. In particular, the contribution of the stakeholders is crucial for the qualification of the risk assessment as discussed in section 3. Specific information and expert judgements to be documented in the risk assessment report are defined in Table 2.

Table 2. Specific information and judgements required during the phases of a standard risk assessment (IEC 60300-3-9).

RA phase	**Decision-maker**	**Referendary**	**Normative Expert**	**Domain Expert**	**Stakeholder**
Scope definition	provides scope in general terms, decision criteria	-	-	-	-
Hazard identification	-	explains the motive, surveys stakeholder feedback	method of brainstrorming / expert panel	qualitative judgements and opinion on priorities	qualitative judgements and opinion on priorities affirms completeness
Risk estimation	-	-	expert judgement protocol	quantitative judgements on quantities	-
Risk tolerability	-	-	conclusions based on the decision criteria	-	-
Analysis of options	-	provides conditions for system (modification) options, surveys stakeholder feedback	method of brainstorming/ expert panel	qualitative judgements and opinion on priorities	qualitative judgements and opinion on priorities affirms completeness

2.2 Risk assessment and decision-making

Ideally, risk assessment should be directed by the decision rules and criteria adopted by the decision-maker (one person or an institutional body) in a specific decision context. Decision rules and criteria should reflect the objectives of the regulatory body or company as comprehensively as possible in order for the risk assessment to be supportive for the decision-maker in making risk-informed decisions, i.e. decisions under uncertainty. Decision criteria imply the scope of the decision context and the risk assessment approach from the point of view of the risk assessment team, and are pre-defined input 'data' for the risk assessment. The risk assessment team consists basically of the experts and the referendary.

The following examples illustrate the implication of the adopted decision rules and criteria (*DC*) on the risk assessment approach:

DC 1: 'Accept the system if the intolerability criterion $F(N) = 10^{-3}/N^2$/yr, where F is frequency and *N* the number of fatalities, is not exceeded'.

DC 2: 'Take the decision option that maximises my (decision-maker) expected utility'.

DC 3: 'Accept the system if the aggregated performance A satisfies $P(A \geq a_c; \mathbf{w}) \geq \alpha$, where **w** denotes the relative weights of the measures of performance, α a pre-defined risk level and a_c a performance treshold level'.

DC 4: 'Accept the system if the intolerability criterion (see *DC1*) is satisfied with precaution' (see Paper I).

DC 5: 'Accept the system if the average failure rate λ_{ave} of the population of units satisfies $P(\lambda_{ave} \leq \lambda_c) \geq \alpha$, where α denotes a pre-defined risk level, and λ_c a performance treshold level'.

The particular type of decision rule in *DC4* is relevant in Paper I, whereas the decision rule in *DC2* is relevant in Papers III and V. The case description in Paper II adopts a decision rule such as *DC2*, whereas Paper VI develops and adopts a decision rule similar to *DC3*. *DC5* illustrates the character of the ‘system’; a system can be viewed as a population of units, which are assumed statistically equal in the risk assessment.

Whether the decision rules are compensatory, as in *DC2* and *DC3*, or non-compensatory, as in *DC1* and *DC4*, different risk assessment approaches have to be applied. Typically regulatory decision rules are non-compensatory, or a mixture of non-compensatory and compensatory decision rules. In the case of a non-compensatory decision rule, acceptance with respect to *all* decision criteria has to be achieved separately in order for the system or decision option to be accepted. For instance, the ALARP-principle (Melchers, 2000) depicted by *DC1*, presupposes that intolerability levels related to each consequence (category) *N*, may not be exceeded (i.e. we have *N* criteria that have to be satisfied simultaneously). On the other hand, the ALARP-principle prescribes rules for trade-offs between risk reduction and cost in the so-called ALARP-region which are applicable when the unconditional criteria are satisfied.

A decision rule maximising expected utility (*DC2*) related to a single-attribute utility function is compensatory, as the distributional form of the consequence is not relevant in the decision rule. In other words, an intolerable (in the ALARP sense) performance with respect to some *N*’ may be levelled out by a risk margin with respect to some *N*’’ to yield acceptance based on compensation. More generally, in the case of a multi-attribute utility function, qualitatively different consequences are evaluated in a compensatory manner. The relative weight between the consequences (or attributes) are determined by standard preference elicitation techniques. (Keeney and Raiffa, 1993)

An important issue pertaining to a decision rule is how different types of uncertainties, as defined in USNRC (1998) and Zio and Apostolakis (1996), are taken into consideration, and how this will affect the risk assessment approach. In the case of precautionary risk assessment and decision-making, as reflected in *DC4*, model uncertainty is systematically assessed, and the direction of bias is determined in order to ensure that modelling assumptions do not underestimate the risk level used in evaluating acceptability / intolerability. This will be further discussed in section 2.4.

2.3 Expert and stakeholder judgement in hazard identification

The notes in this section are not specifically related to any of the Papers in the thesis. They rather reflect the experience of the author in conducting the hazard identification step of FSAs (Rosqvist et al., 1998; Rosqvist et al., 2001; Nyman et al., 2002).

The process of generating, developing and prioritising risk scenarios under time constraints, during the hazard identification step, is a challenging work group facilitation task. The results of this particular step are crucial with respect to the confidence on the results and recommendations of risk assessment, as the results represent the scope of the tasks in the later steps of the risk assessment.

The roles described in section 2.1 are particularly important in the hazard identification step of risk assessment. The interaction between the group and the individual can be better understood, if it is recognised that individual behavior is influenced by the role he or she has assumed, knowingly or unknowingly (Phillips and Phillips, 1993). Thus, by assigning roles in advance, the interaction is more controlled and helps the domain expert in facilitating the work group efficiently.

It is argued in the thesis that two normative experts are in practice needed in the hazard identification task: one directing the process, and one checking that all the information is recorded. The referendary is a natural choice for the second normative expert, if he is familiar with the expert elicitation methods utilised. This argument is based on several hazard identification tasks conducted by the author (Rosqvist et al., 1998; Rosqvist et al., 2001; Nyman et al., 2002), in which a Group Decision Support System (GDSS) was used. The advantages of using a GDSS relate to experts' simultaneous idea generation, commenting on others' ideas, development of ideas to more mature scenarios, and their prioritisation according to some voting scheme (e.g. risk matrix technique). At the same time, two problems known to affect negatively the work group, i.e. groupthink and dominance of strong personalities, are alleviated (Phillips and Phillips, 1993).

In the FSA case studies, and a case study related to investment risk assessment, the number of ideas generated in the hazard identification step were in the range of 50–100, and the number of scenario description in the range of 20–40. These were elaborated during 3–4 hours involving 8–15 people networked by 9–10 PC's (some of the participants shared one PC).

Although progress has been made in facilitating work groups by GDSS, it is, however, not clear what is meant by successful hazard identification in a work group setting. What criteria should be met for the hazard identification step to be 'complete'? In Paper I, it is proposed that a stakeholder survey be conducted to get feedback on the completeness of the results obtained. The feedback would indicate how stakeholders perceive 'completeness', but would not necessarily indicate if the set of stakeholders is complete.

In principle, the stakeholder selection problem is not a problem of the risk assessment team, but a problem of the decision-maker. From the point of view of the risk assessment team, the identified stakeholders are basically treated as 'input data' for the risk assessment. From the point of view of the decision-maker, the selection of the stakeholders has political ramifications that come in to play when risk assessment results are communicated to the public. The stakeholder selection process is thus an important issue closely related to risk assessment. A perspective to the stakeholder selection process is found in Harrison and Qureshi (2000).

The discussion above has implications with respect to the definition of qualification criteria related to the hazard identification step, discussed in Section 3.

2.4 Expert judgement on risk model structure

The main differences in risk assessment approaches relate to how uncertainties pertaining to risk model parameter values, model structure and completeness (USNRC, 1998) are addressed. The uncertainties are dependent on the complexity and understanding of the causal and/or logical relationships of quantities and/or events of the real world (Nielsen and Aven, 2003), and their significance in the decision-making process depends on the adopted decision rules and criteria.

A risk model is basically a model of uncertainties related to a real system. How these uncertainties are addressed qualitatively and quantitatively affects the completeness and the credibility of the risk assessment. The Bayesian statistical approach (Gelman et al., 1995; Lindley, 2000), the alternate hypothesis approach (Zio and Apostolakis, 1996), the adjustment factor approach (Zio and Apostolakis, 1996), the precautionary approach (Paper I), and the predictive Bayesian approach (Nielsen and Aven, 2003) address uncertainties in their own ways. This should be taken into consideration in the definition of the decision criteria adopted in the decision context.

Table 3 briefly describes the type of uncertainty, and how it is addressed in the respective risk assessment approach. Completeness uncertainty is particularly important, as it is crucial for deeming if risk assessment is powerful enough a methodology for scrutiny for the decision-problem in question. Completeness uncertainty pertains to the understanding of the physical and social phenomena under study. In the following, completeness uncertainty is referred to as conceptual uncertainty.

Table 3. Types of uncertainty addressed in risk assessment approaches.

Type of uncertainty	**Description**	**Example**	**Treatment in risk assessment approach**
Conceptual	The qualitative relationships of entities of physical or social phenomena are conceptually undefined / unknown. This reflects ignorance.certainty.	-the impact of organisational changes to work motivation -the effect of global warming on the paths of sea currents	Cannot be addressed by risk analysis methods, needs conceptual analysis and basic research first.
Model	The qualitative relationships of variables/quantities in physical or social phenomena are conceptually known, but practical limitations in describing complex relationships between the variables and/or acquiring detailed information (e.g. by physical measurements) for specifying the risk model induce model uncertainty.	-the loss of stability of bulk cargo under accelerating forces is a complex dynamic event driving many assumptions to be made in the construction of the risk model -the magnitude of the Loss of Coolant Accident event is categorised and realisations are modelled in a bounding way	*Multiple risk models:* -quantified in the AHA and the AFA *Single risk model:* -quantified in the AFP -qualitatively assessed in the PreA -not explicated in the PBA or BSA
Parameter	The true value of the risk model parameter is not known with	-the encounter rate λ of certain ships at a crossing area is uncertain due to scarcity of data. The parameter λ is typically the 'rate' parameter in a (spatial) Poisson process.	Depicted by probability distributions in all approaches: AHA,AFA, PreA, PBA, BSA

AHA = Alternate Hypothesis Approach, AFP = Adjustment Factor Approach, PreA = Precautionary Approach, PBA = Predictive Bayesian Approach, BSA = Bayesian statistical Approach

The different risk assessment approaches have different needs with respect to expert judgement. For instance, in the alternative hypothesis approach all models-of-the world are deliberated by expert collaboration, and a probability has to be assigned for each model denoting the belief of the decision-maker about the model's correctness. In the precautionary approach in Paper I, model uncertainty is qualitatively addressed to check that the decision-maker applies the decision rule in a precautionary way. In the Bayesian statistical model, all uncertainties are related to parameters, and evidence from quantities, especially in the form of expert judgement, is used to update the decision-maker's prior belief to posterior belief.

In Paper I, the decision rule for precautionary decision-making means that the acceptance of a hazardous system is based on such biased risk estimates that it is very unlikely to satisfy the acceptance limit falsely (hypothesising that we get to know the true hazardousness of the system). The opposite outcome, i.e. rejecting the system when in fact it would satisfy the acceptance limit, is possible. The precautionary risk assessment approach entails a systematic, qualitative assessment of bias coupled with modelling assumptions, as described in Paper I.

The development of the precautionary risk assessment approach can be viewed as a response to the need identified by Hattis and Anderson (1999), who say 'it is essential that tools be developed for responsibly describing the degree of bias [in risk estimates]'. Following the argument provided by Hattis and Anderson (1999), the utilisation of 'conservative', 'worst case' or 'biased' risk estimates in risk evaluation pertains to a) checking whether system options are acceptable notwithstanding the conservative risk estimate, and b) prioritising uncertainties subject to further investigation and refinement in the risk model.

It is important to note the role of completeness uncertainty in the quality verification of risk assessment. It is argued in the thesis that the most important objective of quality verification is to verify that the risk assessment team does not simplify risk assessment in the sense that 'what is counted for is what counts'.

2.5 Expert judgement in risk estimation

2.5.1 Behavioral and mechanistic probability aggregation

The use of expert judgement in the risk estimation step of risk assessment aims at producing a single representation, i.e. in practice an aggregated probability distribution of an unknown quantity. A formalised procedure for attaining this is described in Hora and Iman (1989), Keeney and von Winterfeldt (1991), Cooke (1991), Winkler et al. (1992) and Cooke and Goossens (2000). Such a procedure is known as an expert judgement protocol. The main challenge of the protocol is to control cognitive biases inherent in eliciting probabilities (Tversky and Kahneman, 1974).

We will not review each step in the expert judgement protocols described in the above references. We will note that preparations for the expert elicitation session and diagrammatic forms aiding the elicitation are found in e.g. Øien and Hokstad (1998). We will focus on the last step of the protocol, which is the aggregation of elicited probabilities. This particular step has received a lot of attention in the literature. For an annotation, refer to Cooke (1991), Clemen and Winkler (1999), and the references therein.

Expert judgement elicitation and aggregation approaches can be classified into behavioral probability aggregation and mechanical probability aggregation approaches (Clemen and Winkler, 1999). In the behavioral probability aggregation approach, the experts themselves produce the consensus probability distribution. The normative expert only facilitates the process of interaction and debate. Problems related to group dynamics are discussed in Phillips and Phillips (1993). The main objective of the approach is to ensure the achievement of a shared understanding of the physical and social phenomena and/or logical relationships represented by the parameter elicited. It is important to note that this approach induces strong dependence between the experts. This is in fact the aim of the approach, as viewed by the author: unanimousity with respect to the form of the common probability distribution (or the point estimate) is thought to be an indication that all background information is fully assimilated, and similarly interpreted by all experts.

In the mechanistic approach, experts' individual probability distributions are aggregated by the decision-maker after their elicitation. The main challenge is to specify the performance of the experts. Such a specification presupposes at least two assumptions as put forward in the thesis: 1) data for calibrating an expert's performance is available, 2) the expert has not learned from his past performance, and thus uses cognitive heuristics as before.

In the case of 'classical' (non-Bayesian) mechanistic probability aggregation, a proposed way of deriving weights indicating the performance of the experts, is based on applying the 'principles for rational consensus', as described in Cooke (1991), Cooke and Goossens (2000). The method requires the normative expert to define 'seed' variables for establishing the calibration and entropy scores needed to specify the weights of the experts. These seed variables should be selected in a way that trigger in the experts the same heuristics as the target variable.

Pulkkinen (1993) provides an information theoretic foundation for arriving at mechanistic probability aggregation rules: based on surprise measures, Pulkkinen shows that the total surprise experienced by the decision-maker, when experts' probability distributions replace the aggregated distribution, is minimised when the aggregation is based on arithmetic averaging. Similarly, the sum of the experts' surprises, when their probability distributions are replaced by the decision-maker's aggregated probability distribution, is minimised when the aggregation is based on geometric averaging.

Classical probability aggregation and behavioral probability aggregation are utilised in Papers I, II and VI according to a parameter specification procedure described in Paper I. The parameter specification procedure is described in section 2.5.2.

In the case of Bayesian mechanistic probability aggregation, the decision-maker defines the likelihoods of the experts' judgements and treats these judgements as data for updating his prior belief to posterior belief according to Bayes' rule. The data is typically given in the form of a finite number of percentiles or full probability distributions. In the case of full distributions, West (1988) has derived the form of the likelihood function.

The Bayesian mechanistic probability aggregation has been utilised in Papers III and IV. In Paper III, the percentile information of an expert is used to define the median and the variance parameter of the likelihood function, which is a multivariate normal data model of the log-transformed expert judgement on failure time. Paper IV presents an expert model where the parameters of the power law and the log-linear intensity functions are estimated in a way that uses a set of {number of failures, time frame of observation}-pairs elicited from the experts.

2.5.2 Parameter specification procedure

In this section, a parameter specification procedure, presented in Paper I, is discussed. The developed procedure can be viewed as an implementation of the concept ‘combined probability aggregation approach’, as introduced by Clemen and Winkler (1999). Based on the experience from the FSA case studies (Rosqvist et al., 1998; Rosqvist et al., 2001; Nyman et al., 2002), the following claims related to the format and procedures of eliciting expert judgements are put forward in the thesis:

- Empirical data needed for the rational consensus method by Cooke and Goossens (2000) is, in practice, difficult to develop for experts representing different expertise needed to specify the risk model parameters.

- For the same reason as above, it is difficult to specify the likelihood function of a Bayesian expert model on empirical grounds.

- Judgements pertaining to model parameters reflecting very specific phenomena or events may differ significantly due to the limited experience of the experts, making it difficult to achieve consensus based on behavioral aggregation.

- There is usually no prior information processed for the model parameters by the decision-maker or the referendary, as they seem not want to engage in such a task.

- Experts are willing to provide information in the form of two or three percentiles reflecting their uncertainty of an unknown, but observable quantity.

- Experts are willing to show and discuss their elicited percentile information.

The four first claims reflect limitations in the applicability of expert judgement models, wheras the two last reflect possibilities. These observations have partly led to the introduction of the information and parameter specification procedure in Figure 2, as reproduced from Paper I. The diagram shows the integration of the behaviorally and mehanistically based expert judgement protocols. It is left for the domain experts to decide which method to adopt for specific quantities.

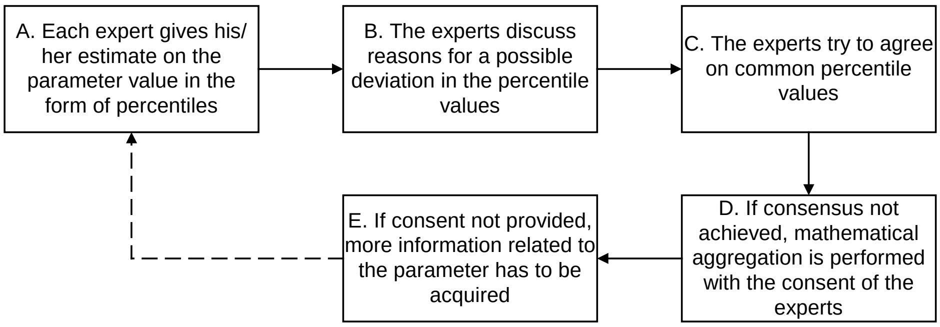

Figure 2. Parameter specification procedure (from Paper I).

In Paper II, the parameter specification procedure in Figure 2 was applied in specifying frequency and probability parameters of a fault tree model. The domain experts were asked to interprete probability in a frequentist way. More specifically, the domain experts where asked to view probability as a limiting relative frequency K/N related to a large population, where K is the number of events realised in a population of N ‘judgemental’ samples, i.e. the number of trials imagined. For instance, we may ask how many manouvers avoiding a sudden grounding hazard will fail (K events) in a sample population of 1000 (N trials) manouvers. The domain expert’s judgement is elicited as percentiles related to the ratio K/N depicting his uncertainty.

Thus, percentile values for both types of parameters in the fault tree model were elicited (Step A in Figure 2). The 5%- and 95%-percentiles were elicited. The median was nor elicited to avoid possible anchoring bias. This was done for all the basic events in the model. After that, the elicited percentile values were marked on a generic logarithmic scale. The normative expert asked each domain expert to describe the ‘sample population’ underlying his judgement, after which the domain experts were asked to reconsider their judgements. This usually resulted in an agreement of common percentiles representing a narrower probability distribution than one represented by the extremist percentiles in the group. If no reconciliation of views was achieved, mechanistic probability aggregation was offered as a solution. The parameter specification procedure in Figure 2 is also utilised as a basis for aggregating probability judgements in Paper VI.

The author argues that there is a seizable potential related to the use of GDSS in implementing expert judgement protocols. Currently, no GDSS’s exist supporting the elicitation of quantitative expert judgement and the aggregation of consensus probability distributions, as depicted by the parameter specification process model.

The last two claims in the bullet point list above are the motivations for the developments of the probability aggregation techniques in section 2.5.3 below.

2.5.3 Useful techniques for aggregation

This section introduces some useful techniques for aggregating expert judgements elicited during the expert judgement protocol. The unknown variable is assumed to be continuous in its domain. The judgements are given in the form of percentiles. We will look at ‘classical’ and Bayesian mechanistic probability aggregation.

<u>Classical aggregation:</u>

Elicit expert k’s q_r%-percentiles (r = 1,…, R): $\left(X_{q_1},...,X_{q_R}\right)_k = \mathbf{X}^k$ related to a quantity $x \in (Q_-, Q_+)$, where the domain is typically $(-\infty, \infty)$, $[0, \infty)$ or $[0, 1]$, and denote the minimal information probability density, specified by the percentile information, by $f_X^k\left(x;\mathbf{X}^k\right)$. The minimally informative probability density is piecewise constant with jumps at the corresponding percentiles[1].

The decision-maker's probability density is efined as

$$p_X(x)=\sum_{k=1}^{K} w_k f_X^k\left(x;\mathbf{X}^k\right). \qquad (1)$$

In practice, $w_k = 1/K \quad \forall k$, expressing that the experts are equally weighted. Thus we have

$$p_X(x)=\sum_{k=1}^{K}\frac{1}{K} f_X^k\left(x;\mathbf{X}^k\right) \qquad (2)$$

which can be identified as the marginal density related to the joint density of a uniformly distributed integer valued random index with the range {1,…,*K*}, and the expert's probability density indicated by the index.

Sampling scheme:

Samples from the decision-maker's probability density is obtained by the following procedure:

i) Set s=1 (sampling iteration)

ii) Take a sample from Uniform{1,…,*K*}: $k^{(s)}$

iii) Take a sample from the expert probability density indicated by $k^{(s)}$

iv) Set $s := s+1$

v) Continue with i-iv) until the histogram $\hat{p}_X(x)$ is stable (e.g. the sample mean does not change more than ε% in 100 consequtive draws).

---

[1] This follows intuitively from the notion of relative information between two probability distributions *Q* and *P* (Bedford and Cooke 2001). *Given* percentile information, the relative information of *Q* with respect to uniform *P* (background distribution) is minimised when *Q* is exactly *P* under the constraint.

Bayesian aggregation:

We utilise a result by West (1988) stating that the likelihood of observing an expert's full probability density $f_X(x)$ is

$$L(f_X(\cdot) \mid y) = e^{-\delta \int_{Q_-}^{Q_+} f_X(x) \ln\left(\frac{f_X(x)}{g_{X|Y}(x|y)}\right) dx} = e^{-\delta D(y)} \tag{3}$$

where $g_{X|Y}(x|y)$ is the decision-maker's belief related to the expert's prediction given the unknown value y. The integral $D(y)$ is seen to be the Kullback-Leibler divergence measure. Typically the domain (Q_- , Q_+) of the target quantity y is $(-\infty, \infty)$, $[0, \infty)$ or $[0, 1]$. The expert model $g_{X|Y}(x|y)$ may be defined as e.g. the normal data model: $g_{X|Y}(x \mid y) = N[h(y), \sigma^2]$ where $h(y) = y$, $h(y) = \ln y$ and $h(y) = \ln(y/(1-y))$, reflecting the decision-maker's belief that the expert is unbiased in the respective domains.

Any choices of $f(x)$ and $g_{X|Y}(x|y)$ have to be such that the integral exists over the whole domain of X for each Y. The parameter δ is a parameter associated with the weight of the expert's judgement.

In the case where the elicited information is a set of experts' percentiles, denote expert k's q_r%-percentiles (r=1,…,R) by $\left(X_{q_1}, ..., X_{q_R}\right)_k = \mathbf{X}^k$ and the corresponding minimally informative probability distribution by $f_X^k\left(x; \mathbf{X}^k\right)$. The likelihood of the K experts' probability distributions $f_X^k\left(x; \mathbf{X}^k\right)$ when the experts are conditionally independent, given y, is:

$$L\left(f_X^1(\cdot), ..., f_X^K(\cdot) \mid y\right) = \prod_{k=1}^{K} e^{-\delta_k D^k(y)} = e^{-\sum_{k=1}^{K} \delta_k D^k(y)} \tag{4}$$

After learning the experts' percentile information, the decision-maker updates his (noninformative) prior probability distribution of y, $p(y)$, according to the Bayes' rule:

$$p\left(y \mid f_X^1(\cdot),\ldots,f_X^K(\cdot)\right) \propto p(y)\cdot e^{-\sum_{k=1}^{K}\delta_k D^k(y)} \qquad (5)$$

The Kullback-Leibler divergence measures $D^k(y)=\int_{Q_-}^{Q_+} f_X^k(x)\ln\frac{f_X^k(x)}{g_{X|Y}^k(x \mid y)}dx$,

k = 1,…,K, have to be obtained by numerical integration in the general case. The parameters specified by the decision-maker's are δ_k and σ_k , k = 1,…,K. These parameters relate to the weight and precision of the unbiased experts, as viewed by the decision-maker. Without empirical data for expert calibration, we set $\delta_k = \delta$, $\sigma_k = \sigma$ $\forall k$, reflecting that the experts' judgements are equally valued. A tentative choice is $\delta = 1$ implying that the divergence measure is not weighted. The choice of σ is more difficult to make and should be subject to sensitivity analyses. In a hierarchical Bayesian model σ would be modelled as a random variable.

Sampling scheme:

In practice, set $p(y)$ = Uniform[Q_-, Q_+] with finite Q_- and Q_+. Samples from the posterior distribution is obtained by acceptance–rejection sampling (Smith and Gelfand, 1992) as follows:

i) Set s=1 (sampling iteration)

ii) Take a sample from $p(y)$ = Uniform[Q_-, Q_+] : $y^{(s)}$

iii) Compute the acceptance probability $\alpha^{(s)}=\frac{L\left(f_X^1(\cdot),\ldots,f_X^K(\cdot)\mid y^{(s)}\right)}{L^*\left(f_X^1(\cdot),\ldots,f_X^K(\cdot)\right)}$ for resampling, where the parameter denoting the expert model, $g_{X|Y}(x|y)$, is replaced by the arithmetic average $f_X^*(x)=\frac{1}{K}\sum_{k=1}^{K} f_X^k\left(x;\mathbf{X}^k\right)$ in $L^*(\cdot)$. Based on the results in Pulkkinen (1993) it can be seen that $L^*(\cdot)$ majorises the likelihood, i.e. $L(\cdot \mid y)\le L^*(\cdot) \quad \forall y$. In the case where

$g_{X|Y}(x|y)$ is normal it appears that the maximum likelihood can be analytically derived making it possible to optimise the resampling.

iv) Accept $y^{(s)}$ into posterior sample with probability $\alpha^{(s)}$

v) Continue with i-iv) until the histogram of the posterior $\hat{p}\left(y \mid f_X^1(\cdot),\ldots,f_X^K(\cdot)\right)$ is stable (e.g. the sample mean does not change more than ε% in 100 consequtive draws).

A related expert model, based on a hierarchic Bayesian network, is introduced in Pulkkinen and Holmberg (1997). This model avoids the specification problem related to the expert model's σ parameter above. Underlying each expert's percentile judgement is a 'hidden' sample population generated by the unknown normal data model parameters. The 'hidden' sample population is conditioned by the experts' percentiles. Gibbs sampler is used to generate the posterior distributions of the unknown parameters.

2.6 Evaluation of fairness among stakeholders

In FSA studies, an important issue of risk assessment is fairness among stakeholders faced with the consequences or risk decision-making. A decision analytic framework is proposed in Paper V, where a social utility function is defined based on weighted compensation-extended utility functions, in a way that the optimal compensation between stakeholders can be deduced if the decision-maker can assume some form for the utility function of each stakeholder.

Denote the utility functions of the decision-maker and the n stakeholders by $u(a_j), u_1(x_{j1}),\ldots,u_n(x_{jn})$, respectively, where $a_j \in A$ denotes the j^{th} decision option. The consequence x_{ji} of a decision option a_j for stakeholder i is an uncertain quantity and associated with the random variable X_{ji}. An additive social utility function is defined by

$$u\left(a_j\right)=\sum_{i=1}^{n}\lambda_i u_i(x_{ji}) \quad x_{ji}\in(-\infty,\infty)\ \forall ij \quad \lambda_i\in[0,\infty):\sum_{i=1}^{n}\lambda_i=1\ a_j\in \mathrm{A} \qquad (6)$$

where λ_i is a (normalised) scaling constant expressing the weight of stakeholder i. The set $A = \{a_1,\ldots, a_m\}$ denotes the decision options in the specified decision-context.

The social utility function $u(.)$ can be adjusted utilising compensation-extended utility functions associated with the stakeholders as

$$u(a_j;\boldsymbol{\delta}_j) \;=\; \sum_{i=1}^{n}\lambda_i u_i(x_{ji}\text{ - }\delta_{ji}) \qquad (7)$$

where $\boldsymbol{\delta}_j=\left(\delta_{j1},\ldots,\delta_{jn}\right)$ and $\delta_{ji}\in(-\infty,\infty)$ is the compensation for stakeholder i related to decision option a_j. The condition $\sum_{i=1}^{n}\delta_{ji}=0 \quad \forall j$ depicts that the amount received by some stakeholders is exactly the amount forfeited by the other stakeholders. In Paper V, it is noted that maximising the social utility function entails solving the Kuhn–Tucker conditions of the optimisisation problem:

$$\begin{aligned}&\boldsymbol{\delta}_j^{*}:\max_{\boldsymbol{\delta}_j}\left\{E_{X_{j1}\times\ldots\times X_{jn}}\left[u(a_j,\boldsymbol{\delta}_j)\right]\right\} \;=\; \max_{\boldsymbol{\delta}_j}\sum_{i=1}^{n}\lambda_i E_{X_{ji}}\left[u_i(x_{ji}\text{ - }\delta_{ji})\right]\\ &\text{s.t. } h(\boldsymbol{\delta}_j)=\sum_{i=1}^{n}\delta_{ji}=0\end{aligned} \qquad (8)$$

It has to be noted that the above problem formulation belongs to the area of convex optimisation. The special case of risk neutral stakeholders yields an undecisive optimisation problem, as any amount of compensation is valued equally among the stakeholders.

Paper V describes a worked example related to different decision contexts, where the probability distributions of the consequences vary among stakeholders with different risk attitudes.

The compensation derived pertains to the predicted distribution of consequences. The ‘a priori’ compensation can thus be viewed as a bargaining factor to engage and commit the stakeholders to a decision. Another problem is whether a redistribution of compensation should be ruled if, or when, the true consequences will be known, or a better prediction is available at some future time point.

It is important to note that no fairness criteria have been proposed in the general risk assessment framework of IEC 60300-3-9 or in the risk assessment application areas FSA or PSA. It is argued in the thesis that fairness criteria can be based on the ‘optimal compensation’ concept described in Paper V. This topic is relevant also from the point of view of quality verification of risk assessment, as discscussed in Paper I and in section 3 of the thesis.

3. Qualification of risk assessment

3.1 Background and motivation

Qualification – 'a condition or standard that must be complied with'
(Merriam–Webster's Collegiate Dictionary at www.britannica.com (2003))

The concept 'qualification' was encountered in a project initiated by Radiation and Nuclear Safety Authorities in Finland (STUK). The objective of the project was to research the conditions for a Probabilistic Safety Assessment (PSA) to be 'qualified', i.e. adequate for risk-informed regulation or decision-making (Chakraborty and Breutel, 2000). The project resulted in a working report, summarised in Simola (2002).

In IAEA (1999), a framework for a Quality Assurance (QA) programme for PSA, is delineated. Two aspects can be identified in the framework: a) Quality characteristics of the PSA study, and b) QA or quality verification tasks. The following quality characteristics are presented: Completeness, Consistency and Accuracy. It is noted in the report that to verify these characteristics, reviews at various levels and stages of risk assessment be conducted. A more detailed linking of the QA tasks and the quality characteristics is, however, not provided, and it is also not clear whether the suggested tasks are deemed sufficient. Furthermore, the QA tasks are integrated into the risk assessment tasks, making it difficult to identify what risk assessment tasks are, and what QA tasks are.

In Apostolakis et al. (1983) the main challenges for a PSA QA programme are identified: i) the definition of the quality characteristics of PSA quality, ii) the organisation of the peer review (nomination procedure, roles, levels of review), and iii) the accountability of the stated qualifications of a PSA.

In Macgill et al. (2000), a checklist-based quality audit framework is presented. The checklist is structured according to a hierarchy of characteristics and subcharacteristics. At the top level five basic quality characteristics are defined; the observational characteristic, the theory-informed methodological characteristic, the risk result characteristic, the validity characteristic, and the peer review characteristic. These are further divided into subcharacteristcs, each associated with an ordinal scale depicting the achievement level. The qulification thus entails detailed judgements on behalf of the peer group. How consensus is achieved with respect to the quality measurements is not discussed.

Another checklist-based approach to evaluate risk assessment quality is presented in Rouhiainen (1992). In this study, effort is laid on developing and validating the checklist approach. The work relates to the development of a tool rather than conceptualising a framework for verifying methodological quality characteristics.

Based on literature, it is argued in the thesis that QA of risk assessment is still in the stage of conceptual development. A mature conceptual framework relating to quality characteristics, peer review organisation and its tasks, and accountability, is a prerequisite for the qualification of risk assessment to serve its aim in risk-informed decision-making. In the next section, a complementary view to the quality audit framework described in Macgill et al. (2000), is put forward, based on Paper I. Further developments are also provided.

3.2 Developments of the qualification of risk assessment

According to IEC 60300-3-9, quality verification of risk assessment shall be assigned to an independent peer group. Independence means that the participants in the peer group have no vested interests in the risk assessment study. Support for this view is found in Macgill et al. (2000) and Apostolakis et al. (1983).

The general definition of quality verification or qualification of risk assessment is reproduced from Paper I with ‘FSA’ exchanged by ‘Risk Assessment’, to emphasise that the process of qualification is not specific to risk assessment application area:

Qualification of risk assessment is an independent peer review process consolidating the decision-maker’s confidence in the results and recommendations of risk assessment.

Thus, qualification of risk assessment is a process with the documented risk assesment as its study object. ‘Independent peer review’ entails that the evaluators involved have no vested interest in the risk assessment conducted, as noted above. ‘Decision-maker’s confidence’ entails that defined *qualification criteria* in the risk assessment are met, as verified by the reviewers. The qualification criteria relate to methodological qualities in each step of the risk assessment. It is argued in the thesis that the qualification criteria should be defined such that i) verifying them entails a binary judgement of ‘yes’/’no’, ii) they are mutually exclusive, iii) they are unambiguous supporting consensus judgements, and iv) their fulfillment is a necessary condition for qualification.

Figure 3 shows a set of methodological qualities that are deemed relevant in risk-informed decision-making. Loosely speaking, the methodological quality characteristics represent qualities that should be met to assure that “right things are done right”. Qualification criteria, implying the quality characteristics in Figure 3, are presented in Paper I, for each step of a FSA. Similarly, qualification criteria related to the risk assessment steps in IEC 60300-3-9 are presented in Table 4.

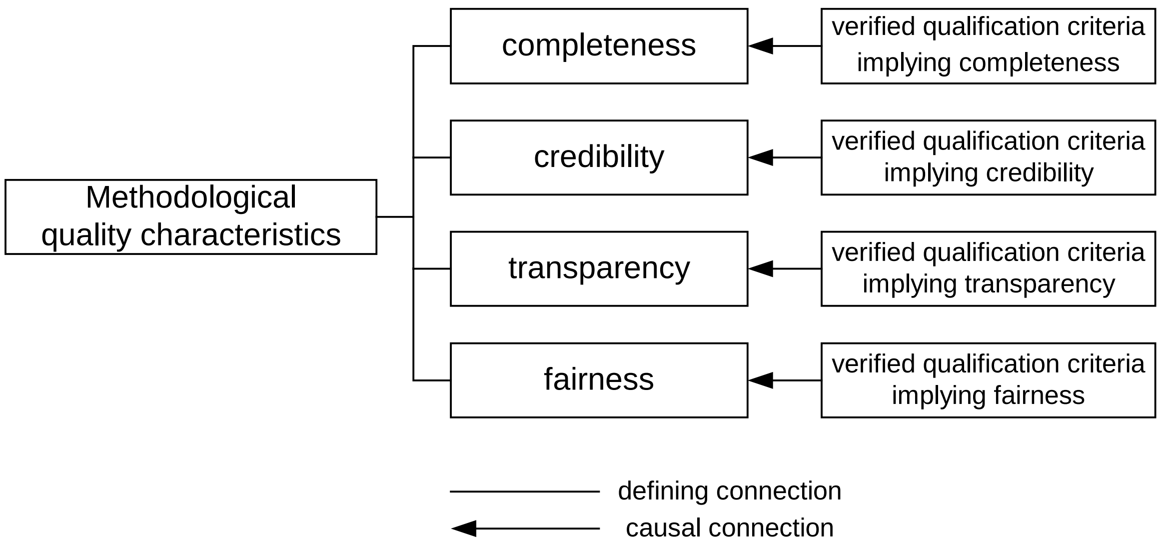

Figure 3. Relationship between methodological quality characteristics and qualification criteria in the qualification process.

Table 4. Methodological quality characteristics and qualification criteria for precautionary risk assessment and decision-making.

Risk Assessment step **(IEC 60300-3-9)**	**Qualification criteria**	**Quality characteristics**
1. Scope Definition	The peer group verifies that the stakeholders have been informed about adopted decision rules and criteria	Transparency
2. Hazard Identification	The peer group verifies that the stakeholders' and the domain experts' feedback on the completeness of hazard identification process is adequately surveyed	Completeness
3. Risk Estimation	The peer group verifies that sensitivity studies, based on parameter uncertainty, are adequate	Credibility
	The peer group verifies that model uncertainty and direction of bias of risk model is adequately addressed	Credibility
4. Risk Tolerability	The peer group verifies that the conclusions drawn as based on the decision rules are consistent	Credibility
5. Analysis of options	The peer group verifies that the stakeholders' and the domain experts' feedback on the completeness of risk control options is adequately surveyed	Completeness
6. Recommendations for the decision maker	A peer review has been conducted	Qualified RA

It is important to note that the qualification criteria defined in Table 4 are related to concrete risk assessment *tasks* in order to satisfy the points i)-iii) in the previous paragraph. Point 6 in Table 4 is different from the other points and not part of risk assessment, but rather a part of risk management, as described in IEC 60300-3-9. The verification of the fulfillment of a qualification criterion is undoubtedly judgemental and dependent on the verifyer's experience on risk assessments. If consensus cannot be achieved in terms of the ‘yes’/’no’ – answers, it may imply that the risk assessment document is inadequate and more details should be provided about the risk assessment.

In general, the adopted decision rule and decision criteria are the basic determinants for defining qualification criteria as depicted in Fig. 4, which shows the relationships between the risk assessment tasks (as documented), the quality verification or qualification, and the decision-making (the graphical notation used is IDEF0 (Feldman, 1998)). The definition of the qualification criteria is thus linked to the risk assessment approach adopted through the choice of the decision rule and criteria. For instance, if the decision rule reflects precaution, as discussed in section 2.2, qualification criteria related to model uncertainty should be defined. Figure 4 is based on the qualification process introduced in Simola (2002).

The inputs to the qualification process are the documented results related to the basic risk assessment steps (IEC 60300-3-9). The outputs are the qualification results and recommendations as judged by the peer review group. The decision-maker makes his decision based on the risk assessment results *and* the quality verification results. The basic decision options for the decision-maker are a) to accept the risk assessment results, or b) to request refinements in the risk assessment, based on the recommendations of the peer review group.

An important issue related to the interaction between the risk assessment and the qualification process is project funding. In Figure 4 a dashed line connects the point of commission of the risk assessment, including the allocated project resources as mandated by the decision-maker, to the qualification process. The status of independence of the peer review process is severely hampered if the project resources affect or steer the qualification process. It is argued in the thesis that this interaction should be taken care of by separately organising funding of the processes.

It is interesting to note that in the quality audit framework of Macgill et al., many of the same methodological quality characteristics as in Paper I are identified. In Paper I and the thesis, the qualification pertains to specific tasks viewed as qualification criteria for the risk assessment approach, whereas in Macgill et al. quality verification relates to more general notions, independent of the adopted decision rule and the risk assessment approach.

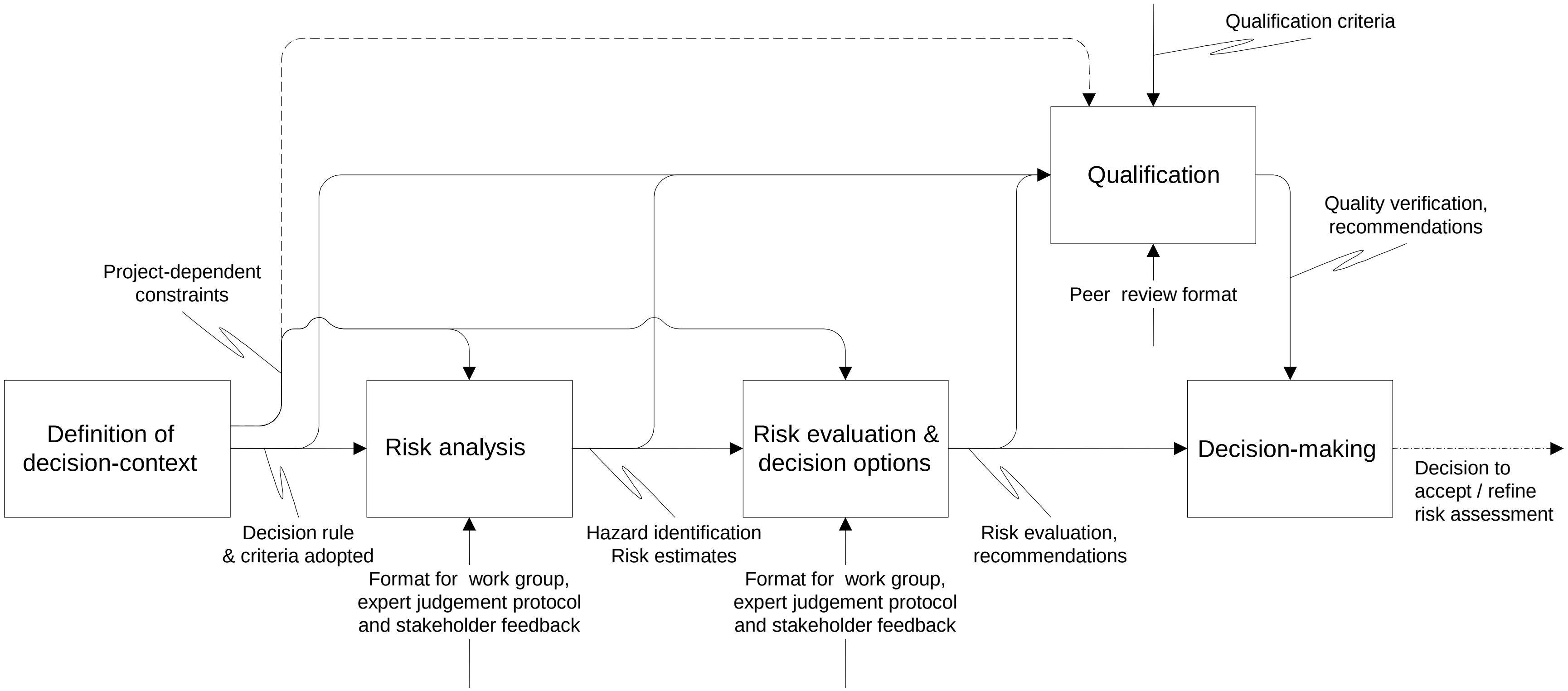

Figure 4. Qualification of risk assessment.

4. Concluding remarks

4.1 Scientific achievements in the Thesis

Table 5 summarises the main scientific achievements in the Thesis i.e. the Summary and the Papers I–VI in the appendices. The research approach in Paper I is exploratory; further applications of the precautionary risk assessment approach and the qualification process are needed to evaluate the contribution of the paper in the risk assessment field. Especially the role and importance of qualification of risk assessment should be subject to further studies. Paper II is a real application where the precautionary risk assessment has partly been adopted. The research approach in Papers II–VI is constructive research with emphasis on conceptual and decision model development. Solid evaluation of the developed methods needs more real case applications.

Table 5. The main scientific achievements in the Thesis.

Summary	The Summary sets the developments in the Papers I–VI in a broader perspective by referencing to selected literature in the area of expert judgement, and furthers the developments in Paper I by introducing a process model of the qualification of risk assessment. Also novel technical solutions on the use of elicited percentile judgments of experts in order to derive aggregated probability distributions are introduced in the Summary.
Paper I	The concept of precautionary risk assessment is introduced. The roles of parameter, model and completeness uncertainties are discussed from the point of view of quality verification of risk assessment. The concept qualification of risk assessment is also introduced. It entails a conceptual framework for communicating quality-related properties of risk assessment.
Paper II	Formal Safety Assessment was applied to assess and evaluate the safety impact of implementing pre-defined VTMIS systems in the Gulf of Finland. The research results convinced the International Maritime Organisation to give support for the implementation. The methodological framework is based on existing methods. The Fault Tree -models, estimating the collision avoidance probability given an encounter situation, were quantified according to the parameter specification procedure developed in Paper I. The quantification task confirmed the hypothesised usefulness of the parameter specification procedure.

Paper III	A stopping time optimisation model is developed to aid the decision-maker to make a decision whether to stop or continue the operation of a system given that an incipient fault has been detected. The developed stopping time optimisation model integrates experts' judgements on failure time to a stochastic cost function such that the stopping time maximising expected utility can be derived. Such decision models have not been developed to date.
Paper IV	An expert model is developed for specifying the failure intensity function of a non-homogeneous Poisson process. Failure intensity functions are needed, for instance, in simulating availability of new system design. Usually, however, failure data is not availbale for the statistical estimation of the parameters. The expert model introduced makes it possible to use expert judgement in the form of {number of failures, chosen time window} -pairs in deriving estimates of the parameters. By eliciting observable quantities the difficulty of directly judging the values of the failure intensity function parameters are avoided.
PaperV	A stakeholder compensation model, based on utility theory, is developed to support the decision-maker in defining incentives, i.e. a priori compensation, for the stakeholders to reach consensus decisions in a specified decision context. The compensation model is intended to be a basis for judging fairness and establishing fairness criteria in the context of Formal Safety Assessment. It incorporates compensation to social welfare functions in a new way not presented in the literature on social utlity to date.
Paper VI	A novel software measurement approach is developed to support decision-making in directing software development efforts and in evaluating acceptability of software. A method to classify the degree of expert consensus is also introduced. The classification method has wider applicability and may prove a valuable development for the research field of facilitated work groups.

4.2 Summary of types of expert judgement in the Papers

Table 6 summarises the types of expert judgement utilised in the Papers I–VI.

Table 6. Expert judgement elicited in the Papers I–VI.

Paper I	Expert judgements on different types of uncertainties are elicited for precautionary risk assessment and decision-making. Especially qualitative judgements pertaining to model uncertainty and bias are discussed. The form of judgements elicited during quality verification of risk assessment is also discussed. It is tentatively suggested that these be given in the form of 'yes'/'no'-answers.
Paper II	Experts are elicited for the specification of probability and frequency parameters of a fault tree model in a case study. Probability is interpreted in 'frequentistic' terms, supporting elicitation of uncertainty in the form of percentiles.
Paper III	Experts' judgements are elicited as percentiles related to an observable, but unknown, failure time. The percentile information is utilised to specify the expert model, which is a multivariate normal data model.
Paper IV	Expert judgements are elicited in the form of (number of failures, time window of observation) -pairs. The judgements are considered as data related to a perturbated, time-transformed, Poisson process.
PaperV	Expert judgements pertain to probability distributions of consequences of stakeholders in decision contexts involving multiple stakeholders.
Paper VI	Expert judgements are elicited in the form of triangular probability distribution functions related to achievement levels of software quality attributes.

4.3 Areas for future research

The conceptual framework presented for the qualification of risk assessment needs further development with respect to qualification criteria and peer review practices consolidating a decision-maker's confidence in the results and recommendations of risk assessment. A way forward may be the collaboration between scholars of risk analysis, risk communication and risk semantics. The framework may be developed in a direction that would support risk communication of risk issues to the public, as well as to the decision-maker.

An area for future research is also the definition and operationalisation of the concept of fairness in the context of risk informed decision-making. Fairness of decision-making with uncertain consequences may presuppose a dynamic scheme for re-allocation of compensation, as uncertainty is resolved.

Group Decision Support Systems have a great potential for supporting the expert judgement protocol and surveying of the stakeholders' feedback in risk assessment. In the case of the expert judgement protocol, graphical formats to aid elicitation, an interface facilitating the parameter specification process described, and an aggregation 'engine' computing and displaying the aggregated probability distribution of the target quantity would need to be implemented in the GDSS. In the case of stakeholder feedback, surveying techniques supporting scoring / voting are readily available in many GDSSs.

As a closing note, if the 'aphorism' in the Introduction is meaningful, what does it say about death?

References

Apostolakis, G., Garrick, B.J., Okrent, D. 1983. On Quality, Peer Review, and the Achievement of Consensus in Probabilistic Risk Analysis. Nuclear Safety, 24(6), 792–800.

Bedford, T., Cooke, R. 2001. Probabilistic Risk Analysis – Foundation and methods. Cambridge University Press, Cambridge.

Chakraborty, S., Breutel, C. 2000. Outlook for risk assessment and risk-informed regulations. Proceedings of PSAM 5th International Conference on Probabilistic Safety Assessment and Management, Osaka, Japan. Pp. 2017–2023.

Cooke, R.M. 1991. Experts in uncertainty. Oxford University Press, Oxford.

Cooke, R.M., Goossens, L.J.H. 2000. Procedures guide for structured expert judgment. European Commission, Nuclear Science and Technology, EURATOM, EUR 18820, Luxembourg.

Clemen, R.T., Winkler, R.L. 1999. Combining Probability Distributions from Experts in Risk Analysis. Risk Analysis, 19(2), 187–203.

Feldman, C.G. 1998. The practical guide to business process reengineering using IDEF0. Dorset House Publishing, New York.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. 1995. Bayesian Data Analysis. Chapman & Hall, London.

Harrison, S.R., Qureshi, M.E. 2000. Choice of stakeholder groups and members in multicriteria decision models. Natural Resources Forum, 24, 11–19.

Hattis, D., Anderson, E. 1999. What should be the implications of uncertainty, variability, and the inherent ‘biases’/’consevatism’ for risk management decision-making? Risk Analysis, 19(1), 95–107.

Hora, S.C., Iman, R.L. 1989. Expert Opinion in Risk Analysis: The NUREG-1150 Methodology. Nuclear Science and Engineering, 102, 323–331.

IEC 60300-3-9. 1995. Dependability management – Part 3: Application guide -– Section 9: Risk analysis of technological systems. International Electrotechnical Commission, Geneva, Switzerland.

IAEA. 1999. A framework for a quality assurance programme for PSA. International Atomic Energy Agency, Vienna. AEA-TECDOC-1101.

Keeney R.L., Raiffa, H. 1993. Decision with Multiple Objectives: Preferences and value trade offs. John Wiley & Sons, New York.

Keeney, R., vonWinterfeldt, D. 1991. Eliciting Probabilities from Experts in Complex Technical Problems. IEEE Trans. on Engineering Management, 38(3), 191–201.

Lindley, D. 2000. The philosophy of statistics. The Statistician, 49(3), 293–337.

Macgill, S.M., Fewtrell, L., Kay, D. 2000. Towards quality assurance of assessed waterborne risks. Water Resources, 34(3), 1050–1056.

Melchers, R.E. 2000. On the ALARP approach to risk management. Reliability Engineering & System Safety, 71, 2, 201–208.

Nielsen, T., Aven, T. 2003. Models and model uncertainty in the context of risk analysis. Reliability Engineering and System Safety, 79(3), 309–317.

Nyman, T., Sonninen, S., Hänninen, S., Rytkönen, J., Tuominen, R., Rosqvist, T., Riska, K., Juva. M., Palonen, A. 2002. The implementation of the VTMIS-system for the Gulf of Finland. Research report VAL34-013153. VTT Industrial Systems, Espoo, Finland.

Phillips, L.D. 1984. A Theory of Requisite Decision Models. Acta Psychologica 56, 29–48.

Phillips, D.F., Phillips, M.C. 1993. Facilitated Work Groups: Theory and Practice. J. Opl Res. Soc., 44(6), 533–549.

Pulkkinen, U. 1993. Methods for combination of expert judgements. Reliability Engineering and Systems Safety, 40, 111–118.

Pulkkinen U., Holmberg J. 1997. A method for using expert judgement in PSA. Säteiluturvakeskus (Radiation and Nuclear Safety Authority of Finland STUK), Technical report STUK-YTO-TR 129, Helsinki, Finland.

Pulkkinen, U., Simola, K. 2000. An expert panel approach to support risk-informed decision-making. Säteiluturvakeskus (Radiation and Nuclear Safety Authority of Finland STUK), Technical Report STUK-YTO-TR172, Helsinki, Finland.

Renn, O. 1992. Risk communication: towards a rational discourse with the public. Journal of Hazardous Materials, 29, 465–519.

Rosqvist, T., Tuominen, R., Jalonen, J., Salmikuukka, J., Reunanen, M. 1998. Formal Safety Assessment: Trial application of FSA to the transportation of dangerous goods on passenger/ro-ro cargo vessels. Report VAL B353. VTT Automation, Tampere, Finland.

Rosqvist, T., Tuominen, R., Reunanen, M., Laaksonen, R., Rintala, S., Sukselainen, J. 2001. FSA study on transportation of iron ore concentrate by pusher-barges on the Luleå–Raahe route. Report RIS C130. VTT Automation, Tampere, Finland.

Rouhiainen, V. 1992. QUASA: a method for assessing the quality of safety analysis. Safety Science, 15, 155–172.

Rowe, W.D. 1977. An Anatomy of Risk. John Wiley & Sons, New York.

Smith, A.F.M., Gelfand, A.E. 1992. Bayesian Statistics without Tears: A Sampling-Resampling Perspective. The American Statistics, 46(2), 84–88.

Simola, K. 2002. Methods of risk analysis (METRI) – Summary Report. In: Kyrki-Rajamäki, R. and Puska E.K. (eds.). FINNUS, The Finnish Research Programme on Nuclear Power Plant Safety 1999–2002, Final Report. VTT Research Notes 2164, Espoo, Finland. Pp. 227–236.

Tversky, A., Kahneman, D. 1974. Judgment Under Uncertainty: Heuristics and Biases. Science, 185, 124–1131.

United States Nuclear Regulatory Commission (USNRC), 1998. An Approach for Plant-specific, Risk-Informed Decisions on Plant-specific Changes to the Licensing Basis, Regulatory Guide 1.174. Washington, D.C.

West, M. 1988. Modelling Expert Opinion. In: J.M. Bernardo, M.H. DeGroot, D.V.Lindley, A. F. M. Smith (eds.). Bayesian Statistics 3, Oxford University Press. Pp. 493–508.

Winkler, R.L., Hora, S.C., Baca, R.G. 1992. The quality of experts' probabilities obtained through formal elicitation techniques. Center for nuclear waste regulatory analyses CNWRA, CNWRA Tech. Rep, San Antonio, Texas.

Zio, E., Apostolakis, G. 1996. Two methods for the structured assessment of model uncertainty by experts in performance assessments of radioactive waste repositories. Reliability Engineering and System Safety, 54(2–3), 225–241.

Øien, K., Hokstad, P.R. 1998. Handbook for performing expert judgement. SINTEF Industrial Management: SINTEF Report STF38 A98419, Trondheim, Norway.

Published by

Series title, number and report code of publication

VTT Publications 507
VTT–PUBS–507

Author(s)
Rosqvist, Tony

Title

One the use of expert judgement in the qualification of risk assessment

Abstract

The increased use of risk assesment in governmental and corporate decision-making has increased the role of expert judgement in providing information for safety related decision-making. Expert judgements are required in most steps of risk assessment: hazard identification, risk estimation, risk evaluation and analysis of options. The use and elicitation of expert judgement is therefore subject to on-going research. Furthermore, expert judgement is also required in the quality assurance or quality verification of risk assessment.

The research presented in the thesis addresses qualitative and probabilistic methods supporting the use of expert judgement in specific decision contexts; introduces a conceptual and procedural framework for quality verification of risk assessment; and presents techniques for the aggregation of probability distributions specified by experts' percentile information.

The methodological view to risk assessment adopted in the thesis is requisite modelling, where a decision and risk model is developed for a certain decision context to support decsision-making under uncertainty, and refined until the decision-maker has confidence in the results and prescriptions obtained from the model.

The decision and risk modelling approaches presented in the thesis are related to specific decision contexts in maritime safety, maintenance management, and software reliability. The modelling approaches presented can, however, be utilised for risk-informed decision-making in other application areas as well.

Keywords
risk assessment, quality of risk assessment, qualification of risk assessment, expert judgement, expert model, probabilistic safety assessment, formal safety assessment, risk modelling

Activity unit
VTT Industrial Systems, Tekniikantie 12, P.O.Box 1301, FIN–02044 VTT, Finland

ISBN			Project number
951–38–6243–7 (soft back ed.) 951–38–6244–5 (URL:http://www.vtt.fi/inf/pdf/)			
Date	Language	Pages	Price
November 2003	English	48 p. + app. 82 p.	C
Name of project		Commissioned by	
Series title and ISSN		Sold by	
VTT Publications 1235–0621 (soft back ed.) 1455–0849 (URL: http://www.vtt.fi/inf/pdf/)		VTT Information Service P.O.Box 2000, FIN–02044 VTT, Finland Phone internat. +358 9 456 4404 Fax +358 9 456 4374	