SINTEF

SINTEF A21560- Unrestricted

Report

Experiences from Using Indicators to Validate Expert Judgments in Security Risk Analysis

Author(s)
Olav Skjelkvåle Ligaarden, Atle Refsdal, and Ketil Stølen

SINTEF ICT
Networked Systems and Services
2012-01-09

SINTEF

SINTEF IKT
SINTEF ICT

Address:
Postboks 124 Blindern
NO-0314 Oslo
NORWAY

Telephone:+47 73593000
Telefax:+47 22067350

postmottak.IKT@sintef.no
www.sintef.no
Enterprise /VAT No:
NO 948 007 029 MVA

KEYWORDS:
Security risk analysis,
Expert judgment,
Indicator

Report

Experiences from Using Indicators to Validate Expert Judgments in Security Risk Analysis

VERSION
Final version

DATE
2012-01-09

AUTHOR(S)
Olav Skjelkvåle Ligaarden, Atle Refsdal, and Ketil Stølen

CLIENT(S)
Research Council of Norway

CLIENT'S REF.
180052/S10

PROJECT NO.
90B245

NUMBER OF PAGES/APPENDICES:
16/4

ABSTRACT
Expert judgments are often used to estimate likelihood values in a security risk analysis. These judgments are subjective and their correctness relies on the competence, training, and experience of the experts. Thus, there is a need to validate the correctness of the estimates obtained from expert judgments. In this paper we report on experiences from a security risk analysis where indicators were used to validate likelihood estimates obtained from expert judgments. The experiences build on data collected during the analysis and on semi-structured interviews with the client experts who participated in the analysis.

PREPARED BY
Olav Skjelkvåle Ligaarden

SIGNATURE
Olav S. Ligaarden

CHECKED BY
Bjørnar Solhaug

SIGNATURE
Bjørnar Solhaug

APPROVED BY
Bjørn Skjellaug

SIGNATURE

REPORT NO.
SINTEF A21560

ISBN
978-82-14-04998-5

CLASSIFICATION
Unrestricted

CLASSIFICATION THIS PAGE
Unrestricted

CONTENTS

Experiences from Using Indicators to Validate Expert Judgments in Security Risk Analysis

Olav Skjelkvåle Ligaarden*†, Atle Refsdal*, and Ketil Stølen*†
* *Department for Networked Systems and Services, SINTEF ICT, PO Box 124 Blindern, N-0314 Oslo, Norway*
E-mail: {olav.ligaarden, atle.refsdal, ketil.stolen}@sintef.no
† *Department of Informatics, University of Oslo, PO Box 1080 Blindern, N-0316 Oslo, Norway*

***Abstract*—Expert judgments are often used to estimate likelihood values in a security risk analysis. These judgments are subjective and their correctness relies on the competence, training, and experience of the experts. Thus, there is a need to validate the correctness of the estimates obtained from expert judgments. In this paper we report on experiences from a security risk analysis where indicators were used to validate likelihood estimates obtained from expert judgments. The experiences build on data collected during the analysis and on semi-structured interviews with the client experts who participated in the analysis.**

***Keywords*-security risk analysis; expert judgment; indicator**

I. Introduction

Much research report on procedures for eliciting expert judgment in risk analysis, decision support, and in general [1], [2], [3], [4]. There is also research that address the quality of expert judgments [5]. In this paper, however, we focus on the validation of the estimates obtained from expert judgments.

One way to validate likelihood estimates based on expert judgments is to use indicators calculated from historical data. Since we base ourselves on historical data, it is in most cases not possible to define indicators from which likelihood values can be inferred directly. For instance, in the case of the unwanted incident "eavesdropper reading a sensitive e-mail", an obvious indicator would be the number of times this has occurred in the past. However, as it is normally not feasible to calculate this from historical data, we will have to make do with other indicators that are less to the point. One potential indicator for this unwanted incident could for example be "the number of encrypted sensitive e-mails sent". Together with knowledge about the total number of sensitive e-mails being sent during a given period of time, this provides relevant input for validating the likelihood estimate.

In this paper we report on experiences from using indicators to validate expert judgments in a security risk analysis conducted in 2010. We build on data collected during the analysis and on semi-structured interviews with the client experts that participated in the analysis.

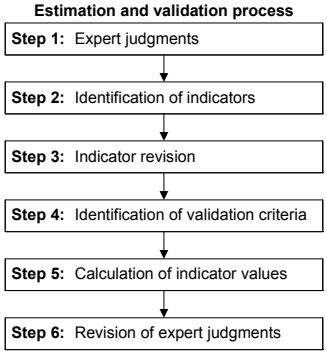

Figure 1. The estimation and validation process

The rest of the paper is structured as follows: in Section II we present the security risk analysis from 2010. In Section III we present the most relevant data collected during the analysis. In Section IV we discuss this data in relation to input from the semi-structured interviews, while in Section V we conclude.

A more detailed description of the security risk analysis case is given in Appendix A. Appendix B presents all the data collected during the analysis. In Appendix C we present the thematic analysis of the semi-structured interviews, while in Appendix D threats to the validity of our conclusions are discussed.

II. Security risk analysis case

Our empirical study was integrated in a commercial security risk analysis based on CORAS [6] conducted in 2010. As the client of this analysis requires full confidentiality we can not report on the system assessed, the risk models obtained, the personnel from the client involved, or the name of the client.

Fig. 1 depicts the fragment of the security risk analysis of relevance for this paper as a process of six steps. Step 1 is the likelihood estimation based on expert judgments. Indicators for validation of the likelihood estimates were identified in Step 2. The analysis team proposed a number of indicators, and these indicators were revised during a meeting with the client experts in Step 3. During this meeting some indicators were rejected, some were subject to minor modifications,

and some new indicators were identified. In Step 4 the analysis team formulated validation criteria for the likelihood estimates in terms of indicators. Each criterion specifies the expected values of the indicators related to the likelihood estimate in question. Here, each criterion makes a prediction about the value of a set of indicators under the assumption that the likelihood estimate in question is correct. Indicator values were obtained by the client experts in Step 5. In Step 6 the validation criteria were evaluated and some of the initial likelihood estimates were adjusted.

In total, an estimated number of 400 hours were spent on the security risk analysis (not including writing the final report) by the analysis team. Three domain experts (E1, E2, and E3) of the client participated in the analysis. The client experts held degrees equivalent to Master of Science and four to ten years of experience in information security and risk analysis.

III. Data collected from case

For each step of the estimation and validation process, depicted in Fig. 1, we collected data, documented in Table I.

In **Step 1**, we came up with 28 likelihood estimates based on expert judgments.

In **Step 2**, the analysis team proposed at least one indicator for each of the 28 likelihood estimates. In total, 68 indicators were proposed.

In **Step 3**, the indicators proposed in Step 2 were revised in a meeting with the client experts. Some of the proposed indicators were rejected during this meeting, because their values were not obtainable within the client's organization. After Step 3, there were 25 out of 28 likelihood estimates with at least one indicator. In total, 57 indicators remained after Step 3 had been conducted.

In **Step 4**, 19 indicators were used by the analysis team to formulate validation criteria for 15 likelihood estimates. For 10 likelihood estimates, validation criteria were not formulated. One of these 10 likelihood estimates was not assigned a criterion because the validation of the estimate was given a low priority by the client experts[1]. For the remaining nine likelihood estimates, the analysis team was not able to come up with good validation criteria, although the indicators were considered to provide relevant information for validating the likelihood estimates.

In **Step 5**, the client experts obtained values for 13 out of the 19 indicators used to formulate the 15 validation criteria. This resulted in that only 10 out of the 15 validation criteria could be evaluated after Step 5.

In **Step 6**, we evaluated 10 validation criteria based on the values obtained by the client experts. The validation criteria were fulfilled for four likelihood estimates, while for two likelihood estimates we could not say whether the criteria were fulfilled or not, because the values of the indicators referred to in the criteria were too uncertain. The criteria were not fulfilled for the remaining four likelihood estimates. For two of these estimates, the client experts decided to adjust the likelihood estimates.

[1]The likelihood estimate was associated with an unwanted incident. For the incident to occur, some technology needed to be used, which was not yet implemented at the time of analysis. Thus, the client considered the likelihood estimate for this incident to be less important.

Table I
Relevant data for the different steps of the estimation and validation process

1	Number of likelihood estimates based on expert judgments after Step 1.	28
2	Total number of indicators after Step 2.	68
	Number of likelihood estimates with at least one indicator after Step 2.	28
3	Total number of indicators after Step 3.	57
	Number of likelihood estimates with at least one indicator after Step 3.	25
4	Total number of indicators used to formulate validation criteria after Step 4.	19
	Number of likelihood estimates with a validation criterion after Step 4.	15
5	Total number of indicators used to formulate validation criteria for which the client experts obtained values after Step 5.	13
	Number of likelihood estimates for which validation criteria could be evaluated after Step 5.	10
6	Number of likelihood estimates with a fulfilled validation criterion after Step 6.	4
	Number of likelihood estimates with a not fulfilled validation criterion after Step 6.	4
	Number of likelihood estimates with a validation criterion where it was undecided whether the criterion is fulfilled or not after Step 6.	2
	Number of likelihood estimates with a not fulfilled validation criterion for which the likelihood estimates were adjusted after Step 6.	2

IV. Discussion

With each of the three client experts, we conducted a semi-structured interview, focusing on likelihood estimation based on expert judgments and the use of indicators to validate these. The transcribed interviews were analyzed by the use of a simplified version of thematic analysis [7]. In this section we discuss the data in Table I in relation to the results from the thematic analysis.

A. Step 1

Experts E1 and E2 were quite confident that they their likelihood estimates were correct, while expert E3 did not want to give a clear yes or no answer to this. Even though they believed the estimates to be correct, expert E1 pointed out that validation in terms of indicators still has a purpose: *"... I think there were one or two such cases where we had to adjust the estimates because of their indicator values. So I think the quality was good anyway but it is still an extra*

quality adjustment when you get acknowledgments or only minor adjustments of the estimates."

B. Step 2 and 3

It was challenging to identify relevant indicators for which values could actually be obtained within the available time and resources for the analysis. Expert E1 supports this: *"It was a challenge in the least because it is terribly difficult to find good indicators of information security, and there were a number of examples where it was actually not possible to find indicators. Even though we had proposals we discovered later that they were not usable. But there were also areas where we came up with indicators that could be used."*

During the revision meeting in Step 3, many indicators were rejected because their values were not obtainable within the client's organization. This resulted in that three likelihood estimates were left without indicators. One might argue that we should have used more time to identify indicators in Step 2, and also that we should have involved the client experts in this step. With respect to the former argument, according to our records we spent about 50 hours to identify indicators in Step 2, which is quite a lot when considering that about 400 hours were spent on the whole security risk analysis. With respect to the latter argument, all three client experts were of the opinion that the analysis team should come up with the initial indicator proposals. Expert E1 even expressed: *"... I also think that when it comes to indicators, it can be a strength that they are proposed by someone else who does not have built-in limitations with respect to ideas."*

On the other hand, we could perhaps have obtained information from the client experts on the kinds of data, in the form of logs and so on, that are available within their company, prior to identifying indicators in Step 2. This would most likely have resulted in fewer indicator proposals being rejected due to their values not being obtainable. On the other hand, proposing relevant indicators where their values are not obtainable at the time of analysis may also prompt the client organization to implement more measurements, as expressed by expert E2: *"It turned out that some of the measurements that had not been carried out should perhaps have been carried out, and that is the experience obtained from what we found here."*

C. Step 4

The analysis team was not able to formulate validation criteria for nine out of 25 likelihood estimates. We do not have the opinions of the client experts on this matter. They were not asked to comment on the formulation of validation criteria in the interviews, since this task was conducted solely by the analysis team.

The indicators of the nine estimates were considered relevant for validating the estimates, but we could not figure out how to link them to the estimates. Common for these indicators is that they are only indirectly linked to the estimates of which they were seen as relevant. An example of such an indicator is "the number of code lines used to produce the web server application" which is indirectly linked with the likelihood estimate of the unwanted incident "hacker takes over the web server by exploiting weaknesses in its code". In many cases it is reasonable to believe that the number of weaknesses will increase with the number of code lines. However, it is not easy to predict how the value of this indicator affects the likelihood estimate since the estimate depends on a lot of other factors as well. On the other hand, the indicator "the number of times the web server was taken over by hackers during the past five years due to weaknesses in its code" is directly linked with the likelihood estimate of the incident. It is not surprising that it is easier to formulate validation criteria based on this kind of more direct indicators than by the use of the more indirect ones. Eight out of the 10 validation criteria evaluated in Step 6 used an indicator that is directly linked to the likelihood estimate. In relation to this it must be pointed out that we would have had seven validation criteria using solely indirect indicators if we had managed to obtain all the indicator values in Step 5 needed for evaluating the 15 validation criteria.

D. Step 5

For five of the validation criteria the client experts did not manage to obtain the indicator values necessary for evaluating the criteria. One reason may be that obtaining all the indicator values required too much effort. The client experts tried to obtain values for 49 out of the 57 indicators remaining after Step 3. Out of the 19 indicators that ended up being used in validation criteria, they managed to obtain values for 13. They may have succeeded for a higher proportion if we had only requested the values for the 19 indicators being used in validation criteria. The reason for requesting all indicator values was that the validation criteria were formulated after the value collection process had been initiated, and before we received the indicator values from the client experts. Thus, we did not know at the time when the value collection process was initiated which indicators we would use to formulate the validation criteria. It would have been better to first identify the indicators needed in the validation criteria, and then ask the client experts to obtain values for those.

Another reason for failing to obtain six of the necessary values may have been that the client experts postponed the task a little too long. This is very likely since we know that many of the indicator values where obtained just before the given deadline. But it can also be the case that the values were not as easily available as first expected. Expert E2 supports the latter: *"... for me the process went pretty smoothly. I got answers if there had been done measurements, but I also got feedback like "we have no*

idea"."

All three experts interviewed believe that indicator values of high quality were obtained. It is however a bit uncertain whether this was actually the case. We know, for instance, that some of the values obtained were just new expert judgments by other experts. Expert E2 told us that he obtained indicator values by asking other people working at the company: *"The hardest part is to find the right person who has the right competence. It was pretty easy to find the answers for those indicators where there were numbers if we found the right person. In our case there were actually two-three persons who answered all."* It is however a bit uncertain how many of the obtained indicator values that were just new expert judgments.

E. Step 6

Two likelihood estimates were changed by the client experts as a result of their validation criteria being falsified. When changing the likelihood estimate of an unwanted incident, its risk level will often change as well, since the risk level depends on the likelihood value and the consequence value of the unwanted incident. A change in the risk level will often have consequences for the type of treatments that are implemented. In our case, however, the risk levels associated with the two unwanted incidents did not change when their likelihood estimates were updated.

In the case of a validation criterion being falsified we can not straightforwardly conclude whether likelihood estimates should be changed or kept as they are. For instance, although we manage to obtain correct indicator values, it may be that the validation criterion does not capture what we believe it does. In the risk analysis we had two cases where the client experts decided not to adjust the likelihood estimates of two unwanted incidents, even though their validation criteria were falsified. In the case of the first incident, the client experts kept the likelihood estimate because the value of the indicator used in the criterion did not represent a typical value. In the case of the second incident, its likelihood estimate was, according to its validation criterion, equal to zero since some technology required for the incident to occur was not in use at the time of analysis. As a consequence of this the incident should have been removed from the threat model. The client experts wanted the threat model to reflect the situation were the technology was in place, and the threat model was therefore not changed. Also, it was no harm in keeping the incident, since it did not result in any unacceptable risks needing treatments, due to a low likelihood estimate.

V. Conclusion

In this paper we have presented experiences from using indicators to validate likelihood estimates based on expert judgments in a security risk analysis conducted in 2010.

The use of indicators brought forward new information resulting in two out of 28 likelihood estimates being changed.

We also identified some challenges that need to be addressed in order to get the most out of indicator-based validation. First, it is challenging to identify indicators for which it is feasible to obtain values within the available time and resources for the analysis. For a number of the indicators identified, their values were not obtainable within the client's organization. By having some knowledge on the kinds of historical data that are available within the organization and whose responsible for the different kinds of data, it should be easier to both identify indicators and obtaining their values. Unfortunately, it may be difficult to obtain this knowledge since data is often spread across the organization and since few, if any, have a complete overview of the data available. Second, it is challenging to formulate validation criteria for likelihood estimates in terms of indicators. It is especially difficult to predict how indicator values affect a likelihood estimate when the indicators are only indirectly related to the estimate in question. This will typically be a problem when formulating validation criteria for likelihood estimates of incidents that that are not easily observable. Third, the indicator values obtained from an organization may vary when it comes to correctness. In order to get the most out of the validation, the uncertainty of the values should be taken into account. Moreover, one should strive to reduce uncertainty by using several independent indicators to validate the same estimate.

Acknowledgments

The research on which this paper reports has been carried out within the DIGIT project (180052/S10), funded by the Research Council of Norway, and the MASTER and NESSoS projects, both funded from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreements FP7-216917 and FP7-256980, respectively.

References

[1] M. A. Meyer and J. M. Booker, *Eliciting and Analyzing Expert Judgment: A Practical Guide*, ser. ASA-SIAM series on statistics and applied probability. SIAM, 2001.

[2] R. M. Cooke and L. H. J. Goossens, "Procedures Guide for Structured Expert Judgment in Accident Consequence Modelling," *Radiation Protection and Dosimetry*, vol. 90, no. 3, pp. 303–311, 2000.

[3] L. H. J. Goossens, R. M. Cooke, A. R. Hale, and L. Rodic-Wiersma, "Fifteen Years of Expert Judgement at TUDelft," *Safety Science*, vol. 46, no. 2, pp. 234–244, 2008.

[4] H. Otway and D. von Winterfeldt, "Expert Judgment in Risk Analysis and Management: Process, Context, and Pitfalls," *Risk Analysis*, vol. 12, no. 1, pp. 83–93, 1992.

[5] F. Bolger and G. Wright, "Assessing the Quality of Expert Judgment: Issues and Analysis," *Decision Support Systems*, vol. 11, no. 1, pp. 1–24, 1994.

[6] M. S. Lund, B. Solhaug, and K. Stølen, *Model-Driven Risk Analysis: The CORAS Approach*, 1st ed. Springer, 2011.

[7] D. Ezzy, *Qualitative Analysis: Practise and Innovation*, 1st ed. Routledge, 2002.

[8] V. Braun and V. Clarke, "Using Thematic Analysis in Psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.

APPENDIX A.
SECURITY RISK ANALYSIS CASE

This appendix provides an extended description of the security risk analysis case, briefly presented in Section II. The analysis involved an analysis team of two analysts and three client experts (E1, E2, and E3) that participated on behalf of the client. In the analysis the client experts presented the analysis team with a business objective that they need to comply with. To comply with this business objective, a number of requirements need to be satisfied. The purpose of the analysis was to document the different risks that may arise and harm assets of the client if the requirements are not satisfied. To conduct the analysis, the CORAS approach [6] was used. CORAS provides a method, a language, and a tool for asset-oriented risk analysis. The analysis was conducted at the following three levels:

- **Level 1:** Identify vulnerabilities with respect to the requirements. Select the vulnerabilities that require further analysis.
- **Level 2:** Document possible high-level threat scenarios that may arise if the selected vulnerabilities from Level 1 are exploited. Select the threat scenarios that require further analysis.
- **Level 3:** Conduct a detailed analysis of the selected threat scenarios from Level 2.

As seen in Fig. 2, the first four steps of the CORAS method were conducted as part of the Level 1 and 2 analyses, while the four last were conducted during the Level 3 analysis. As part of the Level 3 analysis we validated likelihood estimates, obtained from expert judgments, by the use of indicators. In the following we focus our attention on the steps of the CORAS method relevant for the Level 3 analysis and their relation to the steps of the estimation and validation process. For more details on the different steps of the CORAS method, see [6].

The relation between the Level 3 analysis and the estimation and validation process is depicted in Fig. 2. In the following we refer to Step *X* of the CORAS method and Step *Y* of the estimation and validation process as Step CM*X* and Step EVP*Y*, respectively.

In Step CM5 we used the selected threat scenarios from the Level 2 analysis as a starting point for identifying and documenting unwanted incidents as well as threats, vulnerabilities, and the threat scenarios leading up to the unwanted incidents, by the use of CORAS threat diagrams. In Step CM6 the client experts estimated likelihoods for the unwanted incidents identified in Step CM5, and their consequences with respect to different assets. All the estimates were based on expert judgments. Thus, Step EVP1 was conducted as part of this step.

After conducting Step EVP1, the analysis team identified indicators in Step EVP2 for validating likelihood estimates assigned to the different unwanted incidents. The indicators identified by the analysis team were revised during a meeting with the client experts in Step EVP3. During this meeting some indicators were rejected, some were subject to minor modifications, and some new indicators were identified.

In Step EVP4 the analysis team formulated validation criteria for validating likelihood estimates, based on expert judgments, in terms of the identified indicators. As previously explained, each criterion specifies the expected values of the indicators related to the likelihood estimate in question. Here, each criterion makes a prediction about the value of a set of indicators under the assumption that the likelihood estimate in question is correct.

Indicator values were obtained by the client experts in Step EVP5. In Step CM7 we estimated risk levels for the different risks. As part of this step we conducted Step EVP6 where we validated likelihood estimates, based on expert judgments, by the use of the validation criteria. Some of the likelihood estimates were adjusted as a result of their validation criteria not being fulfilled. In Step CM8 we identified treatment options for the risks classified as unacceptable in Step CM7.

Table II shows estimates for the number of hours that were spent on each level of the analysis, as well as estimates for the number of hours that were spent on the steps of the Level 3 analysis and the estimation and validation process. The estimates are based on the analysis team's own notes from the analysis. For the client experts we only have numbers for meetings. Thus, we do not know how much time they spent between meetings, but we know that it was considerably less than the time spent by the analysis team. In total, an estimated number of 400 hours were spent on the full analysis (not including writing the final report) by the analysis team. About 60% of these hours were spent on the Level 3 analysis. The table shows that a large amount of the hours spent on the Level 3 analysis were spent in relation to indicators.

As already explained, three domain experts participated on behalf of the client. Table III shows the education of the client experts, as well as their experience within information security and risk analysis.

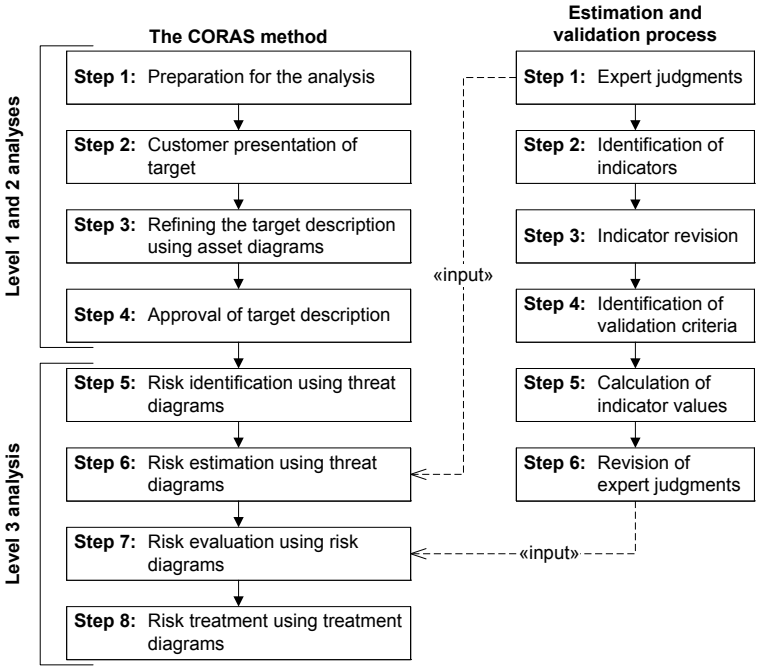

Figure 2. The CORAS method, the estimation and validation process, and their relations

Table II
ESTIMATES FOR THE NUMBER OF HOURS SPENT ON THE DIFFERENT LEVELS AND THE STEPS OF THE LEVEL 3 ANALYSIS (CM 5–8) AND THE ESTIMATION AND VALIDATION PROCESS (EVP 1–6)

	Estimated number of hours spent by	
Level	**Analysis team**	**Client experts (meetings only)**
1	75	10
2	75	10
3	250	65
Total	**400**	**85**
Step(s)	**Analysis team**	**Client experts (meetings only)**
CM5	35	15
CM6 / EVP1	40	15
EVP2	50	-
EVP3	20	15
EVP4	45	-
EVP5	-	-
CM7 / EVP6	30	10
CM8	30	10
Total	**250**	**65**

APPENDIX B.
DATA COLLECTED FROM CASE

Table I only presents some of the data collected for the different steps of the estimation and validation process. In this appendix we present all the data collected. We provide explanations for all the different data collected, even the data explained in Section III. The data is presented in Table IV.

Table III
EDUCATION AND EXPERIENCE OF CLIENT EXPERTS

Expert	**Education and experience**
E1	Degree equivalent to Master of Science. Ten years of experience in information security. Ten years of experience in risk analysis.
E2	Degree equivalent to Master of Science. Ten years of experience in information security. Four years of experience in risk analysis.
E3	Degree equivalent to Master of Science. Ten years of experience in information security. Eight years of experience in risk analysis.

Some of the rows in Table IV refer to rows in Table I to make it easier for the reader to understand which of the data items that are found in both tables.

A. Data for Step 1

The risk analysis resulted in 28 (ID 1.1 in Table IV) likelihood estimates based on expert judgments, where each estimate is associated with an unwanted incident.

B. Data for Step 2

In Step 2 the analysis team identified one or more indicators for each of the 28 (2.3) likelihood estimates. In total, 81 (2.1) indicators were proposed by the analysis team, of which 68 (2.2) were unique[2]. Even though it has not

[2]A number of the indicators were used for more than one likelihood estimate.

Table IV
ALL THE DATA COLLECTED FOR THE DIFFERENT STEPS OF THE ESTIMATION AND VALIDATION PROCESS

Step 1: Expert judgments			
ID	*Definition*	*Value*	*Row no. Table I*
1.1	Number of likelihood estimates based on expert judgments after Step 1.	28	1
Step 2: Identification of indicators			
ID	*Definition*	*Value*	*Row no. Table I*
2.1	Total number of indicators after Step 2.	81	-
2.2	Number of unique indicators after Step 2.	68	2
2.3	Number of likelihood estimates with at least one indicator after Step 2.	28	3
2.4	Number of likelihood estimates after Step 2 with indicators that are directly linked to the estimates.	19	-
Step 3: Indicator revision			
ID	*Definition*	*Value*	*Row no. Table I*
3.1	Total number of indicators after Step 3.	68	-
3.2	Number of unique indicators after Step 3.	57	4
3.3	Number of likelihood estimates with at least one indicator after Step 3.	25	5
3.4	Number of likelihood estimates with zero indicators after Step 3.	3	-
3.5	Total number of new indicators added during Step 3.	3	-
3.6	Number of unique new indicators added during Step 3.	2	-
3.7	Total number of indicators rejected during Step 3.	16	-
3.8	Number of unique indicators rejected during Step 3.	13	-
3.9	Number of likelihood estimates after Step 3 with indicators that are directly linked to the estimates.	11	-
Step 4: Identification of validation criteria			
ID	*Definition*	*Value*	*Row no. Table I*
4.1	Total number of indicators used to formulate the validation criteria after Step 4.	20	-
4.2	Number of unique indicators used to formulate the validation criteria after Step 4.	19	6
4.3	Number of likelihood estimates with a validation criterion after Step 4.	15	7
4.4	Number of likelihood estimates with a validation criterion after Step 4, where indicators that are directly linked to the estimates are used in the criteria.	8	-
4.5	Number of likelihood estimates without a validation criterion after Step 4, due to that the validation of the likelihood estimate was given a low priority.	1	-
4.6	Number of likelihood estimates without a validation criterion after Step 4, due to the analysis team not being able to formulate validation criteria based on the indicators associated with the likelihood estimates.	9	-
Step 5: Calculation of indicator values			
ID	*Definition*	*Value*	*Row no. Table I*
5.1	Number of unique indicators used to formulate validation criteria for which the client experts obtained values after Step 5.	13	8
5.2	Number of likelihood estimates for which validation criteria could be evaluated after Step 5.	10	9
5.3	Number of likelihood estimates for which validation criteria could not be evaluated after Step 5.	5	-
5.4	Number of likelihood estimates for which validation criteria could be evaluated after Step 5, where indicators that are directly linked to the estimates are used in the criteria.	8	-
5.5	Number of unique indicators for which the client experts tried to obtain values after Step 5.	49	-
5.6	Number of unique indicators for which the client experts managed to obtain values after Step 5.	37	-
Step 6: Revision of expert judgments			
ID	*Definition*	*Value*	*Row no. Table I*
6.1	Number of likelihood estimates with a fulfilled validation criterion after Step 6.	4	10
6.2	Number of likelihood estimates with a not fulfilled validation criterion after Step 6.	4	11
6.3	Number of likelihood estimates with a validation criterion where it was undecided whether the criterion is fulfilled or not after Step 6.	2	12
6.4	Number of likelihood estimates with a not fulfilled validation criterion for which the likelihood estimates were adjusted after Step 6.	2	13

been stated explicitly, we only focus on unique indicators in Table I. In Table IV we distinguish between the number of unique indicators and the total number of indicators.

For 19 (2.4) likelihood estimates, the analysis team proposed indicators that are directly linked to the estimates. The distinction between indirectly and directly linked indicators is explained in Section IV-C. As explained in Section IV-C, indicators that are directly linked to a likelihood estimate have often an advantage over the indirect ones when it comes to validating the likelihood estimate.

C. Data for Step 3

The indicator proposals were revised in Step 3 during a meeting with the client experts. After the revision had been conducted, 25 (3.3) out of the 28 likelihood estimates were associated with one or more indicators. In total this left us with 68 (3.1) indicators, of which 57 (3.2) were unique. During the revision meeting, three (3.5) new indicators, of which two (3.6) were unique, were added. 16 (3.7) of the proposed indicators were rejected during the same meeting, mostly due to their values not being obtainable within the client's organization. Of these indicators, 13 (3.8) were unique. The result of rejecting these indicators was that three (3.4) likelihood estimates were left without indicators.

As we can see from the data of Step 2 and 3 in Table IV, almost 20% of the proposed indicators were rejected. Also, a number of indicators that are directly linked to the estimates of which they were seen as relevant, were rejected in Step 3. The result of this was that only 11 (3.9) out of 19 (2.5) likelihood estimates were left with indicators that are directly linked to the estimates after Step 3 had been conducted.

D. Data for Step 4

Validation criteria were formulated by the analysis team, in Step 4 of the estimation and validation process, for 15 (4.3) out of the 25 likelihood estimates with indicators, of which eight (4.4) of the validation criteria, or more than 50%, use indicators that are directly linked to the estimates. For the remaining ten likelihood estimates, validation criteria were not formulated. One (4.5) of these estimates was not assigned a criterion due to that the validation was given a low priority by the client experts[3]. For the other nine (4.6) likelihood estimates, the analysis team was not able to formulate validation criteria based on their indicators, although the indicators were considered to provide relevant information in terms of validating the estimates. To formulate the validation criteria the analysis team used 20 (4.1) indicators, of which 19 (4.2) are unique. Thus, only 33% of the unique indicators left after the revision meeting in Step 3 was used for formulating validation criteria.

[3]The likelihood estimate was associated with an unwanted incident. For the specific incident to occur, some technology needed to be used, which was not yet implemented at the time of analysis. Thus, the client considered the likelihood estimate for this incident to be less important.

E. Data for Step 5

In Step 5, values were obtained for a number of the indicators remaining after the revision meeting in Step 3. Note that the analysis team did not restrict the client experts' collection of indicator values to the 19 (4.2) unique indicators used in the validation criteria. The validation criteria were formulated after the collection process had been initiated, and before we received the indicator values from the client experts. Thus, we did not know at the time the collection process was initiated which indicators we would actually be able to use in the validation criteria.

The client experts tried to obtain values for 49 (5.5) out of the 57 (3.2) unique indicators that remained after Step 3 had been conducted. Of these 49 unique indicators, the experts managed to obtain values for 37 (5.6). Thus, the experts managed to obtain 75% of the values. 13 (5.1) of the obtained values belong to indicators used in the validation criteria. This means that six of the values needed in the evaluation of the validation criteria were not obtained. The result was that five (5.3) of the likelihood estimates with validation criteria could not be evaluated. This meant that we could only validate ten (5.2) of the likelihood estimates, of which eight (5.4) use indicators in their validation criteria that are directly linked to the estimates.

F. Data for Step 6

Ten likelihood estimates were validated in Step 6. For four (6.1) of the estimates the criteria were fulfilled, for two (6.3) we could not say whether the criteria were fulfilled or not because the indicator values used in the criteria were too uncertain, while for the remaining four (6.2) the criteria were not fulfilled. For two (6.4) out of these four likelihood estimates, the client experts decided to adjust the estimates, while the two other estimates remained unchanged.

Appendix C. Thematic analysis

A. Procedure for collecting data

We conducted a semi-structured interview with each of the three client experts that participated in the security risk analysis. In these interviews the experts were asked open-ended questions related to estimation of likelihood values based on expert judgments and the use of indicators to validate these. Each question had a number of prompts (follow-up questions), that were asked if the client experts did not answer them as part of the open-ended question. All the interviews were recorded and conducted in Norwegian. The Norwegian Social Science Data Services[4] has been notified about the interviews, as a result of personal data being collected, recorded, and stored.

[4]See http://www.nsd.uib.no/nsd/english/index.html and http://www.nsd.uib.no/nsd/english/pvo.html for more information.

The two analysts had different roles in the interviews. One analyst had the role as the interviewer, while the other acted as an observer. Almost all of the interaction was between the interviewer and the interviewee. The main tasks of the observer were to administer the recordings of the interviews, to take additional notes if necessary, and to make sure that the necessary information was collected. The observer only interacted with the interviewee in cases where he felt that additional answers were needed.

Each interview was introduced by the interviewer explaining the purpose of the interview and the terms under which the interview would be conducted. The interviewees were informed that the purpose of the interview was to collect empirical data to be used in an evaluation report on the security risk analysis they had participated in. They were also told about our intention to publish this report, that they would appear anonymous in this report, and that the report would not contain any confidential information. The interviewees were also informed that they could withdraw from the interview at any time, without giving any form of explanation, and that the interview would be recorded and that a non-confidential transcript of each recorded interview would be stored as background material for the evaluation report for a period of ten years. Before starting each interview, the interviewer asked the interviewee whether he/she accepted the terms, including that the interview would be recorded.

An interview guideline, given below, was used for conducting the interviews. Only the interviewer and the observer had access to the guideline. In the guideline, *Q* stands for open-ended question, while *P* stands for prompt. Also, the guideline states the topic addressed by each open-ended question, and whether a handout was provided to the interviewees for the purpose of triggering their memory. The interviewees were only allowed to have a short look at the handouts, to ensure that they did not become too focused on specific details.

- First, we want you to answer the following:
  - What kind of educational background do you have?
  - How many years of experience do you have with risk analysis?
  - How many years of experience do you have with ICT security?

- **Q1:** *[Topic: General question about estimation.] [Handout: Printout of all the threat diagrams.]*
  As you know, a risk level depends on the frequency and consequence of an unwanted incident. A part of the risk analysis focused therefore on estimating the frequency and consequence of the identified incidents. How did you experience that part of the process - that is, the frequency and consequence estimation?
  - **P1:** To what extent do you think we generally came up with correct estimates?
  - **P2:** Can you say something about possible sources to uncertainty with respect to the correctness of the estimates?
  - **P3:** Was there estimates that were particularly difficult or easy to do? If so, which ones and why?
  - **P4:** Was there any estimates where you had more confidence in their correctness than others? If so, which ones and why?
  - **P5:** What do you think about the approach used for estimation?

- **Q2:** *[Topic: About identification of indicators.] [Handout: Overview of unwanted incidents and their final indicators.]*
  For some of the incidents we identified indicators to support the estimation of their frequencies. How did you experience this sub-process - that is, the identification of relevant indicators?
  - **P1:** What worked well/badly?
  - **P2:** How difficult was it to identify relevant indicators?
  - **P3:** Was there any incidents of which identification of indicators were particularly easy or difficult? If so, why?
  - **P4:** Do you think in general that there exists categories of unwanted incidents of which indicator identification is particularly easy or difficult?
  - **P5:** We (the analysis team) came up with proposals for indicators that were presented and discussed at a meeting. What difference do you think it would have had if we had asked you to think through whether these were good indicators between two meetings, and possibly discussed them with colleagues?

- **Q3:** *[Topic: About obtaining indicator values]*
  After having identified relevant indicators, the next step was to obtain the indicator values. Can you tell us how you experienced this part of the process?
  - **P1:** To what extent was it difficult to obtain indicator values?
  - **P2:** Was there any indicators for which it was particularly easy or difficult to obtain values?
  - **P3:** To what extent do you trust that the obtained indicator values were correct?
  - **P4:** What kind of approach do you think we should use for identifying indicators that it is possible to obtain values for?

- **Q4:** *[Topic: From indicator values to frequency estimates.]*
  Indicator values were used to support the frequency

estimation for some of the unwanted incidents. How did you experience this part of the process?
  - **P1:** How difficult was it to estimate frequency on the basis of indicator values?
  - **P2:** Was there any cases where is was particularly easy or difficult to estimate frequency on the basis of indicator values?
  - **P3:** Do you have any suggestions for methods that could be used to come up with estimates based on indicator values?
  - **P4:** One possible way for coming up with indicator-based estimates could be to define the frequency values as a function of one or more indicators. For example, one could say things like that "If $I1 < 100$ and $I2 < 50$, then the frequency is Seldom, Otherwise ... etc.". Do you think such an approach would work well? Why/why not?

- **Q5:** *[Topic: Indicator-based estimates versus estimates based solely on expert judgments.]*
  Some of the frequency estimates were based on a combination of indicators and expert judgments, while others were based solely on expert judgments. To what extent do you think this difference is important for the quality of the estimates?
  [Make sure to get an explanation.]
  - **P1:** Does the difference have any relevance for the correctness of the estimates? Why/why not?
  - **P2:** Does the difference have any impact on yours or others' confidence in the correctness of the estimates? Why/why not?

- **Q6:** *[Topic: Possible benefit of indicator values as documentation.]*
  In the analysis report, indicator values can be included as documentation for the foundation for frequency (and possibly consequence) estimates. To what extent do you think this will affect the quality of the report? In what way could it be affected?
  - **P1:** Imagine that you are going to read a report from an analysis you have not participated in. Do you think it will affect your confidence in the correctness of the estimates if the report documents the indicators that were used for the various estimates, and their values?
  - **P2:** Imagine that you are participating in an analysis and contribute to making the estimates. To what extent do you think it is important to document the foundation for the estimates in terms of indicator values?

- **Q7:** *[Topic: About the importance of good estimates.]*
  The estimation of frequency and consequence values for unwanted incidents - and thus risk level - is only a part of the whole risk analysis. How important do you think it is that you arrive at good/correct estimates?
  - **P1:** Can you give some indication of the amount of resources you think is right to use on this part of the analysis in relation to the rest?
  - **P2:** What do you think is reasonable to expect from the documentation in an analysis report regarding the foundation of the estimates?

B. Procedure for analyzing data

The recorded interviews were transcribed not long after the interviews had been conducted. Based on these transcripts, non-confidential transcripts were created by removing all the text mentioning the system assessed, the risk models obtained, the personnel from the client involved, and the name of the client.

The data set (the three non-confidential transcripts) have been analyzed by the use of a simplified version of thematic analysis [7], which is a method for identifying, analyzing, and reporting patterns (themes) within data. In [8], theme is defined as follows: *"A theme captures something important about the data in relation to the research question, and represents some level of patterned response or meaning within the data set."* The data was analyzed with the following overall research question in mind: *"To what extent may indicators be used to validate likelihood values obtained from expert judgments in a security risk analysis?*

Thematic analysis offers a lot of flexibility as a method, i.e. that it can be performed in many different ways. What separates one thematic analysis from another are, among other things: whether the themes focus on the whole data set or only parts of it; whether themes are identified by an inductive or theoretical approach; whether themes are identified at the semantic or latent level; and the research epistemology used, i.e. what you can say about your data and how you theorize meaning.

In our thematic analysis we do not try to give a rich thematic description of the entire data set. We rather focus on giving a detailed account of a group of themes within the data, which are of relevance to the overall research question. We can therefore say that our themes are identified in a theoretical way. Furthermore, we identify themes at the semantic level. This means that themes are identified within the explicit or semantic meanings of the data. We do not try to go beyond what the interviewee said in the interview. In the case of research epistemology we have followed an essentialist/realist approach. The research epistemology states what we can say about the data and it informs how we theorize meaning. In the essentialist/realist perspective, meaning and experience inhere with individuals.

Inspired by the process for thematic analysis outlined in [8], our thematic analysis was conducted as follows:

Table V
THEMES, IDENTIFIED IN THE THEMATIC ANALYSIS, AND THE DATA EXTRACTS DISTRIBUTION OVER THESE THEMES

Theme – Description	No. of data extracts
Theme 1: Estimation – Reflections about estimation of likelihood and consequence.	18
Theme 2: Indicator identification – Reflections about the identification of indicators.	24
Theme 3: Obtaining indicator values – Reflections about obtaining indicator values.	11
Theme 4: Validation – Reflections about validation of likelihood estimates by the use of indicators.	14
Theme 5: Quality – Reflections about the quality of expert judgments, likelihood estimates, and indicators.	24

1) We familiarized ourselves with the data by transcribing the recording interviews. From these transcripts, we created non-confidential transcripts.
2) We performed an initial coding of data extracts found in the different data items (non-confidential transcripts). The coding was performed for identifying interesting features of the data. For a number of the data extracts, more than one code was assigned. The coding also resulted in coded data extracts of varying sizes. Some of the data extracts refer to only a part of a sentence, while others refer to a number of sentences.
3) We identified themes based on the initial coding of the data, and assign the different data extracts to these themes. Some data extracts were assigned to more than one theme, while others were not assigned to any theme at all.

C. Results from the thematic analysis

The thematic analysis resulted in 5 themes. These are shown in Table V. Each theme represents a topic that the interviewees talked about. The data extracts assigned to the different themes have been used for discussing the data collected from the case.

The interview data on which the results are based comprise 89 data extracts. These data extracts differ much in size and for some extracts, more than one theme is associated. Their distribution over the different themes is shown in Table V.

APPENDIX D.
THREATS TO VALIDITY

In this appendix we present threats to the validity of our conclusions. The main threat to the validity of the conclusions is, of course, the fact that the investigation was carried out in a single risk analysis case. We should therefore be careful when generalizing the results. It is possible that a different case would have yielded completely different results. However, we do not believe this is the case, for a number of reasons. First, based on experiences from our previously conducted risk analyses we believe that the analysis is representative with respect to target and scope. Second, based on the same experiences we believe that the client experts are representative with respect to experience and background, as well as are their behavior and roles in the analysis.

Based on the above considerations, in the following we present what we believe are the most significant threats to the validity of our conclusions:

- This was the first time the analysis team used indicators in an industrial setting. On the other hand, the analysis team has some experience with indicators from an academic setting. It is therefore possible that the analysis team has more knowledge about indicators than most other analysts that are first and foremost skilled in the area of risk analysis.
- In a risk analysis we most often use experts with different areas of expertise. It is therefore not unthinkable that experts having indicators and security metrics as their field of expertise may participate in risk analyses involving the use of indicators. Based on this we cannot say that the client experts are representative when it comes to understanding the relation between indicator values and likelihood estimates, and when it comes to knowledge about indicators.
- In our particular case the client experts knew quite often who to ask when they needed data. This does not need to be the case in other organizations. It may also be the case that the client organization addressed in this paper is different from other organizations when it comes to logging of risk and security related information. Based on this we cannot say that the client organization is representative with respect to the amount of risk and security related information logged and with respect to having an overview of this data, i.e. knowing what kind of data that is logged and knowing whose in charge for the different data.
- This was the first time the analysis team documented the experiences of the participating experts by the use of semi-structured interviews. Since the analysts are not skilled in interviewing, it is possible that a more professional interviewer would have asked different questions and could perhaps have extracted more interesting information from the interviewees, where as this information could have lead to a different understanding and interpretation of the data collected during the analysis. However, we do not believe that this would have resulted in big differences, since the data collected during the analysis can mainly speak for itself.

Technology for a better society

www.sintef.no