An adequate method to quantify probabilities

Case studies in the context of early reliability prediction using Bayesian networks restricted to subjective data

By J.V. Visée

Student identity number 0602050

in partial fulfilment of the requirements for the degree of

Master of Science
in Innovation Management

Supervisors:
Dr. Ir. P.J.M. Sonnemans, TUE, ID/BPD
Prof. Dr. M.J. Newby, TUE, OPAC

Company supervisor:
Dr. G. Stollman, Philips Medical Systems, BU CV
Subject headings: product creation process, reliability, Bayesian networks, probability, node probability table, experts, opinions, probability, elicitation, quantitative, subjective
Abstract

Use is made of Bayesian networks to investigate early reliability prediction during the product creation process phases at Philips Healthcare. Two case studies have been performed. The first case study investigates which probability elicitation method is the most adequate for quantifying the probabilities for the Bayesian networks, according to experts in the manufacturing and installation phase. The resulting probability elicitation method is checked in the second case study for its generalization over all product creation process phases and consistency. The internal consistency (in terms of accuracy) is checked by using the obtained probability estimates as reference point, due to the lack of historical data. The direct probability elicitation method with table formats is accepted by experts.
Preface

This master thesis finishes eight years of study. Although the Aeronautical Engineering degree has a different focus, this Master thesis is certainly a valuable addition.

I want to thank all those who made this study possible, especially thanks to the supervisors, Peter Sonnemans, Martin Newby, and Ph. D. student Maurits Houben for guiding me throughout the Master thesis project. Providing feedback, critical comments, and iterations of the chapters have helped me to keep me on track.

In addition, I want to thank my supervisor at Philips Healthcare, Guillaume Stollman and all the participating experts for making the research possible. Especially critical feedback and involvement has been appreciated.

During the Master Innovation management I have performed many projects with a dedicated group of companions. Hereby I want to thank them for four years of close cooperation and other social activities.

Finally, I want to thank my family and (future) family in law for interest, warm support and motivation at the times when writing the Master thesis was the most challenging. My parents always backed me up in all situations during my study.

Special thanks to my fiancé, Annelies Kraaijeveld, for her continuous support, endless patience, and encouragements when most needed.

John Visée
Utrecht, September 2009
Management summary

Preconditions
This Master thesis project is an investigation within the Ph. D. project “Early reliability prediction during the PCP” (Houben, 2010). The business environment of this Master project is Philips Healthcare, business unit Cardio Vascular X-ray. This leads to the following preconditions.

The choice within the PhD-project has been made to use Bayesian Networks (BN). For quantifying nodes of the BN, many probability elicitation (PE) methods are available. PE methods are designed to obtain subjective data from experts.

Research problem
It is required to have a PE method that can quantify a BN on the basis of subjective data only. All the current PE methods make use of subjective and objective data and do not describe how they are used. The problem to be tackled is: which PE method to use in the given research environment, in the absence of objective data and how to check the quality of the resultant PE method?

Main research question
The main research question for this Master thesis is the following:

“Which probability elicitation method can be used in this research environment to quantify the node probability tables on the basis of experts’ opinions only?”

Three separate research questions are formulated to investigate the main research question.
1. Which probability elicitation method is found most adequate to use in this research environment, according to experts?
2. Can the resulting probability elicitation method be generalized?
3. Is the resulting probability elicitation method internally consistent, consistent over time and mutually consistent?

Research approach
The Master thesis project is divided into four sequential parts.
1. The first part contains the preparations for both case studies. It starts with the development of the theory. Research strategies are determined and how two case studies are conducted is explained. Finally, preparations are taken for the case studies by explaining what kind of problems and issues have to be taken into account when conducting the case studies.
2. The second part contains the first case study (obtaining an answer to the first research sub question). A pre-selection of the PE methods that are going to be tested is made. A product creation process (PCP) phase is selected in which these PE methods are tested. This is followed by designing a PE document that will be used for the collection of the data for the first case study. The data collection, analyses and findings are conducted and described. The resulted PE method of this first case study serves as input for the second case study.
3. The third part contains the second case study (obtaining an answer to the second and third research sub questions). The gained experience of the first case study and literature will serve as input for designing the PE process used during the second case study. The second case study is conducted and described.
4. Finally, conclusions are drawn and recommendations done.
Answers to the research questions

Answer to the main research question: Direct PE method with a table format is accepted by the experts. This can be concluded from both case study 1 and case study 2.

Answer to the first research sub question: Direct PE methods with table formats are the most adequate to use in this research environment, according to experts. From the literature study can be concluded that direct PE methods are the most adequate in the given research environment. Case study 1 concludes that Direct PE methods with table formats are the most adequate to use.

Answer to the second research sub question: Direct PE methods with table formats can be generalized. With case study 2 the direct PE method is tested for all three PCP phases. 12 out of 13 experts were able to cope with both the PE process as well as the PE method. The table format is accepted over all three PCP phases. As a result, the table format of the direct PE method can be generalized.

Answer to the third research sub question: The conclusion from case study 2 is that no proof has been obtained for the overall consistency of direct PE methods with table formats (method 3).

- Internal consistency: By making use of evaluation through the use of bar charts, the experts are found to be internally consistent (in terms of accuracy) with the probabilities assessed. Process feedback is used to evaluate the ranking of conditional probabilities. Although a couple of experts indicated some changes in ranking, the majority agree with the probabilities assessed by making use of the probability elicitation process.
- Consistency over time: Case study 2 contains no research on this issue. The complicating factors of stability in the answers that play a role with the repeatability and reproducibility of elicitations is the reason. Instead the elicitation has been performed with many experts to gather a data set which can be used as reference point for future research.
- Mutual consistency: Case study 2 shows that there are large differences between the experts over all three phases of the PCP. Within one child node both mutually consistent as well as mutually inconsistent answers can be found.

Recommendations

The most important recommendations of this research are related to the consistency:

- The internal consistency (in terms of precision) could be investigated.
- The consistency over time could be investigated.
- The mutual consistency through use of consensus, making use of the individual input of each expert, could be investigated.
# Table of contents

Abstract ........................................................................................................................................ V
Preface ......................................................................................................................................... VI
Management summary ........................................................................................................... VII
Table of contents .................................................................................................................. IX
List of abbreviations ............................................................................................................... X

1 Introduction .......................................................................................................................... 1
  1.1 Introduction to the problem definition ............................................................................. 1
  1.2 Preconditions .................................................................................................................. 2
  1.3 Problem definition ......................................................................................................... 2
  1.4 Outline of thesis ............................................................................................................. 2

2 Theoretical framework .......................................................................................................... 3
  2.1 Bayesian networks ......................................................................................................... 3
  2.2 Building Bayesian networks .......................................................................................... 4
  2.3 Probability elicitation methods ..................................................................................... 5

3 Research outline .................................................................................................................. 11
  3.1 Research environment ................................................................................................. 11
  3.2 Research questions ....................................................................................................... 13
  3.3 Research design ........................................................................................................... 14
  3.4 The quality criteria ...................................................................................................... 18
  3.5 Thesis outline ............................................................................................................... 19

4 Issues in data gathering ....................................................................................................... 22
  4.1 General considerations: heuristics and biases ............................................................... 22
  4.2 Before elicitation .......................................................................................................... 24
  4.3 During elicitation .......................................................................................................... 27
  4.4 After elicitation ............................................................................................................ 28
  4.5 An overview of the issues ............................................................................................. 29
  4.6 Case study preparations ............................................................................................... 30

5 Probability elicitation method selection .............................................................................. 33
  5.1 Pre-selection of the probability elicitation method ....................................................... 33
  5.2 Selection of the Product creation process phase ......................................................... 36
  5.3 The what’s and why’s of case study 1 ......................................................................... 36
  5.4 Case study 1: probability elicitation document ............................................................ 37
  5.5 Case study 1: data gathering ......................................................................................... 38
  5.6 Case study 1: analysis of answers and observations .................................................... 40
  5.7 Case study 1: findings .................................................................................................. 43

6 Implementation of the selected probability elicitation method ............................................ 46
  6.1 The what’s and why’s of case study 2 ......................................................................... 46
  6.2 Case study 2: probability elicitation process ............................................................... 47
  6.3 Case study 2: data gathering ......................................................................................... 49
  6.4 Case study 2: generalization ...................................................................................... 50
  6.5 Case study 2: consistency ........................................................................................... 50
List of abbreviations

BN  Bayesian Networks
BU  Business Unit
CD  Concept and Definition
CV  Cardio/Vascular
DI  Disposal
DD  Design and Development
FMEA  Failure Mode Effect Analysis
MI  Manufacturing and Installation
NPT  Node Probability Tables
OM  Operation and Maintenance
PCP  Product Creation Process
PE  Probability Elicitation
PH  Philips Healthcare
1 Introduction

This chapter starts with an introduction to the problem definition (1.1). Because this Master thesis is conducted under specific circumstances, some preconditions have to be stated (1.2). Following the stated preconditions, the problem definition is given (1.3). Subsequently, the outline of this thesis (1.4) will be provided.

1.1 Introduction to the problem definition

Before the problem definition can be given, the context of the problem has to be explained. The research of this thesis is in the context of early reliability prediction. The vision on reliability has changed from component reliability to product reliability (Brombacher and Sander, 1999). Customers’ expectations are rising, resulting in an increase of required warranty time. In order to increase the warranty time one needs higher product reliability. Reliability is defined as “the ability of a system or component to perform its required functions under stated conditions for a specific period of time” (Hardy et al., 2007). So, it is not known how reliable a product is until it is at the end of the specific period of time. However, estimations of reliability are needed during all the product lifecycle phases.

Bedford et al. (2006) defines the product life cycle in five phases.
- Concept and definition (CD).
- Design and development (DD).
- Manufacturing and installation (MI).
- Operation and maintenance (OM).
- Disposal (DI).

The first three product lifecycle phases define the product creation process (PCP) in which the product is made and installed at the customer. The last two phases are influenced by the end user.

A product is unreliable, if it does not perform its required functions within the specific period of time. The later faults, that influence the reliability of a product, are discovered, the higher the expenses are to resolve these (Yates & Beamen, 1995). Because reliability is moving towards the responsibility of the company selling the products (Sander & Brombacher, 2000), companies are interested in how they can influence the product reliability as soon as possible. Changes at the beginning of the PCP are relatively low on costs and changes can be made more easily compared to later in the product lifecycle (Yates & Beamen, 1995). As a result, companies focus on how to influence reliability by changes made during the PCP. However, the effects of changes during the PCP are subject to uncertainty.

To be able to gain insight in the reliability during the PCP, the types of data that are available during the PCP have to be investigated. Garthwaite et al. (2005) identified three types of data: literature, historical data and subjective judgement. Literature contains abundant data about reliability, but little can be used as a reference to quantify probabilities. The data that is available lack the information on how the data has been retrieved, or cannot be used (v. d. Gaag et al., 2002). In case of new product development little historical data is available. Many companies even lack the resources and/or knowledge to make optimal use of gathered historical data. In this case, the only data source is subjective expert judgement. This stresses the importance of the use of experts for a prediction on reliability.
Expert knowledge can be made available in the form of subjective data that can be obtained by using elicitations. However, which method can be used to cope with subjective data under uncertainty? The next section will present a method and which stage of that method is used.

1.2 Preconditions

This Master thesis project is an investigation within the Ph. D. project “Early reliability prediction during the PCP” (Houben, 2010). The business environment of this Master project is Philips Healthcare, business unit Cardio Vascular X-ray, and is further described in chapter 3.1. This leads to the following preconditions.

The choice within the PhD-project has been made to use Bayesian Networks (BN), because BN’s are able to reason under uncertainty (Neil et al., 2000; Sigurdsson, 2001; Nadkarni & Shenoy, 2004; Khodakarami et al., 2007). BN’s can handle both objective and subjective data (Fenton et al., 2007). And by making use of probabilities, BN’s are able to model the uncertainties present in both objective and subjective data (O’Hagen & Oakley, 2000). Probabilities can also give an indication of the amount of uncertainty.

Generally, building a BN consists of three stages.

- Structuring the network.
- Quantifying the probabilities.
- Interpreting the results.

The first stage, structuring the network, has been done by prior research that will be published in the near future (Houben, 2010). This thesis will focus on the second stage, namely quantifying the probabilities.

For the quantitative stage of the BN, many probability elicitation (PE) methods are available. PE methods are designed to obtain subjective data from experts. It is not clear whether the nodes in the BN can be quantified without objective data as a reference and how this should be done. Now, all the components are present to define the problem.

1.3 Problem definition

It is required to have a PE method that can quantify a BN on the basis of subjective data only. All the current PE methods make use of subjective and objective data and do not describe how they are used. The problem to be tackled is: which PE method to use in the given research environment (chapter 3.1), in the absence of objective data and how to check the quality of resulted PE method?

1.4 Outline of thesis

The problem defined will be addressed by the following steps. First of all the theoretical framework of this thesis will be presented (chapter 2). Subsequently, the research outline is given (chapter 3). Chapter 4 deals with preparations regarding the case studies. In chapter 5 a first case study is conducted. This case study investigates which PE method is found adequate by experts. The actual quantification of the required probabilities for the BN by the selected PE method is done in the following chapter (chapter 6). The final chapter (chapter 7) provides answers to the research questions and gives recommendations for future research. The next chapter presents the theoretical framework.
2 Theoretical framework

From the introduction, one of the preconditions is to work with a BN. First of all the BN will be discussed in more detail (2.1). Secondly, a closer look at the various steps within the three stages (i.e. structuring the network, quantifying the probabilities and interpreting the results) of a BN follows (2.2). The problem definition is focussed on the second stage, quantifying the probabilities. Three types of PE methods are identified and explained (2.3).

2.1 Bayesian networks

A BN is a network in which the identified variables are ordered according to cause-effect relationships. Each identified variable is represented by a node in the network. The arrows between the nodes indicate the direction from cause to effect. Figure 1 shows an abstract BN as example. The causal nodes are so called ‘parent-nodes’ (nodes A and B), while the node on which it has effect is a ‘child-node’ (node C). The two dotted arrows (at node A) show that each parent node can be subject to other nodes. In that situation the other nodes are the parent-nodes of node ‘A’. Each node can be in several states, i.e. the possible values or value-ranges of that particular variable (e.g. states A1, A2, B1, B2, C1 and C2). Each of these states is expressed by the probability that such a state occurs. To understand the relationships between the nodes, and states of each node, BN’s make use of probability calculus and Bayes’ theorem (Woolridge, 2003).

![Figure 1. From Bayesian network to node probability table (NPT).](image)

The states for each node are expressed as probabilities of occurrence, often expressed in so-called node probability tables (NPT), as given in Figure 1, e.g. the probability of node C to be in state C1, or C2 (given the states A1 and B1) is 80 % and 20 %, respectively. The probabilities (80, 20) represent the conditional probabilities for the hypothetical situation (A1, B1). The law of probability states that the sum of the probabilities is 100 percent (state B1 + state B2 = 100%). The child node is affected by the parent nodes (A and B). This is represented in the NPT, where each row indicates the possible combination of node ‘A’ and node ‘B’. Column ‘C’ represents the two different states (C1 and C2) for node ‘C’ which should horizontally sum up to 100% to comply with the laws of probability. Each probability is based on the reasoning of situations that already took place. Woolridge (2003) states that estimated unobserved conditional probabilities (i.e. conditional probabilities that are based on hypothetical situations which did not take place) are strengthened because they are related to hypothetical situations that have taken place. All the conditional probabilities for one node together form the ‘node probability table’.

Because the sizes of the NPT’s depends on the number of parents and the number of states of each parent, the size of the NPT’s increases exponentially with the number of involved
parents and their states (Druzdzel & v.d. Gaag, 2000; Wang, 2004). For this reason, the number of parents and states should be kept to a minimum (Wang, 2004). On the other hand, limiting the size of the network prevents detailed modelling. Although the use of BN is generally accepted, it is difficult to obtain the probabilities for the NPT’s (Fenton et al., 2006; v.d. Gaag et al., 2002). The steps for building a BN are described in the next section.

### 2.2 Building Bayesian networks

Many articles have been written on how to build a BN (e.g.: Spetzler & Staël von Holstein, 1975; Pearl, 1988; Apostolakis, 1990; Sigurdsson, 2001; v.d. Gaag et al., 2002; Nadkarni & Shenoy, 2004). The number of stages to build a probabilistic network for a specific domain varies between two and four stages. Because articles focus on particular areas, more steps are taken in those stages. Summarizing, the following three stages are used in all the papers (appendix A).

1. Structuring the network.
2. Quantifying the probabilities.
3. Interpreting the results.

Each of these three stages consists of steps that have to be taken to construct a BN. Figure 2 presents the seven steps as described by Sigurdsson (2001).

The first stage, **structuring the network**, contains three steps: identify the variables, identify the network structure and express as statistical variables. Qualitative elicitation is used to identify the relevant variables (step 1). The causality of the relevant variables is used to construct a BN (step 2). Each variable should be expressed as a discreet or a continuous variable to capture the statistical relationships between variables probabilistically (step 3). Steps 2 and 3 of the process are considered iterative, because iterations are required to refine the initial BN before the model can be considered robust.

Robust has not been defined explicitly in literature. However, robust can be seen as the state in which the model is unambiguous and all the parties involved can identify their opinions with the model. This is a requirement before continuing to the next stage (v.d. Gaag et al, 2002).

During the second stage, **quantifying the probabilities**, the conditional probabilities are specified (step 4). Conditional probabilities are defined by: “the probability that an event takes place, given that (an) other event(s) already took place”. The quantification may be done on the basis of data and/or experts’ opinions. The latter is considered the most difficult, because experts find it hard to express their opinions in terms of probabilities (Sigurdsson, 2001). In this thesis only experts’ opinions are available.

Figure 2. Flowchart for building a BN (Sigurdsson, 2001).
In the final stage, interpreting the results, the evidence is filled in throughout the network and results are interpreted (steps 5 to 7). The evidence in this project is represented by the experts’ knowledge about both the probabilities and conditional probabilities of the states of certain variables. By feeding the BN with these probabilities, an insight is created in how the nodes (variable) and the states of each node influence the reliability during each stage of the PCP, according to the each expert.

This Master project focuses on the second stage, quantifying the probabilities through probability elicitation (PE). Therefore, the next section will present three types of common PE methods that are described in literature.

### 2.3 Probability elicitation methods

A BN makes use of probabilities. In this thesis only subjective data (expert’s opinions) is available, which has to be obtained from experts by making use of quantitative PE. This can be done by making use of PE methods. The PE methods discussed in literature can be divided into three different types.

- Direct methods.
- Indirect methods.
- Parametric methods.

This section will shortly introduce these PE methods.

Several methodologies are described in literature that help the experts transfer their opinions into probabilities. Three large distinctions can be identified between the identified PE methods. Spetzler & Staël von Holstein (1975), Cooke (1991) and van der Gaag et al. (2002) describe two PE methods; direct and indirect. The direct PE methods provide numbers to the asked questions. With the indirect PE method choices have to be made between alternatives or bets, from which their probabilities are inferred. Cooke (1991) adds a third PE method, which is the parametric PE method. In some applications the nature of the quantities suggests a particular class of probability distributions that needs to be elicited. The parametric PE method suggests a way to cope with such a situation.

The best applicable method depends on the overall goal the researcher has with the BN (Renooij, 2001b). Each PE method is discussed in one of the following three sections. The final section (2.3.4) presents an overview of the advantages and disadvantages as described in literature.

#### 2.3.1 Direct probability elicitation methods

With direct elicitation, the respondent is asked direct questions about the relation between a parent- and child-node, which have to be answered by either values or probabilities (Spetzler & Staël von Holstein, 1975). Although literature differs in opinion whether direct elicitation leads to accurate probabilities, problems have been raised, refuted and renewed (Renooij, 2001b; v.d. Gaag et al, 2002; Spetzler & Staël von Holstein, 1975, Druzdzel & v.d. Gaag, 2000). The most common direct PE methods mentioned in literature will be introduced, i.e. numerical probability scales, verbal scales and a combination of both.

**Numerical probability scale**

The probability response can be given as percentage, absolute number or as fraction (one out of five). Fractions have been found very practical when discussing probabilities close to zero (Spetzler & Staël von Holstein, 1975). A probability scale (figure 3) can support the
respondent to think in terms of visual proportions instead of precise numbers. The probability scale can be presented horizontally and vertically, showing anchors on a fixed or variable interval. By measuring the distance between ‘0’ and the line or cross (no circles, because these are not accurate enough) the subjective probability estimated by the expert can be determined. However, the ‘centering effect’ and ‘spacing effect’ are possible biases that can occur and negatively influence the results (Renooij, 2001a). Numerical anchors have the advantage of being precise, in a fixed-ranked order and enable calculations. However verbal scales can also be used.

![Probability scale with a fixed interval and two indicators: the circle which is not accurate enough and the line as an appropriate indicator.](image)

**Verbal scale**

Instead of a scale which is numerical, a verbal scale can also be used (figure 4) (Renooij, 2001a). The anchors on a verbal scale are dependent on the words chosen. Opposite to the numerical anchors, verbal anchors are less precise and unsuitable to be used for calculations. However, they are able to convey the vagueness of opinions of the experts. Hereby verbal anchors give the respondent the opportunity to give a more natural answer, expressing their uncertain opinions in an understandable way, which makes communication easier (Spetzler & Staël von Holstein, 1975; v.d. Gaag et al, 2002; Renooij, 2001a; Witteman & Renooij, 2003). Copying words without a research on which words are accepted in the research field may lead to unknown biases, although the words in figure 4 have already been used with success in several research fields (Renooij, 2001a).

![Verbal scale with a variable interval.](image)

**Combination of numerical and verbal scale**

Comparing the numerical and verbal anchors, it cannot be said that either one results in a better communication medium for probabilities. After trying several methods, Druzdzel and van der Gaag (2000) discovered how much time the quantification process was taking. As response, Renooij (2001a) has determined a methodology to determine probabilities, in an easy and time constrained manner. Because the focus was on the ease and time, the accuracy of the results was not checked. Common verbal terms were aligned and placed next to the numerical scale (figure 5). Because respondents in different fields should be able to understand the stated question, van der Gaag et al. (2002) have added the fragments of text, to the verbal-numerical scale, in terms of likelihood instead of frequency. The experts elicited in the research of van der Gaag et al. (2002) found this method more easy to use than any of the preceding methods that had been used. Although the preparation time for the elicitors has increased, the ease with which probabilities are elicited has also increased considerably. A third method is therefore introduced that combines the numerical anchors, with variable intervals, with verbal words as guidance, called a verbal-numerical scale (Witteman & Renooij, 2003). Witteman and Renooij’s (2003) experience shows that this method is a proper method to retrieve a large number of coarse probabilities in a limited amount of time.
Next to the direct PE methods, where numerical values are directly asked, also indirect PE methods are described in literature. Indirect PE methods make use of graphical representations of the required probabilities and/or preferences of the experts for alternative states. This is further explained in the next section.

2.3.2 Indirect probability elicitation methods

With indirect elicitation techniques the respondent is asked questions about alternatives or bets. The elicitor then infers the probabilities from the given answers. This prevents the need to make numbers and/or verbs explicit by the elicitor, or educate the expert on probability assessment. This methodology is very effective with experts who lack a clear intuition about numerical probabilities (v.d. Gaag et al., 2002).

Probability wheel

The probability wheel (figure 6) is considered to be the most popular tool for graphically encoding indirect responses from subjects (Spetzler & Staël von Holstein, 1975; Wang, 2004). The wheel, in this example shown in figure 6, is represented by a disk with two adjustable sections, each with its own colour (purple and blue), and an arrow. Although literature (a.o. Renooij, 2001b; Wang, 2004) mainly uses two sections, it can be imagined that it has several sections (three or four). After spinning the wheel by the elicitor, the arrow will stop in one of the two coloured sections. The relative size of the two sections can be changed and thereby also changing the probability of the arrow pointing to one of the two sections when stopped. The expert is now to adjust the two sections in the wheel until he finds that the sections represent the corresponding probabilities according to his feeling. His feeling is about the likelihood of events, represented by the two colours. The coloured areas represent the probabilities of the corresponding events (Spetzler & Staël von Holstein, 1975; Renooij, 2001b; Wang, 2004).

Pair-wise comparison

Another method, originally developed for utility elicitation, is the analytical hierarchy process, or pair-wise comparison, discussed by Renooij (2001b). Pair-wise comparison compares two events (i.e. states of variables) at a time. For example, is the likelihood of event A higher than that of B? To obtain the utility of the pair, event A and B, a score is assigned from a scoring table (e.g. event A is absolute more likely than event B, results in a 9). If this should be done for all the events in a BN, the number of comparisons will explode with the size of the BN. In some cases, events that are compared will be so different that they cannot be judged.
As with all the indirect methods mentioned, the respondent should be able to imagine the events, even when they seem rare or unethical. This also applies to the probability wheel, and pair-wise comparison (Cooke & Goossens, 2008). Furthermore, another possibility is that probabilities follow a distribution. How to act on such a situation is discussed in the next section.

2.3.3 Parametric probability elicitation method
In contrast to the previous PE methods, parametric elicitation is based on the idea that probabilities can be defined by a distribution. Instead of asking for a single (conditional) probability, experts indicate that the probability follow a continuous distribution (Cox, 2006).

Continuous distributions
Continuous distributions can appear in multiple shapes. Two steps are proposed by Cooke in case of a lognormal distribution (1990):

1. Elicit a best estimate for the failure frequency in question.
2. Elicit the degree of uncertainty about the given value.

The first step, elicit the best estimate, can be done by asking for the median estimate. The second step, the degree of uncertainty is derived by asking the expert to state the upper, 95%, and lower, 5%, confidence bounds of the estimated median.

A more precise 'bell-shaped' distribution is acquired if the mean, derived from expert opinions, is interpreted as the median of the distribution (Apostolakis, 1990). The error between the mean and median can partly be ascribed to the anchor bias that experts are influenced by.

Although the shape of the probability is relevant for the accuracy of the model, retrieving the type of distribution and the parameters will take more time, effort and understanding about probabilities from the expert. The outcomes can be more, but also less accurate, depending on the correctness of the assumptions. These methods are therefore considered less robust (Freedman, 2000).

Three different types of PE methods have been presented: direct-, indirect and parametric PE methods. The next section presents an overview of the advantages and disadvantages found in literature for each PE method.

2.3.4 An overview
Section 2.3 has shown that there are three different types of PE methods to quantify the BN: direct-, indirect- and parametric PE methods. Direct PE methods elicit numerical values. This can be done by making use of different types of scales, numerical and/or verbal scales, to support the expert during the elicitation. Indirect PE methods are based on the preference experts have, possibly supported by graphical tools. Parametric PE methods are used if the probabilities follow a continuous distribution. A distribution is elicited by eliciting particular characteristics of the type of distribution, a possible expedient can be different graphs, or by eliciting points on the distribution. Visée (2008) has addressed advantages and disadvantages of the three types of PE methods mentioned in this chapter. An overview is presented on the next page. The first column presents the types of PE methods that correspond with 2.3.1, 2.3.2 and 2.3.3, respectively. The second column shows the specific PE methods. In the third and fourth column the advantages (positive) and disadvantages (negative) of the specific method are addressed. The final column shows where the mentioned advantages and/or disadvantage can be found in literature. For example, the direct elicitation method (first column) "verbal scales" has as advantage that these can be used to convey the vagueness of
the experts (third column), however the assessed probabilities will be less precise (fourth column). This advantage and disadvantage are discussed by Renooij (2001a) (fifth column).

This chapter dealt with the theoretical framework in relation to the research problem as defined in the introduction, i.e. the elicitation of experts to quantify the probabilities required for the BN. The next chapter will present the research outline.
### Overview of probability elicitation (PE) methods

<table>
<thead>
<tr>
<th>Types of PE methods</th>
<th>PE methods</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Numerical scales</td>
<td>Easiest to implement, Supported by graphs a higher accuracy, Fractions easy to use close to zero, easy to understand, fast method for elicitation, Use for large probabilities, Use of rough, centering effect, spacing effect bias, scale intervals.</td>
<td>Probabilities not directly observable, least reliable, when respondents are not familiar.</td>
</tr>
<tr>
<td></td>
<td>Direct</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Verbal scales</td>
<td>Convey the vagueness of the experts, more natural answer, understandable way, Current used words accepted over research fields</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combination</td>
<td>Labeled nodes easier to identify by experts, Spread of filling in for non mathematicians, self-assured assessments, limited differences between three tools/scales</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indirect</td>
<td>Most popular for graphical encoding, Based on utility, Visualize the probabilities, Different formats, Based on utility</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Probability Wheel</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pair-wise comparison</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parametric</td>
<td>Degree of uncertainty through median and variance, Mean interpreted as Median</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distribution</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### References

- Witteman & Renooij, 2003
- Numerical scales
- Renooij, 2001b
- Spetzler & Steal von Holstein, 1975
- Wang, 2004
- Freedman, 2000
- vd Gaag et al., 2002
- Fenton et al., 2006
- Drudzel & vd Gaag, 2000
- Cooke, 1991
- Apostolakis (1990)
- Apostolakis (1990)
- Cooke, 1991
- Drudzel & vd Gaag, 2000
- Fenton et al., 2006
- Renooij, 2001a
- Renooij, 2003
- Rentfrow, 2016
- Rentfrow, 2003
- Soddy & Steal von Holsten, 1975
3 Research outline

The research outline starts with a description of the research environment (3.1). Secondly the research questions are defined (3.2). Subsequently, the research design to investigate the formulated research questions is presented (3.3). This is followed by the quality criteria (3.4) with which the research design is checked. Finally, the thesis outline is presented (3.5).

3.1 Research environment

The research environment of this Master thesis project is confined both by the business environment in which the research takes place, and also by preconditions as mentioned in the introduction. First the business environment is presented followed by the specific BN’s used in this research.

3.1.1 Business environment

This Master thesis project is conducted at Philips Healthcare (PH), which is part of Royal Philips electronics. PH makes medical scanning devices, accessories and provides services to hospitals worldwide and is amongst others focussed on the reliability of the products they deliver. This can be noticed in their quality policy. “We (PH) satisfy their (the customers) needs by delivering effective, safe and reliable products through our business processes”.

Royal Philips electronics was founded in the year 1891, as a lamp factory. The mission statement is defined as: "Improve the quality of people’s lives through the timely introduction of meaningful innovations". Royal Philips electronics is sub-divided into three sectors; ‘Consumer Lifestyle’, ‘Healthcare’ and ‘Lighting’. Philips Healthcare is the product division that is concerned with capital goods. The advantage of capital goods over fast consumer goods is that the reliability of the product is becoming more important in the after sales period, in relation to consumer goods. This is, among other things, due to the large difference in expected lifetime of the product.

Figure 7 shows the organisational chart for the sector ‘Healthcare’. The research is conducted at the department ‘CV Innovation’. This department is part of the business line ‘Imaging Systems’, and business unit Cardio/Vascular X-Ray (CV). The aim of PH is “to improve patient outcomes while lowering the burden on the healthcare system”. One of the burdens can be considered the unreliability of products.

Figure 7. The business structure of Philips Healthcare at the end of 2008.

1 pww.philips.com, last visited on February 2009
CV and all the departments are located at Best in the Netherlands. CV is the location where the Allura Xper product line is still an in-house activity. Appendix B gives a limited description of the Allura Xper product line, with a focus on the latest developed subsystem, the AD-7 table. Figure 8 gives an illustration of what the Allura Xper product line can look like. Because the Allura Xper product is a combination of many systems each product is customized (i.e., the customer chooses which systems they want (monitor, patient support (AD-7 in this case), Lateral- and/or frontal stand, etc, etc)). The combination of all the systems defines the Allura Xper product. Further explanation can be found in appendix B. The complete PCP of the Allura Xper is controlled from Best. The PCP used in this Master thesis project can be seen as a representation of the PCP used at CV.

![Figure 8. Allura Xper product.](image)

The opportunity to investigate a PCP in a company, where production is still in-house, can be seen as a unique experience. Use can be made of experts with a large variety of expertises in different areas of the PCP. Experts from design, manufacturing, marketing, sales and service are all within the range of the project. This research is limited to the experts of the business unit CV. Now the business environment has been explained, the BN’s already derived from this business environment in the PhD-project, are discussed in the next section.

### 3.1.2 Bayesian networks

As discussed in the introduction, this Master thesis project is confined by two preconditions coming from the PhD-project (i.e. the use of BN’s and focus on quantifying the probabilities). The first stage of building a BN has already been done. Because the second stage (quantifying the probabilities) is dependent on the robustness of the BN’s, the realization of the BN’s is shortly discussed in this section.

Although the product lifecycle is defined in five phases (CD, DD, MI, OM and DI), the business unit CV is only interested in the first three phases (CD, DD and MI) that define the PCP. The reason for this is that CV is able to proactively steer the reliability performance of the products they are making mainly during the PCP. Furthermore, during the last two phases
(OM and DI) the reliability of the product is also influenced by the end user, which is out of control of CV. The effect of each PCP phase on the reliability is the central topic of each BN. The qualitative elicitation has resulted in more than 300 variables from approximately 20 experts. A selection from all the identified variables has been made, followed by determining the cause-effect relationships between the variables for each PCP phase. As a result, for each PCP phase a BN has been constructed (appendix C). Figure 9 shows the BN for the MI phase, which is taken as an example. The size of each BN has been minimized, but detailed enough to investigate the effects of controllable variables (e.g. ‘experience with respect to installation’, ‘quality of testing process’ and ‘quality of the production assembly-process’), which are all the parent nodes in the figure. The parent nodes influence the child nodes, with in the centre of the figure the ‘Reliability effect during the MI phase’.

![Bayesian network of the manufacturing and installation (MI) phase](image)

Because the aim of the Ph. D. project is to gain insight in early reliability prediction, the constructed BN’s are predictive instead of reactive. Predictive implies that the models are created early during the PCP, where only experts’ opinions are available. No feedback loops are taken into account, as would be appropriate in case the BN’s are reactive. With a reactive BN the experts’ opinions are influenced by feedback from tests during the PCP and/or field data, gaining new insights in the relationships between nodes in the BN’s. As a result, the probabilities for the BN’s should be adjusted during the development of products for a reactive BN. Because the BN’s under investigation are predictive, they are quantified once.

This Master thesis project is bound to the three constructed BN’s and the controllable variables used. Now the research environment has been discussed, the research questions are defined in the next section.

### 3.2 Research questions

In the introduction the following problem definition has been formulated: It is required to have a PE method that can quantify a BN on the basis of subjective data only. However, all the current PE methods make use of subjective and objective data and do not describe how they are used. How to tackle the problem of which PE method to use in the research environment under investigation (chapter 3.1) and how to check the quality of the resulted PE method?
In order to find the PE method that suits the research environment under investigation, research is needed. In this thesis the following main research question will be investigated:

"Which probability elicitation method can be used in this research environment to quantify the node probability tables on the basis of experts’ opinions only?"

Because the main research question cannot be investigated at once, it is divided into three separate research sub questions. The first research sub question focuses on the theoretical research, where the applicability of the theoretical framework is examined within PH. The second and third research questions focus on the practical applicability.

The research sub questions are as follows:

1. Which probability elicitation method is found most adequate to use in this research environment, according to experts?

The resulted PE method has no value if it is not checked on its performance by checking some quality criteria. One of these criteria is how generally accepted the resulted PE method is. The broader the scope at which the PE method is accepted, the more the PE method can be generalized. From this the second research sub question follows.

2. Can the resulting probability elicitation method be generalized?

A second quality criteria on which the PE method can be tested, is the consistency of the assessed probabilities from the resulted PE method. Because literature was always able to check the consistency by means of objective data (a.o. Cooke, 1991; Renooij, 2001a), this research is dependent on subjective data only. As a result, checking for consistency is of importance to prevent that assessed probabilities are a random indication of experts, but that the assessed probabilities are made on the basis of experts’ experience. The consistency can be checked on three levels, i.e. checking if experts are consistent with their own assessed probabilities (internally consistent), checking if experts are consistent with themselves over time (consistency over time) and by checking if experts agree or disagree with one another (mutual consistency). This results in the third research sub question.

3. Is the resulting probability elicitation method internally consistent, consistent over time and mutually consistent?

Expected deliverables of this research are:

- A PE method that is most adequate according to the experts, together with its performance on some quality criteria that will be formulated in chapters 5 and 6.
- NPT’s for each of the three BN’s.

This section has formulated one main research question, with a number of research sub questions. The next section will present the research approach that is used to answer the research questions.

3.3 Research design

In the previous section the main research question has been divided into three research sub questions. The first research sub question investigates which PE method is found the most adequate by experts. This will result in one adequate PE method.
Research sub questions two and three will investigate to which extent the selected PE method can be implemented and will elaborate on the quality criteria of only the resulted PE method. Two separate research strategies are determined. The first research strategy is for the first research sub question. The second research strategy is for the second and third research sub questions. Each research strategy is followed by how to conduct each of the research strategies.

For the selection of a research strategy, Verschuren and Doorewaard (1995) have proposed a procedure to follow. This contains the following five steps for the selection of a research strategy.

1. Is the research focussed on depth or breath?
2. Is the research qualitative or quantitative?
3. Is the research based on literature or empirical evidence?
4. Which of the following five strategies fits the previous three questions best?
   - Survey
   - Experiment
   - Case-study
   - Literature based approach
   - Desk research
5. Choose a specific type of research strategy.

This paragraph will evaluate the steps for the first research sub question (Which PE method is found most adequate to use in this research environment, according to experts?).

1. PE methods are a specific topic within the construction of BN’s. The first research question investigates the adequacy of available PE methods, according to a small number of experts. As a result, the research strategy is more focussed in depth than in breadth.

2. Qualitative research is particularly important when studying persons, groups or organisations, whereas quantitative research is focussed on the amount or number of these qualities (van Aken et al., 2006). Since the number of experts is limited within PH, it is unlikely to gather enough data for a significant quantitative research. As a result, the first research strategy is a qualitative research.

3. Although the first research strategy is based on a theoretical framework, the theory is tested in a business environment. Empirical evidence will point which PE method is found the most adequate by experts.

4. According to the previous three answers, Verschuren and Doorewaard (1995) suggest to perform a case-study. A case study that relies on theoretical concepts to guide the design and data collection is one of the most important strategies (Yin, 2003; van Aken et al., 2006).

5. For the fifth step, several types of research strategies for case studies have been proposed. Verschuren and Doorewaard (1995) make a distinction between single case research and multiple cases research. Yin (2003), on the other hand, describes six types of case-studies: exploratory, descriptive and explanatory research, each of them can be performed with single or multiple cases. This Master thesis project is interested in the applicability of a specific PE method within PH. This can be described as an exploratory research. A case study is especially applicable for exploratory research (Graziano & Raulin, 2004). Whether the research is performed with single or multiple cases is dependent on the unit of analysis (Yin, 2003). The unit of analysis can be defined as “the type of object that is the focus of interest” (van Aken et al., 2006). The number of objects investigated determines if it is a single case or multiple cases
research. The type of object can be based on available literature, previous research, or other units of analysis (Yin, 1984). Although multiple experts will be elicited during this case study, the object under investigation is the most adequate PE method. Because the most adequate PE method can be considered a single type of object for the unit of analysis, this results in a single case exploratory research (Yin, 1994).

The purpose of the first case study is to explore which of the PE methods is found to be the most adequate by the experts at PH, for translating their opinions into probabilities. In order to achieve this purpose, a small multi-method case study (i.e., a case study in which multiple PE methods are tested under similar conditions to be able to make a comparison) will be conducted with a selected number of experts from one phase. Not all PE methods are tested, because many are not applicable in this research environment. A selection of PE methods will be made in section 5.1. Because PH is a large firm with a lot of experts, a selection has to be made on which experts to use. The PCP under investigation consists of three phases (CD, DD and MI) one of these phases and accompanying experts is selected (section 5.2) to conduct the first case study with. A PE process needs to be developed, because experts at PH are unknown with quantifying their opinions by making use of PE methods (section 5.3, 5.4). Data will be gathered on what the experiences are of the experts with the selected PE methods by making use of semi-structured questions, observations and documentation of the answers. Making use of semi-structured questions has as benefit that they gather the richest amount of information (Verschuren and Doorewaard, 1995), because it provides the opportunity to elaborate on questions asked. Thus the appreciation of the experts for a certain PE method defines which method will be chosen as the “most adequate”.

To gain insight on the experience of the experts from several perspectives, the semi-structured questions will be formulated within the following four groups.

- The expert’s experience with estimating probabilities.
- The expert’s appreciation about the methods used.
- The expert’s appreciation about the accuracy of their answers.
- Which knowledge the experts lack to define probabilities.

The obtained answers from the individual elicitations are then analysed by grouping them with the four groups in which the questions have been asked. Each group is then analysed. From the first, third and fourth group (the expert’s experience with estimating probabilities) a general view will be given. The second group (the expert’s feelings about the methods used) has a different approach, because answers from this group will be the closest related to the research question in comparison to the others. The second group will be ranked by interpreting if answers are of a positive or negative nature towards the methods used. A mutual comparison of the rankings will point out which method will be the best accepted. An overall view of the four analysed groups will point out which method is the most adequate to use within the research environment.

Like done at the beginning of the section for the first research sub question, this paragraph will discuss the steps to determine the research strategy for the second (Can the resulting probability elicitation method be generalized?) and third research sub questions (Is the resulting probability elicitation method internally consistent, consistent over time and mutually consistent?).

1. The second and third research sub questions take the research create more depth, by taking the result of the first research question and look at the implementation of the
PE method and some quality criteria within the discussed business environment (section 3.1).

2. The second step is equal to the first research strategy in which not enough significant data can be gathered for a quantitative research strategy. As a consequence this research strategy will also be a qualitative research.

3. The second and third research questions are based upon empirical evidence of the first research question. Through implementation of the resulted PE method from the first research strategy, results will be based on empirical evidence.

4. A case study answers the needs of the first three answers. As a result a case study is proposed by Verschuren and Doorewaard (1995) for the second research strategy.

5. Like with the first case study, this case study is also an exploratory research. To determine if this research is a single or multiple case study research dependents on the unit of analysis (i.e. the type of object that is the focus of interest) (Yin, 2003). The second research strategy investigates the quality of the resulted PE method. This is done by investigating if the resulted PE method can be generalized and if it is consistent (i.e. internally consistent, consistent over time and mutual consistent). To investigate this multiple experts need to be elicited and analysed. To be able to say anything about the quality of the method, the answers of multiple experts need to be investigated. Consequently is each expert a unit of analysis. As a result, the second research strategy is a multiple case exploratory research (Yin, 2003).

The purpose of the second case study is to investigate the quality of the resulted PE method from the first case study. This will be done by investigating two quality criteria. The first criterion is if the resulted PE method can be generalized (research question two). The second criterion is if the assessed probabilities are internally consistent (i.e. do experts agree with the assessed probabilities), consistent over time (i.e. do experts give the same answers over time) and mutual consistent (i.e. are experts agree with one another).

To be able to investigate the two criteria elicitation have to take place with experts. An elicitation is more than a PE method only, but also needs some other elements (e.g. training on probabilities, an instruction on how to use the method, etc.). As a result a PE process (i.e. all the elements needed to be able to elicit an expert) will be created on the basis of literature and gained experience from the first case study (i.e. by applying a PE process in the first case study experts were able to indicate their experience with the PE method, but also with the PE process).

Next to the PCP phase used for the first case study, the other two phases of the PCP are also used for this case study. While the first case study was focussed on gaining insight in which PE method is the most adequate to use in this research environment (3.1), this case study will investigate to which extent other experts agree with the experts used during the first case study. By making use of all the three phases of the PCP (CD, DD and MI) the number of experts participating will approximately triple. Consequently, the numbers of elicitation have to be performed increases significantly, which limited this research by being able to perform only one per expert. In the situation that all the experts are able to quantify their opinions by making use of the PE method, results can be generalized over a wider scope. In case not all experts are able to quantify their opinions by making use of the PE method, an explanation will be sought for.

All the three types of consistency (internal consistency, consistency over time and mutual consistency) can only be determined if probabilities have been obtained. To check the
internal consistency, the obtained probabilities have to be assessed within a small period of
time from the moment the probabilities have been obtained to prevent the experts being
influenced by external factors (Kynn, 2008). External factors, i.e. biases, can influence the
way how experts think about assessed probabilities due to new insights. In the situation that
too much is between assessing the probability and checking the probability with the expert, it
is not clear if inconsistencies are as a result of the PE method used, or as a result of new
insights by the expert elicited. Because of these reasons, the assessed probabilities will be
checked directly after the elicitation. To prevent experts recognising their own assessed
probabilities, feedback will be given in a different format. This will be done by making use of
bar charts instead of probabilities. How this has been done exactly will be elaborated on in
chapter 6.5.1. With the consistency over time and the mutual consistency the obtained
probabilities will be compared. With measuring the consistency over time the probabilities
are at least obtained twice from the same expert, preferably under the same conditions. The
obtained probabilities are compared and the variance is the measure of consistency. The
smaller the variance the higher the consistency over time, which indicating that the PE
method assesses the probabilities the same, each time. The mutual consistency can be seen as
an indication on how much experts agree with one another. This will be done by comparing
obtained probabilities from different experts. The variance between the experts assessed
probabilities for the same conditional probability is an indication for how much they agree.
The results for mutual consistency are however highly dependent on the consistency over
time. In case the consistency over time is low, the mutual consistency will vary a lot, which
will result in no significant conclusions. However in case of a high consistency, significant
conclusions can be drawn because experts stick to their assessed probabilities. In the latter
situation differences of opinions for specific relationships within the BN’s can be indicated.

This section has determined two research strategies and the accompanying research design to
answer the three research questions under investigation. Now the research design has been
presented, Yin (1994) presents four criteria to check the quality of the research designs. The
quality criteria and also how both case studies answer these quality criteria will be explained
in the next section.

3.4 The quality criteria

The quality of any research design can be checked by judging the design on four quality
criteria. These criteria are: construct validity, internal validity, external validity and reliability
(Yin, 1984; van Aken et al., 2006). The following paragraphs will explain what these four
criteria imply and how they will be addressed during both case studies.

The **construct validity** refers to the quality of how the research has been performed (van
Aken et al., 2006). Triangulation is a common used tactic, in which multiple sources of
evidence are used to research the same question. Documentation, observation of experts and
interviews are examples of common used tools to obtain results from different perspectives
(van Aken et al., 2006). For the first case study semi-structured questions, observations and
documentation of the answers are used. Semi-structured questions are used during the
elicitation to deepen insights and explore the experience the experts have with the PE
methods faced with. For the second case study multiple cases are conducted and analysed
under similar conditions.

**Internal validity** is concerned with how results are obtained throughout the research.
Explanation-building is a tactic used to infer events that cannot be directly observed (Yin,
In the situation that conclusions are made about causal relationships, alternative causes of actions are investigated and blocked through the use of inference (van Aken et al., 2006). For the first case study the participating experts are all confronted with the selected PE methods in the same way. The obtained data of this case study is compared by looking for similarities and differences between the experiences the experts have with the PE methods. For the second case study an elicitation process is developed on the basis of explanation building.

**External validity** refers to the generalization of research results (Yin, 1984; van Aken et al., 2006). Although results of case studies cannot be statistically generalized (as a result of the small sample size), case studies rely on analytical generalization. With analytical generalization the investigator tries not to generalize over companies, but to generalize over a broader underlying theory (Yin, 1994; Graziano & Raulin, 2004). For the first case study the external validity is expected to be limited, because a limited number of experts from a specific PCP phase (MI phase) participate in this research. Results are therefore presumed to be applicable for all experts that operate within the PCP of PH. The resulted PE method from the first case study is input for the second case study. The external validity of the second case study is therefore likely to increase. The second case study tests the PE method with all the PCP phases and some quality criteria for the method are tested. In case the method performs positive for the tested quality criteria, the external validity increases, because the used PE method and process are generalized over a wider scope.

The **reliability** of research refers to the extent in which the research can be conducted at a later stage and still obtain the same results. Case studies rely on results obtained from humans that are highly dependent on subjective data that vary over time (van Aken et al., 2006). A case study is more reliable and valid than an experiment, because results from a case study are based on experts that base their answers on experience and raw data (Jacobs & Moll, 2007). Another way of increasing the reliability is by documenting the elicitation processes and the obtained results (Yin, 1984). With controlling biases the quality of the results increases (Meyer & Booker, 1990).

For the first case study all procedures followed and steps taken during the case study are documented. As a result the case study and argumentation used can be repeated at a later stage. For the second case study all steps taken documented in the elicitation process, making it possible to reproduce the same situation. However, as a result of subjective data it is not expected that exactly the same answers will be retrieved.

This section has discussed how the four criteria (i.e. construct validity, internal validity, external validity and reliability) proposed by Yin (1984) and van Aken et al. (2006) will be addressed during both the case studies. The next section will present the thesis outline.

### 3.5 Thesis outline

Section 3.3 has shown that two case studies have to be performed to find answers for the research sub questions. The previous section has discussed measures that will be taken to with each case study to the quality criteria as discussed by Yin (1984). The way in which the case studies will be conducted, is sequential and described by Verschuren and Doorewaard (1995) as a snowball effect. This effect represents the way in which the resulted PE method of the first case study is further investigated in the second case study. The chapters of this Master thesis project will be discussed in the same sequential order. Figure 10 gives a
The Master thesis project is divided into four sequential parts with two feedback loops.

5. The first part is represented by the first three boxes which contain the preparations for both case studies. It starts with the development of the theory (chapters 1 and 2). This is followed by chapter 3, where the research strategies are determined (determine case) and how these are conducted is explained (design data collection protocol). Finally, preparations are taken for the case studies by explaining what kind of problems and issues have to be taken into account when conducting the case studies (chapter 4). After these preparations the first feedback loop (1) follows. The first case study can now be conducted.

6. In the second part the first case study is conducted. A pre-selection of the PE methods that are going to be tested will be made in section 5.1. In the next section (5.2) a PCP phase is selected in which these PE methods are tested. This is followed by designing a PE process that will be used for the collection of the data (sections 5.3, 5.4) for the first case study. The data collection, analyses and findings are conducted in respectively the sections 5.5, 5.6 and 5.7. The first case study is documented in chapter 5. The resulted PE method of this first case study will serve as input for the second case study (the second feedback loop).

7. The third part contains the second case study. The gained experience of the first case study and literature will serve as input for designing the PE process used during the second case study (sections 6.1, 6.2). After which the second case study is conducted (sections 6.3, 6.4, 6.5). The second case study is documented in chapter 6.

8. Finally, conclusions will be drawn in chapter 7.

This chapter has presented the research outline of this Master thesis project. First the research environment has been described. This was followed by the formulation of three research questions. The first research question is interested in the most adequate PE method to use within this research environment, according to the experts. The second and third research questions investigate the resulted PE method on two quality criteria, namely the generalization of the resulted PE method and the consistency of this PE method. The procedure proposed by Verschuren and Doorewaard (1995) has been used to determine the research strategy for the research questions. The use of this procedure resulted in two case studies and an explanation how to conduct them. The quality of the research design has been
checked on the basis of four quality criteria as proposed by Yin (1994). Finally this section has described the thesis outline. The next chapter will present preparations that have to be taken into account before conducting the case studies.
4 Issues in data gathering

This chapter will discuss the problems and issues that arise during the elicitation process. In section 4.1 the general considerations heuristics and biases, are addressed because they are of influence on the whole PE process.

The following sections are following three the stages defined by Kynn (2008);
- Before the elicitation (section 4.2)
- During the elicitation (section 4.3)
- After the elicitation (section 4.4)

Thus these three consecutive stages are used to structure this chapter. In section 4.5 the issues to be addressed before starting the actual elicitation are discussed. Additionally a table is given with an overview how all the issues (in this chapter) are addressed. In section 4.6 the case study preparations are conducted.

4.1 General considerations: heuristics and biases

The elicitation process is influenced by three potential sources of bias (Kynn, 2008; Graziano & Raulin, 2004; Yin, 2004); the role of the elicitor, the role of the respondent (the expert) and the methods used. Both problems and issues can be caused by each of the three sources of bias.

In literature (for example, Spetzler & Staël von Holstein, 1975; Meyer & Booker, 1990; Cooke, 1991; Renooij, 2001; v. d. Gaag et al, 2003; Kynn, 2008) it is stated that the mental aspects (i.e. the way experts think) are of a large influence with the PE process.

The focus of this thesis will be on finding a PE method that serves in the given business environment. Heuristics and biases are discussed as a general consideration, not to be forgotten in any stage of the PE process. The use of any method should be aimed at minimizing the influence of heuristics and biases.

An explanation for the deficiency of human’s probability estimates is that they use a series of heuristics, i.e. ‘rules of thumb’, for judging probabilities. Kynn (2008) describes heuristics as the correct probability calculus or in the worst case a ‘guesstimate’ of the answer. Because this research is solely based on experts’ opinions, heuristics and biases should therefore be taken into account. The problem of heuristics and biases is that both are the cause of that people cannot be relied on to give accurate probability assessments in many contexts. The heuristics and biases in research programs illustrate cognitive imperfections. They do not discuss the processes that should be followed with PE, nor do they show when biases will appear and disappear (Gigerenzer, 1996). The problems with heuristics are that in most cases these take place, subconsciously (Spetzler & Staël von Holstein). In case respondents are influence in their subconscious, this is very difficult to notice, let alone how to measure.

Many authors have discussed heuristics and biases. Kynn (2008) however has made an overview of heuristics and biases that appear in expert elicitation, when assessing probabilities. Kynn (2008) has been thorough in her work, examining heuristics and biases over diverse fields (statistics, psychology, decision and management science), and has focussed on heuristics and biases with PE. In table 2 the list with heuristics and biases, as proposed by Kynn (2008) are shown. These heuristics and biases will be used in this section.
Kynn (2008) is not exhaustive, but addresses many established papers (for example, Tversky and Kahneman, 1974, 1983; Kadane and Wolfson, 1998; Miller, 2003; Newell, 2005). Although other authors may have addressed other heuristics and biases, in many cases these can be categorized under one of the heuristics as proposed by Kynn (2008). Kynn (2008), referring to Tversky and Kahneman (1983), takes the following three heuristics as base:

- **Representativeness** is caused by the tendency to judge the probability that X is linked to Y, by how similar or representative X is to Y. For example in case a random person (e.g. Peter) seems highly representative of a bank director, then the probability that the person (i.e. Peter) is a bank director is judged to be high even though the base rate of bank directors in the population is low.

- **Availability** is caused by the tendency to judge the frequency of an event by the ease of remembering specific examples. For example in case the last three problems in a factory had to do with the machines used, then the probability that the next problem is related to a machine will be judged high, even though a machine brake down is a rare event.

- **Anchoring and Adjustment** is caused by the tendency to rely too heavily on an initial estimate and to make adjustments to this estimate to reach the probability estimate. For example in a negotiation deal, the person who sells the product will ask a price that is higher than the price he is expecting to get. The counter party will most likely give a counter offer that is related to the initial stated asking price. The initial stated price is in this case the anchor.

An overview of the heuristics and their biases are presented in table 1. The left column shows the heuristics, as discussed above. The right column shows biases to which the heuristics might lead.

**Table 1: Heuristics and biases (Kynn, 2008), referring to Tversky and Kahneman (1983)**

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Representativeness</strong></td>
<td>- Base rate neglect/fallacy</td>
</tr>
<tr>
<td></td>
<td>- Insensitivity to sample size</td>
</tr>
<tr>
<td></td>
<td>- Insensitivity to predictability</td>
</tr>
<tr>
<td></td>
<td>- Illusion of validity</td>
</tr>
<tr>
<td></td>
<td>- Misconceptions of regression</td>
</tr>
<tr>
<td><strong>Availability</strong></td>
<td>- Disproportionate risk assessment</td>
</tr>
<tr>
<td></td>
<td>- Illusory correlation</td>
</tr>
<tr>
<td><strong>Anchoring and adjustment</strong></td>
<td>- Bias of overestimation when judging conjunctive events</td>
</tr>
<tr>
<td></td>
<td>- Bias of underestimation when judging disjunctive events</td>
</tr>
<tr>
<td></td>
<td>- Miscalibration</td>
</tr>
</tbody>
</table>

The heuristics and biases mentioned above are primarily seen from the perspective of the respondent. However, Graziano and Raulin (2004) looked at the perspective of the elicitor and identified two types of influence the elicitor can have. These are called ‘experimenter reactivity’ and ‘experimenter bias’. Experimenter reactivity is defined by any action taken by the elicitor that tends to influence the response of the respondent. Experimenter bias, on the other hand, is defined by any effect that the elicitor’s expectations might have on the obtained results of these observations. Both definitions are looking at the effect of actions taken by the elicitor. The prior definitions are the actions taken, while the latter definition is looking at the expectations. As a result, the elicitor should be aware of the impact of his/her own conscious- and unconscious influences.
4.2 Before elicitation
This section will focus on the issues that should be addressed before the PE by experts.

Issue 1: Whom to consider an ‘expert’?
Although we limit the discussion on whom to consider an ‘expert’, more reading can be done in Kadane and Wolfson (1998), where the expert is defined as "the persons who have thought harder, and over a longer period of time, about a subject at hand than others may have done.” Jenkinson (2005) refers to two types of experts, the substantive and normative experts. The former, that are present in this research, refers to people that have expertise within the application and whose probabilities are of the most interest. The latter refers to people whose expertise is in elicitation methods or have good numerical skills. A respondent of probabilities must be knowledgeable both of the subject to be analyzed and of the theory of probability (Apostolakis, 1990, Renooij, 2001b). A critical note must follow, because a subject-matter expert cannot be considered equal to a statistical or probability expert (Kadane and Wolfson, 1998).

Issue 2: Expert for qualitative and quantitative elicitation
When selecting experts for the PE, it is advised to use the same experts for the qualitative as also for the quantitative stage (Renooij, 2001; v. d. Gaag et al., 2002). Multiple experts have been participating during the qualitative elicitation (Appendix D). Within a business environment it cannot be expected that the same experts remain at the disposal of the project at all times. Even though experts may have been used during the qualitative stage, it is not guaranteed that all identified variables will be used in the model. This is due to the used method, grounded theory (Houben, 2010). As a result, even though the same expert may have participated in the qualitative and quantitative stage, it is very likely that the expert will be confronted with ‘new’ variables during the elicitation. This cannot be prevented. Not every expert can be satisfied. Tackling of this issue will be discussed under issue 7.

Issue 3: Prior biases
Decisions made during the qualitative stage directly influence the quantitative stage (Spetzler & Staël von Holstein, 1975; Sigurdsson, 2001; Renooij, 2001a). Drudzel and v. d. Gaag (2000) warn for bias in the sample data and/or model that is used as input in this research. If biases have already been introduced during the first stages of building a BN, these biases may increase during the second stage (quantifying the BN). Since this research concentrates only on quantifying BN’s, the following assumptions can be made about the qualitative elicitation.

- Each BN constructed during the qualitative elicitation (i.e. the BN’s used as input), covers all the relevant variables (nodes) that are needed to investigate reliability during that stage of the PCP.
- With the determination of the relevant variables during the qualitative elicitation, the size of the resulting NPT’s has been taken into account with the construction of the BN’s.
- The cause-effect relationships represented by the BN’s are assumed to be true.

Issue 4: The selection of experts
While Kadane and Wolfson make a distinction between experts, Drudzel and v. d. Gaag (2000) make the critical note that if no or few reliable data is available, the knowledge and experience of the expert in the domain of applications remains the only source of probability assessment. The selection of experts is the most controversial part of using expert opinions...
(Apostolakis, 1990). In an ideal situation, the experts should be selected that have the necessary domain knowledge and are familiar with assessing probabilities (issue 2). Due to the nature of expertise this is a scarcely available (Renooij, 2001b). To obtain insight from multiple points of view it is also better to have multiple experts involved (Renooij, 2001b). The fundamental principle that underlies the use of multiple experts is that multiple experts know more than one individual (Winkler & Clement, 1999). In the case multiple experts can be used, heterogeneity and several stakeholder groups are highly desirable (Apostolakis, 1990; Winkler & Clement, 1999). Winkler and Clement (1999) state that in the situation where experts are very similar (i.e. experts have the same philosophy, modelling style and access to the same data) redundancy of information happens sooner. There analyses point out that in this situation the optimal number of experts, to obtain an average insight, is between three to five experts (Winkler and Clement, 1999).

The selected experts should be divided into two groups. This first group should contain a small number of experts on which a dry run is performed (Cooke, 1991; Meyer & Booker, 1990). Pilot testing with sample experts gains insights in the logistics of the elicitation. It will provide information on the time it takes to perform one elicitation session, but also if the experts are able to cope with the provided documentation. This will provide information on whether or how the elicitation needs to be revised (Meyer & Booker, 1990). The second group contains the rest of the experts with whom to perform the research. Even though the dichotomy has been presented in literature, a ratio is not found.

**Issue 5: Willingness to cooperate**

Bedford et al. (2006) state that reliability is one of the requirements experts are confronted with during their work. As a consequence experts may see research towards this topic as criticism of their work. Experts should therefore be approached with respect as described by Meyer and Booker (1990). They propose a formal procedure on how to address respondents. Because the research is solely dependent on experts, it is important that the experts are willing to cooperate (Renooij, 2001b).

**Issue 6: Familiarizing experts with method**

When the experts have been selected and are willing to cooperate with the research, the next step is to familiarize the experts with the elicitation method (Renooij, 2001b; Garthwaite et al., 2005). A brief explanation should be given about the reason for the elicitation, what is expected of the expert, how the BN structure works and a short training should be given on estimating probabilities (Cooke, 1991; Renooij, 2001b; Jenkinson, 2005; Kynn, 2008). Next to the preparations of the PE, an explanation should be given in terms of possible use of calibration, scoring rules and evaluations (Cooke, 1991; Kynn, 2008). The first two cannot be performed in this research due to a lack on seed variables, see section 6.5 for further explanation. Renooij (2001b) summarizes these aspects by “Elicitation method should be straightforward, easy to handle and not difficult to learn”.

**Issue 7: Unambiguous**

To be able to conduct a PE, the BN should be unambiguous (Spetzler & Staël von Holstein, 1975; Cooke, 1991; Renooij, 2001b). Although unambiguousness is a prerequisite, it is very hard to formulate unambiguous questions (Cooke, 1991). The formulation of questions is preferably neutral, not positive, nor negative. If this is no option, one can try both options to double check (Kynn, 2008).
Renooij (2001b) state that to make the PE process unambiguous the questions should be structured and definitions of the variables and states should be documented. These documents should be within reach during the assessment of the probabilities for the respondent. In case the experts are used for both the qualitative and quantitative elicitation (issue 2), the experts are more independent of these documents (Renooij, 2001b) than experts who have not participated in the qualitative elicitation. However, unambiguousness is a means to make sure that any respondent understands the stated questions. Unambiguousness can be tested by using the clairvoyant test (i.e. if a clairvoyant could reveal the probability without requesting clarification), as described by (Spetzler & Staël von Holstein, 1975). Next to the difficulty of defining the variables and states as unambiguous variables, the states and scales used should also be meaningful to the respondents (Spetzler and Staël von Holstein, 1975). This can be achieved by relating the questions to the area of expertise (Kynn, 2008).

**Issue 8: Making use of subjectivity**
The number of tests that can be performed on an expert is a delicate topic, because collecting data from humans is completely different from collecting test-data from products/machines. In previous research a limited number of experts were used for the research (Renooij, 2001a), or PE methods are used in combination with hard data sets (among others, Fenton et al., 2006)

In this research all the probabilities for the complete network have to be assessed. The difference between subjective measurements (from experts) and objective measurements (from machines/products) is the lack of memory with the latter. Humans have two disadvantages. First humans have a memory, which makes it very hard to conduct an experiment without knowing the previous test (Does et al., 2001). Secondly, individual responses differ, because humans will react to the environmental stimuli. (Meyer & Booker, 1990). Whereas a machine can be tested under identical conditions, this is not possible with subjective data (Does et al., 2001).

**Issue 9: Individual- vs. group elicitation**
Because each participating expert has his own kind of knowledge, choices have to be made on how to collect the data from the selected experts. Use can be made of individual-, group elicitation or both to come to one assessment, as proposed by Renooij (2001b). When each expert is elicited individually, results are combined at the end. The individual assessment has as advantage that it can be extended mathematically, but presumes that the interaction with other experts does not add value. However, the group elicitation has as advantage that the knowledge of each expert is taken into account when coming to a consensus (Garthwaite et al., 2005). The disadvantages that are encountered, next to the individual problems, are group problems. An example is that a strong personality has too much weight in the discussion, or that overlapping experiences are over weighted due to repetition in discussion (Meyer and Booker, 1990; Renooij, 2001b; Garthwaite et al., 2005). For the elicitation of group problems a knowledgeable and experienced facilitator is required who is able to oversee all the problems encountered in a group elicitation (Garthwaite et al., 2005). Dependent on the method chosen each option has an own process to follow during the PE (Meyer and Booker, 1990).

This section presented an overview of nine issues. These issues have to be taken into account before PE can take place. The next section will focus on the issues that are of importance during the PE.
4.3 During elicitation
The previous section has described the issues that have to be taken into consideration before the PE of experts. The focus in this section is on issues that appear when obtaining the probabilities. Solely obtaining the probabilities is discussed, not the verification of the elicited probabilities. The issues discussed in this section should be addressed before the actual PE takes place. In this way the elicitor is prepared for the elicitation and aware of possible difficulties that may appear during elicitation.

Issue 10: Influence of the elicitor
The elicitor is of a major influence on how the PE is performed (i.e. the elicitor determines the procedure, the formats used, the number of experts, etc.). One of the choices the elicitor has to make is about his own participation in the elicitation. In case the elicitor is absent, this will automatically result in a passive form, while a present elicitor can both be passive or active. An active elicitor participates in the PE by experts, which is the case in all the indirect PE’s. Another advantage is that the elicitor is able to continue asking for a deeper insight on how the experts have reasoned their probabilities. However, an active elicitor should prevent coaching the expert, because he/she is not the expert (Cooke, 1991) A passive elicitation, on the other hand, has as advantage that the experts receive a similar approach. This can be done by making use of, for example, a PE document in which each expert receives the same input. This makes it less likely that results are influenced by the elicitor.

Issue 11: Tasks of the elicitor
In the case the elicitor is present during an elicitation, it is expected that he/she performs a number of tasks. Renooij (2001b) has defined the following tasks the elicitor should perform during the elicitation:
- Clarify inevitable problems the experts will come across. Examples are wrong interpretation of problems, definitions, variables, states, and so on.
- Record all the answers for which the answer format does not provide space, but may still be of use.
- Sometimes some context for the estimation of certain variables is wrong, or unnecessary. In these situations the expert has to adjust the structure of the elicitation process.
- In the expectation that biases are introduced, the elicitor should make the respondent aware of this.
- The elicitor should control the time. The duration is discussed in issue 13.

Issue 12: Reasoning of the respondents
Measuring experts’ opinions is considered the most difficult, because experts find it hard to express themselves in terms of probabilities (Sigurdsson, 2001) and Cooke (1991, p61) states: “Probability thinking is much more subtle and tricky than ordinary logical thinking”. To get insight in how respondents reason, the assessed probabilities are therefore important (Spetzler & Staël von Holstein, 1975). A view on the reasoning process of the respondent can be gained through:
- Explicit reasoning
- Observations

Insight in the reasoning process of the expert has as advantage that possible heuristics used by a respondent can be prevented (Kynn, 2008). For example, the respondent returns to something he has already done, which could indicate that the respondent used different underlying assumptions. As a result, reasoning by the expert, observations and questions
stated by the elicitor can make that obtained probabilities are more supported by the experts’ opinions. This results in a higher validity of the answers (chapter 6).

**Issue 13: Duration**
This issue concerns the duration of the PE. Multiple authors have addressed this issue (Kynn, 2008; Renooij, 2001; Cooke, 1991; Spetzler and Staël von Holstein, 1975). Duration of the elicitation depends on the importance and complexity of the uncertain quantity and the experience of the respondent with PE (Spetzler and Staël von Holstein, 1975). Different times have been mentioned. In case only a limited number of variables had to be identified, 10 minutes were sufficient. However, the duration that an expert can focus on the assessments and pay attention has been found to be at most ninety minutes. The elicitation has been found to be more intensive for the respondent than for the elicitor (Cooke, 1991).

**Issue 14: Inform about possible biases**
During an elicitation, experts are often not aware of the heuristics and biases that have influence on their own performance (Spetzler a& Staël von Holstein, 1975; Meyer & booker, 1990). It is therefore difficult to control the bias people use. If the experts are made aware of possible biases the expectation is that the expert is conscious and is able to control its effects to a greater extent (Spetzler & Steal von Holstein, 1975).

This section has addressed five issues that are important during PE. The next section will address issues that are of importance once the probabilities have been obtained from the experts.

### 4.4 After elicitation
The previous section has discussed the issues that are of importance during the PE. Once all the probabilities for the NPT’s are assessed, these should be checked. This section will discuss some issues that are of importance at the end of the elicitation.

**Issue 15: Evaluation of the estimated probabilities**
Once results are obtained, the elicitor is interested in finding to what extent the expert agrees with their own probabilities. Verification is among others proposed by Sigurdsson (2001) and Renooij (2001b) as a proper method. However verification is defined by Renooij (2001b) as: “the process of checking whether the probabilities provided by the expert are well calibrated (conform to observed frequencies), obey the laws of probabilities (are coherent) and are reliable”. The reliability of the method is controlled by the elicitation process which is carefully documented. Coherent answers cannot be determined without aid, for example a computer (Kynn, 2008). The problem however is that in our case there is no available data to calibrate the experts with. As a result, verification of the obtained probabilities is not possible due to a lack of the possibility to calibrate the experts. Kynn (2008) proposes to offer process feedback, for example by making use of different representations of probability (say graphical), to give a summary and allow experts to reconsider probability estimates or confirm the stated probability estimates.

**Issue 16: Accuracy of estimated probabilities**
Current research checks the accuracy of estimated probabilities by making use of historical data.
In literature the accuracy of the data obtained with a PE method is always checked with historical data. (among others, Cooke, 1991; Renooij, 2001a; Fenton et al., 2006; Marquez et al, 2007; Norrington et al., 2008) The historical data are used to construct seed variables. These seed variables are variables (nodes) from which the probabilities are elicited. The obtained probabilities are subsequently compared with the historical data as indication for the accuracy of the expert (Cooke, 1991). However, in this research the data set is dependent on experts’ opinions (subjective data) only. Checking the accuracy of obtained probabilities with historical data is therefore not applicable.

The data retrieved from experts cannot be checked on accuracy by using other available data, as has been shown in the previous paragraph. The accuracy of the elicitation method during this research might be evaluated by an instrument to check the accuracy of objective data. A practical example of measuring the accuracy of objective data is the well known ruler. The ruler is a measurement device that has been used over the years. The accuracy, the ease of use and the variance of this measurement method are well established. This measurement method has been proved through multiple tests. Does et al. (2001) describes factors that have to be investigated when analysing a measurement method. These factors are:

- Resolution
- Calibration
- Repeatability
- Reproducibility
- Stability
- Linearity

The resolution is the accuracy with which results can be indicated by respondents. Calibration refers to the method that is chosen to gain a zero-point to serves as base. At every moment within the use of the method, this can be used to set the method to the zero-point. Repeatability is the variance that is caused by the measurement method. Reproducibility on the other hand, is the variance as a consequence of changes in the environment within which is measured. Stability is an indicator that looks at the average and variance of a method in the long-term. Linearity is also applicable with non-linearity and looks at the bias over the complete range of the method.

### 4.5 An overview of the issues

In this section an overview (table 2) is shown of all the issues discussed in the previous three sections. The first column (seen from left to right) states the issue numbers. The second column shows the name of the issues. In the third column an answer has been given to the question if that specific issue has been tackled (‘yes’ or ‘no’). In case the issue has been tackled, a reference is given in the fourth column. Otherwise an explanation has been given in the final column.
### Table 2. Overview of tackled issues.

<table>
<thead>
<tr>
<th>Issue</th>
<th>Issue name</th>
<th>Tackled?</th>
<th>How tackled?</th>
<th>Why not tackled?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Whom to consider an ‘expert’?</td>
<td>Yes</td>
<td>Proper definition obtained</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Expert for qualitative and quantitative elicitation</td>
<td>Yes</td>
<td>Used the participating experts from the qualitative stage as input for the quantitative stage (4.6)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Prior biases</td>
<td>Yes</td>
<td>By making assumptions</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>The selection of experts</td>
<td>Yes</td>
<td>The selection has been based on the same experts used during the qualitative stage (4.6)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Willingness to cooperate</td>
<td>Yes</td>
<td>By informing experts before PE (p.38)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Familiarizing experts with method</td>
<td>Yes</td>
<td>By creating a PE process/document for each case study (5.4 &amp; 6.2)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Unambiguous</td>
<td>Yes</td>
<td>Using a PE document and list of nodes and their definitions for both case studies (5.4.1 &amp; 6.1)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Making use of subjectivity</td>
<td>Yes</td>
<td>Only individual elicitations</td>
<td></td>
</tr>
</tbody>
</table>

### 4.6 Case study preparations

The previous sections have described the issues that play a role in conducting a research. This section addresses how preparations for both case studies (1 and 2) have been made. These preparations ought to be done before starting with the case studies on quantifying a BN.

**Knowledge transfer between qualitative and quantitative stages**

A ‘peer mentoring’ relationship is used to gain insight in how the qualitative stages have been performed (Boer et al., 2004). Through the use of peer mentoring the transfer and creation of knowledge is stimulated between the qualitative elicitor and quantitative elicitor (Peroune, 2007). As a result, the quantitative elicitor is not able to change the results of the qualitative phase (i.e. variables in three BN’s), but is aware of the possible places where biases may have been introduced in the previous stage. Therefore issue 3 cannot be prevented, but the quantitative elicitor is informed about procedures followed during the qualitative stages able to act on this information.

**Selecting the experts**

This research is a continuation of previous research (Houben, 2010). As a result, experts have already participated in the qualitative elicitation. An overview is presented in appendix D, of the participating experts, their functions and number of variables they identified for each
stage of the PCP. For this research, it is tried to use the same experts again for the quantitative stage to meet the conditions stated in issue 2. Because this research is taking place in a business environment, it is not always possible to comply with issue 4.

Based on issue 4 (Selection of experts) two groups are made from the participating experts. The first group serves as test panel and is only a select number of experts. These experts participate in both case study 1, where the selection of a response mode takes place, and in case study 2 for the approval of the elicitation process. These experts differ from background so as to have a better representation of the sample. Due to the limited number of experts participating, the complete sample size cannot be represented in case study 1. The second selection participants participate in case study 2, where the selected response mode is implemented in the elicitation process to gain insight in how valid and reliable retrieved quantitative data is.

**Willingness to cooperate in practice**
In the first place experts are selected that already took part in the research. If necessary, other experts are approached to participate in this research. A presentation is given at the beginning of this research to inform the participants both of the results of the qualitative stages and of what is expected of them in the future. This gives experts an insight into the process in which they are involved, creating a base of support for this research.

Because PH has an informal business culture, the formal procedures, proposed by Meyer and Booker in issue 6 (willingness to cooperate), are not followed step by step. However, the experts are approached with respect by keeping in mind that participation is on a voluntary base. As a result experts are informed of the aim of the research before the meeting takes place, meetings can be re-scheduled and interviews are limited by the duration planned in the agenda’s varying from 30 minutes up to 2 hours (including a break) to comply with the duration (issue 13). This shows that every business environment has its own approach.

**Individual vs. group elicitation**
The aim of PE is to transfer experts’ opinions into quantitative data. This can be done individually and/or by conducting group sessions (Renooij, 2001b; Garthwaite et al., 2005; Meyer and Booker, 1990). As indicated in issue X, there are advantages and disadvantages for both methods. However, group elicitations are more difficult to manage in a business environment (v.d. Gaag et al., 2002). Compared with individual elicitations, the number of problems that have to be controlled increase significantly (among others, Meyer and Booker, 1990; Renooij, 2001b; Garthwaite et al., 2005). To be able to evaluate the accuracy of the PE method (issue 16) individual results are necessary (issue 9). In the three group elicitations, discussed by Jenkinson (2005), two group elicitations first investigate experts individually. A problem with group elicitations is that it is not clear whose opinion is measured (Gairthwaite et al., 2005). This makes the reproducibility more difficult. Next to this all, the qualitative research has also been performed individually (Houben, 2010).

The approach has to serve the research questions and is, like the willingness to cooperate, also dependent on the business environment. Individual elicitations have a couple of advantages in this feasibility study. PH experts are already used to this method, individual elicitation is easier to manage for the elicitor, it is easier to make appointments and group elicitations are still possible in future research, but not the other way round. As a result, to gain as much insight as possible about the PE method and its applicability, each elicitation will be performed individually.
Unambiguousness in practise
With multiple experts participating in the research, a prerequisite is that the variables and states applicable to an individual should be unambiguous (issue 7) (Renooij, 2001b; Cooke, 1991; Spetzler and Staël von Holstein, 1975). Due to the different background knowledge of each individual expert this is difficult to accomplish (Cooke, 1991). As a result unambiguousness was found to be more an aspiration. This research has reached the level what was feasible within PH. Each individual’s opinion cannot be taken into account nor can the BN be customized for each expert. As a result choices have been made which opinion to adopt in the model. Averaging is not possible, because in such a situation none of the experts agrees with the model.

To prevent biases from not complying with unambiguousness (issue 7) the following actions were performed. Firstly a document is constructed which contains the definition of all the applicable variables and states in a structured manner. Secondly this document is always present during elicitation and can be checked at all times by both the elicitor and respondent to comply with the conditions for unambiguousness as stated by Renooij (2001b). The first action, constructing a document, has been done in two steps. The first step is to formulate the definitions of variables and states to pass the clairvoyant test as described by Spetzler & Staël von Holstein (1975). The second step is to cross reference these definitions with at least two participating experts to check for similar interpretation. The final action taken is that at least one expert is present during the elicitation to explain the definitions if misinterpreted (issue 11, 12) (Renooij, 2001b).
5 Probability elicitation method selection

In this chapter case study 1 is described. This chapter contains seven sections. Section 5.1 contains a description of the pre-selection of the type PE method used for the first case study. In the next section (5.2) the choice for one of three PCP phases is discussed. In section 5.3 what’s and why’s of case study 1 are given. In section 5.4 the process of putting together the PE document is described. In section 5.5 the data gathering process is described. In section 5.6 the data obtained during the interviews, held with experts, are analyzed. Finally, in section 5.7 the first research sub question (see 3.2) is discussed.

5.1 Pre-selection of the probability elicitation method

A distinction has been made between three different types of PE methods (Chapter 2): direct-, indirect- and parametric elicitation. A selection of one of the three types of PE methods has to be made. This is necessary because investigation of all three types of PE methods is not feasible (lacking of time and experts). The choice for one type of PE method is underpinned by selection criteria.

No large multi-method studies have been found where experts had to assess a large number of probabilities with every single method (Renooij, 2001a). Therefore, to find out which type is adequate, it is necessary to conduct a multi-method study on a limited scale, which is the subject case study one.

5.1.1 Parametric elicitation

Parametric elicitations are based on continuous variables. The BN’s used in case study 1 are based on variables, which have discrete, as well as a non-parametric state space. As a result the parametric-type of PE methods are not applicable for the BN’s used in case study 1 and will not be studied or discussed further.

5.1.2 Indirect elicitation

Indirect elicitation methods have been developed to suppress the influence of biases (section 4.1) (Renooij, 2001a). These indirect methods aid the experts in assessing their probabilities in probabilistic terms, limiting what they need to know about probability theory (Kadane & Wolfson, 1998). The use of indirect elicitation methods is confined by the number of probabilities that have to be assessed and the available time, since indirect elicitation methods can take up to 30 minutes per probability (Druzdzel & v. d. Gaag, 2000). This suggests that if more time is used to assess a probability, the answer becomes more precise. This trade-off between available time and precision has been doubted in literature (Kadane & Winkler, 1988). The time an expert has available for elicitation, therefore limits the choice of PE methods applicable. For our BN’s between the 189 and 287 probabilities have to be elicited within a maximum of ninety minutes.

This research is dependent on the number of experts, who had to be attracted to participate in this research, and the available time they have. Participating experts are not willing to spend $30 \times 189 = 5670$ minutes (94.5 hours) on PE. Therefore indirect PE methods are not used in this research for this pragmatic reason.

5.1.3 Direct elicitation

In this case study direct PE methods are used for the reasons mentioned under 5.1.1 and 5.1.2. Case study 1 focuses on finding out which of these methods is the most adequate,
according to experts. Direct elicitation is said to be unreliable (e.g. Apostolakis, 1990), Renooij (2001a) and van der Gaag et al. (2002) coped with this problem by using a special format for direct elicitation. Those formats are containing questions and answers. Their questions (Renooij (2001a) and van der Gaag et al. (2002)) were transcribed mathematical notations of the relationships between nodes. Their answering format is a verbal-numerical scale on which experts indicate their assessed probability. This method has proven itself, using medical experts. It is tested in this research to find out if it will hold, using technical experts.

Figure 11 shows an example BN. This BN is introduced to illustrate the selected direct PE methods. The BN in figure 11 shows two parent nodes ‘experience with respect to installation’ and ‘following installation procedures’, influencing one child node ‘quality of installation’. The parents have two states each, ‘much’, ‘little’ and ‘yes’, ‘no’, respectively, and the child node has three states, i.e. ‘excellent’, ‘average’ and ‘bad’. The different PE methods used are set in this format.

![Bayesian network example](image)

Figure 11. Example Bayesian network to determine the quality of installation.

The workload involved in eliciting the BN in figure 11 can be deducted from the total amount of probabilities that need to be elicited. For the parent nodes this implies four estimations (i.e. the probability for each state (‘much’, ‘little’, ‘yes’, and ‘no’)). Since the parent nodes have two states each, and the child node has three states, this results in 2x2x3 = 12 estimations for the child node. In total this is 12+2+2=16 estimations in order to do a complete elicitation for this simple example BN. It is easy to see that the expert’s effort will explode in more complex BN’s.

The NPT for the child node of the example BN (figure 11) can be elicited by making use of three direct PE methods. By the direct PE method (method 1 “scale”), as proposed by Renooij (2001a), each probability is elicited with a separate question (see “circle 1”). With the second direct PE method (method 2 “open answers”), the number of questions is reduced by asking one question per situation (see “circle 2”). The number of questions asked for the child node is thus reduced from 12 (2x2x3=12 probabilities) to four (i.e. for each situation of the child node). Using the third direct PE method (method 3 “table”), the number of questions is even further reduced by asking one question per node (see “circle 3”). With one question the total NPT for one node is elicited.

**Elicitation per probability (method 1)**

The question and answer format as used by Renooij (2001a) is shown in figure 12. Here, the conditional probability is assessed for achieving an excellent quality of installation, given the
conditions that the installation procedures are followed (state ‘yes’) and the expert has much experience with respect to installation (state ‘much’) (Figure 11 and Figure 12).

Consider that a system has been installed at a customer site following the installation procedures, by an expert with much experience with respect to installation. How likely is it that the quality of the installation can be considered excellent?

<table>
<thead>
<tr>
<th>Probability</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(almost) impossible</td>
</tr>
<tr>
<td>15</td>
<td>impossible</td>
</tr>
<tr>
<td>25</td>
<td>uncertain</td>
</tr>
<tr>
<td>50</td>
<td>fifty-fifty</td>
</tr>
<tr>
<td>75</td>
<td>probable</td>
</tr>
<tr>
<td>85</td>
<td>(almost) certain</td>
</tr>
<tr>
<td>100</td>
<td>certain</td>
</tr>
</tbody>
</table>

Figure 12: Question and answer format on a scale as used by Renooij (2001a)

The expert will give his opinion/estimation on the verbal-numerical scale as given in Figure 12. One question and answer (i.e. an indication on the scale) needs to be executed for each probability.

Elicitation per situation (method 2)
To reduce the number of questions that are put to an expert, the choice is made to test another method in which the same question is asked as in the previous method, but instead of only asking for the conditional probability of an excellent quality of installation, the conditional probabilities for all three states (excellent, minor and bad) of the quality of installation are asked. Instead of presenting three scales, the expert divides 100% (or 1) over the three conditional probabilities (Figure 13). Although the expert still has to estimate 16 probabilities, this is done by asking 6 instead of 16 questions. This method is also easier for both expert and elicitor to check for the laws of probability (∑=1 or 100) after each question.

Consider that a system has been installed at a customer site following the installation procedures, by an expert with much experience with respect to installation. How likely is it that the quality of the installation can be considered excellent; ...?

excellent; ... minor; ... bad; ...

∑ = 1 or 100

Figure 13: Transcribing probabilistic mathematical notations as fragment of text with open answers

Elicitation per node (method 3)
Because the selection of a method is influenced by the ease of use for experts and elicitors (issue 6; familiarization with method) (Renooij, 2001a), a table/matrix format is also investigated. Since the experts involved in our research have a technical background, it is expected that they are comfortable using and understanding table/matrix formats. The table/matrix format has, to the knowledge of the author, never been used before for direct elicitation. Table 3 shows such a table, in which the 12 probabilities of the child node are assessed at once. In the first (most left) column, the parent node ‘Experience wrt installation’ is shown, with the two states, ‘much’ and ‘little’. The second column shows the parent node ‘Following installation procedures’. To describe each possible situation for the child node, the states ‘yes’ and ‘no’ are placed behind both states of the other parent node (‘much’ and
‘little’). The child node ‘Quality of installation’ is shown in the top row. The three possible states, ‘excellent’, ‘minor’ and ‘bad’ of the child node are shown in respectively the third, fourth, and fifth row. Each row represents a situation and the possible outcomes in which the child node has to be evaluated. In this way a 3x4 matrix is obtained in which the 12 estimated probabilities can be filled in.

Table 3. Example table format in which two parent nodes, each with two states, influence a child node with three states.

<table>
<thead>
<tr>
<th>Experience wrt Installation</th>
<th>Following installation procedures</th>
<th>Quality of installation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much</td>
<td>Yes</td>
<td>Probabilities</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.2 Selection of the Product creation process phase

Case study 1 has been focussed on one of the three PCP phases. This has been done to use the available experts as well as their time as efficiently as possible. Thus choice for one PCP phase has been made. The selection criterion for making that choice is the optimal reduction of PE time involved. The PE time is determined predominantly by the complexity of the given BN linked to each PCP phase. The complexity of a BN is determined by:

- the number of nodes
- the number of cause-effect relationships between the nodes
- the size of NPT’s (on average)

The BN belonging to the third PCP phase, “Manufacturing and Installation” (MI) is considered the least complex, because it contains the lowest number of nodes, the lowest number of cause-effect relationships as well as the smallest size of NPT’s (on average).

5.3 The what’s and why’s of case study 1

In this section what’s and why’s of case study 1 are described keynote wise. Table 4 shows in the left column what’s (actions) that have been undertaken and a reference to a section. In the right column the reasons why are shown as well as a reference to a section.

Table 4. The what’s and why’s of case study 1.

<table>
<thead>
<tr>
<th>WHAT’S (actions)</th>
<th>WHY’S (reasons)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preconditions</strong></td>
<td></td>
</tr>
<tr>
<td>- Test three direct PE methods (5.1)</td>
<td>- Testing direct PE methods are the only feasible (5.1)</td>
</tr>
<tr>
<td>- Test in the third PCP phase (5.2)</td>
<td>- Use available experts and their time as efficient as possible (5.2)</td>
</tr>
<tr>
<td>- Contact the experts of the MI phase (4.6)</td>
<td>- Willingness to cooperate (issue 5, 4.2)</td>
</tr>
<tr>
<td><strong>Actions</strong></td>
<td></td>
</tr>
<tr>
<td>- Construct a PE document (5.4.1)</td>
<td>(5.4.1)</td>
</tr>
<tr>
<td></td>
<td>- Preventing unambiguousness</td>
</tr>
</tbody>
</table>
Experts reason from gained experience with the PE methods. Limit the influence of the elicitor and limit the influence of biases.

- Conduct elicitation (5.5)
- Conduct interviews with semi-structured questions and observations (5.5)

Analyse results
- Analyse obtained answers from experts (5.6)
- Draw conclusions from analyses (5.7)

**5.4 Case study 1: probability elicitation document**

In this section the development of the PE document, used for case study 1 is described. The PE document is used as guidance for the expert during the PE and as guidance for the interviewer to conduct the interview. The PE document contains elicitation questions, evaluation questions (with regard to the appreciation of the PE methods used, by the expert) (appendix E) and a list of nodes and their definitions (appendix F).

**5.4.1 Contents of the PE document**

The use of a standard document and defined nodes has four benefits. The first benefit of the use of a document and list of nodes and their definitions is that every expert interprets the elicitation in the same way (Renooij, 2001a), preventing ambiguousness (Issue 7). The second benefit is that experts are able to argue their choices for the different elicitation methods on their gained experience with elicitation. Thirdly, the document makes it possible to limit the influence of the elicitor on the process (issue 10) (limit the introduction of biases), creating the opportunity for the elicitor to fully focus on observing and questioning the expert. Finally, each elicitation was performed face-to-face in a closed office with at least one elicitor present to indicate the importance of the research (Cooke, 1991; Renooij, 2001b) and to prevent the expert becoming distracted by giving them the opportunity to ask questions, and observe the expert during the elicitation.

The PE document (appendix E) contains the following parts:
- Introduction: explaining the purpose of the interview.
- Semi-structured questions about the experts experience with estimating probabilities
- Three questions to show the two direct PE methods (method 1 and 2) to the expert
- Semi-structured questions about the appreciation of the two PE methods by the expert
  (interview moment 1)
- Seven questions to show two direct PE method (method 2 and 3) to the expert
- Semi-structured evaluation questions (appreciation as well as an estimate of the accuracy by the expert of the actual elicitation) (interview moment 2)

**5.4.2 Contents of the interview**

During the interview in first instance the purpose of the interview is explained to the expert. This is followed by questions related to their experience with making probability estimations. Following this the elicitation is conducted.
During the elicitation questions are asked at two different moments in time. The first moment is after method 1 and method 2 have been used. The expert has gained experience in elicitation with these methods. Questions are asked about the experience with the methods used and the expected accuracy of their answers. The second moment is after method 3 has also been used. At this point, the complete BN has been elicited. Then, some evaluation questions follow. The gained experts experience in elicitation has now shifted from experience with few elicitation questions (assessment of 16 probabilities) to experience with a higher number of elicitation questions (assessment of 65 probabilities elicited). Again questions are asked about the experience with the methods used and the expected accuracy of their answers. This is done to investigate whether experts think differently about the use of methods under different workloads.

The two moments can be described as follows.

**Moment 1:** The experts have experienced two PE methods (method 1 and 2). Afterwards questions are asked relating to the experience with the PE methods and the expert’s ability to perform the same action in a table format. Finally experts are asked whether they think the methods will influence their accuracy.

**Moment 2:** The experts have performed the complete elicitation. They are now familiar with the use of method 1, 2 and 3. At this point in time the evaluation is targeted at all the experiences they have with the PE methods. The experts are asked to compare the three PE methods. Finally the question is asked if they are lacking any knowledge to conduct such an elicitation.

The next section is an example of an interview during the elicitation to show how responses were given and how deeper insights were created as a result of the open structure of the interview.

### 5.5 Case study 1: data gathering

In this section the actual data gathering is described. Four experts are participating during case study 1.

An example of one of the four interviews held is given below (because the questions are asked primarily in Dutch, Dutch answers are freely translated). All the answers of the interviews are schematically shown in appendix G. The quantitative data, gathered during the actual elicitations, are not discussed in this section, because they are irrelevant for the answering of the first research sub question (3.2).

**Example interview with expert d**

- **After moment 1** -

  **Question-1:** With the questions from the previous page, were you aware that you were filling in a matrix?

  Answer: Answering questions is straight forward, because one situation is given at a time that needs to be answered. Without the overall view, it is easier to answer the questions.

  **Question-2:** How consistent do you think your answers are without the overall view?

  Answer: By making use of questions you fill in your first thought. With a table you adjust your answers according to the other answers given for one relationship within the BN. Questions give more space for answering.

- **After moment 2** -
Question-3: Looking back, what is your experience?
Answer: It is more efficient to fill in a table, but your first thought is influenced. You try to find your own logic.

Question-4: At which point did you adjust the table?
Answer: First you fill it in, but along the way you see the inconsistencies of your own answers. By the introduction of fixed intervals between variables, you are able to differentiate the answers the way you are comfortable with. I adjusted the answers both ways. I adjust my opinion and think the table is also a reliable method. However, elicitation in general is scary, because answers are based on gut feeling, instead of facts.

Question-5: What was your experience with the scale?
Answer: Hard to fill in, due to gut feeling.

Besides asking questions during the elicitation, the elicitor made observations. The observations have been documented and an overview is presented in the next paragraph.

Observations made during the case study
The experts were asked to use a document which would guide them through the elicitation of the MI phase. This gave the elicitor the opportunity to observe the experts during the elicitation. The following observations have been made:

- Experts hesitate in giving answers which are not based on factual data.
- The first table is often not understood without an oral introduction, even though the first table has been introduced by previous questions and supported by a picture of the BN.
- The verbal-numerical scale is no problem, but new to experts. With the same ease they divide 100%.
- Each expert made use of a 0 to 100% scale. No one made use of a 0 to 1 scale to indicate his probabilities.
- The scales have a high ease of use with the majority of experts, in the case of a variable with binary states. However it has been found hard to interpret the marker, ‘/’, given by experts when not on an anchor. This results in a rough estimate. With a binary state only one marker and question is requested, however with more states the numbers of scales increase per question. The laws of probability are also more difficult to check, because no absolute probability is given on a scale. As a result interpretation is left to the elicitor.
- Experts erase more with tables than with the open answers format. This is done, amongst other reasons, to be consistent in answers and to keep proportions as experts think they should be.
- Experts have trouble interpreting the questions each time. Every time they have to imagine the new situation, instead of imagining the same situation with only one variable changing at the time, as in the case of the table format.
- After moment 1: experts find the open answers format easy and can keep the overview (4 questions).
- During moment 2: you experience experts confronted with their own probability estimations which cause experts to work more systematically.
- During moment 2: experts are concentrated throughout the table, imagining one situation in which they vary the conditional probabilities.
- Experts filled in tables with different strategies: from top-down (sequential), extreme values first and then the others, most-likely values first and then the others.
- Experts had no trouble filling in tables.
With a table format, experts want to make a distinction between the different conditional probabilities.
Experts must be told that there is a list with nodes and states and their definitions, which they should use.

Now the answers and observations have been gathered, the gathered data is analysed in the next section.

5.6 Case study 1: analysis of answers and observations

After the data have been gathered the observations and answers are grouped and analyzed. This section uses the groups in which the questions are asked during the elicitation (section 3.3) as base to analyse the gathered data.

The expert’s experience with estimating probabilities
By asking experts as to their experience, experts are not asked in how good they are. The majority (3/4) indicated having experience with estimating probabilities. The experts found it easy to divide the 100 point scale and know how to do this. The majority indicated not to be very experienced in using probabilities during daily practises. Failure mode effect analysis (FMEA) is indicated twice as experience. During a FMEA experts have to imagine failure modes and estimate among other things the likelihood of occurrence. However, this type of question does not say anything on how good experts are in estimating, it does show that experts are familiar with imagining hypothetical situations and estimating the likelihood of occurrence of such situations.

The expert’s appreciation about the methods used
One of the criteria on which to judge the obtained results is by grading the responses of the experts (Yin, 1994). Each PE method is judged by the responses given by each expert. Responses of the expert are graded as a positive or negative response. The number positive responses are measured as a positive indication for the use of that method, and vice versa.

The character of the answers obtained from the experts has been presented in table 4. In the left column the three direct PE methods are presented (scale, open-answers and table from section 5.1). The results of the table are divided in two sections. The second and third column show the answers given at moment 1, the fourth and fifth column show the answers given at moment 2 (section 5.4). The answers given by experts, at both moments, have been characterized positively or negatively. However, at moment 1, experts have only experienced the use of scales and open answers. Therefore their answers about the use of the table are based on assumptions, at moment 1. At moment 2, the experts have experienced the three direct PE methods. Their answers about the use of the scale and open answers formats, under moment 2 (5.4.2), are partly based on assumptions. Besides the assumptions made, the answers are also based on their experience gained during moment 1.

To show how the answers to the semi-structured questions are characterized, two examples are given that relate to the example interview in the previous section:

Example-2: With “Question-1” the expert states: “Answering questions is straightforward”. This is a positive response for the use of the open answers format and results in a positive response for method 2 at moment 1.

Example-1: With “Question-4” the expert states: “I adjust my opinion and think the table is a reliable method.” This is a positive response for the use of the table format as
probability elicitation method and results in a positive response for method 3 at moment 2

Like these examples, all the feedback of the experts is evaluated in 5. However, during the elicitation not all questions asked were explicitly answered by the experts. In some cases the expert did not explicitly answer what they thought about a method. Instead experts indicated advantages and disadvantages they experienced with using the methods. In the latter case the elicitor interpreted the answers, taking into account the context (observations and environment) in which answers were stated.

Table 5: Graphical representation of the answers given

<table>
<thead>
<tr>
<th>PE methods:</th>
<th>Moment 1</th>
<th>Moment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Method 2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Method 3</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 5 presents an overview of all the ranked answers given by the experts. Looking at the row of the open answers PE method, answers differed at moment 1 and 2. Experts indicate after moment 2: “Tables are more in balance. Estimations made as a result of an overview result in a higher internal consistency”. Therefore an overview is seen as an advantage by experts. After the first moment experts have not yet assessed a child node with many different conditional probabilities (combination of the state of two parent nodes). As a result, the workload has been low when answering the questions. At the second moment, experts have assessed many conditional probabilities of one child node. Multiple different probabilities are assessed at once, which makes it possible for experts to make a distinction between the conditional probabilities asked. As a consequence, experts’ adjust their initial estimated probabilities, to make sure that the proportions between the conditional probabilities are in accordance with their feeling. This results in using systematics for filling in the answers. Experts indicate that the systematics used is based on experience. On the other hand, in case the experts have been assessing a limited number of conditional probabilities, the use of the open answer PE method is found to be an accurate translator for experts’ opinions. In case many conditional probabilities are assessed, the table is found to be more adequate than the questions.

The expert’s appreciation about the accuracy of their answers

The accuracy of the answers has been indicated, by the experts, to be subject to variance. Instead of accuracy, reproducibility of the answers is measured. Experts indicated that they are not able to reproduce their answers with any of the PE.

When scales are used by experts both the anchors on the scales and the line between anchors are used to indicate the assessed probability. Although anchors are absolute values, research confirmed that answers marked between anchors convey the vagueness of the opinion the expert holds (Renooij, 2001). However, the markers ’/’ are freely interpretable by the elicitor. As a consequence the marked probabilities are interpreted and quantified by the elicitor, instead of the expert (see figure 14). This makes the accuracy of assessed probabilities less accurate. One expert found it hard to fill in an answer, because his knowledge of the node was limited.
The first response experts give to the use of open answers is that experts indicate that answering a question is based on the first feeling they have. The answer to a question would therefore be closer to the feeling an expert has than with a table format. Although the majority of the experts indicated that they have problems reproducing the answers in any format, one expert stated that he is able to reproduce the answers of the questions exactly the same in any format by making use of the same structured reasoning. After moment 2 (5.4), experts are glad that the tables are not in an open answers or scale format where the relation between the questions would get lost. They also indicated that the use of method 2 leads to inconsistencies if the number of conditional probabilities would increase to the number of conditional probabilities asked during method 3, during moment 2.

With a table the experts were able to compare their answers with probabilities stated for other conditional probabilities. A table presents the overview of all the relevant conditional probabilities, which made experts reconsider their assessed probabilities. This made experts’ move from an initial answer close to their feeling (as with questions and scales) to an answer in proportion to the other answers given. With reconsidering their thoughts, for example:” I am thinking too negatively”, experts analyse their answers and adjusted probabilities until they are satisfied. Experts indicate that the obtained probabilities obtained through the use of tables result in a higher internal consistency and reflect their opinions just as well.

To summarize, experts expressed the accuracy of their answers with the ease in which they are able to reproduce their answers. Reproducibility is indicated as troublesome with all of the formats. As a consequence, the internal consistency became an indicator for the adequacy of the PE methods. Methods 1 and 2 represent the first impression experts hold for conditional probabilities better than method 3. However, in the case many conditional probabilities are assessed, method 3 is found to be a better estimator than method 1 and 2. The overview presented with a table leads to revision of initial estimations to increase the internal consistency.

Which knowledge the experts lack to define probabilities
All experts indicated no lack of knowledge and are capable enough to divide the 100 point scale over the states. This indicates that all participants fall within the category that feels capable of giving direct numerical probability assignments (Spetzler & Staël von Holstein, 1975). However in order to interpret the first table presented, with the keywords used to denote the variables, some practice is needed on how to read the table.

Other observations
Besides the four groups of questions which are used in this section to analyse the gathered data, this section describes an aspect found that does not belong to one of the mentioned groups. From the observations made, experts’ concentration varies with the method they are using. The concentration of the experts can be measured on the basis of two variables (table 6). The rows represent the time (short vs. long) it takes them to obtain an answer. The columns represent the concentration span (low vs. high) necessary to come to an answer. The
concentration span varied with the methods used. With open answers and scales the concentration span is short and high, because experts have to imagine the conditional probability for each question asked. However, the table presented an overview of all the conditional probabilities which made experts see the relation between the questions as a whole. This results in a long and low concentration before all questions are answered. The difference between the scale, open answers and table formats are shown in Table 6.

<table>
<thead>
<tr>
<th>Table 6: Concentration matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration</td>
</tr>
<tr>
<td>Short</td>
</tr>
<tr>
<td>Long</td>
</tr>
</tbody>
</table>

Each the four types of questions have been analysed on the outcomes of questions asked during the elicitation. In the next section the conclusions of the performed analyses are drawn. Beside the conclusions some comments are made that have to be taken into account when performing a PE.

5.7 Case study 1: findings

In this section each of the direct PE methods has been evaluated and conclusions for the first research sub question have been drawn. First the direct PE methods (scale-, open answers- and table format) are evaluated in a consecutive order in the following paragraphs. Secondly, the research sub question is answered and finally some comments about PE have been made. These comments should be taken into account when implementing the found direct PE method.

Evaluation of PE method 1

The majority of the participants found the scale easy to use. Experts primarily used the anchors and one used the line in between. The experts indicated not being able to reproduce their answers. Perhaps this was a reason for anchoring. Renooij (2001a) did not explain how to interpret markers given in between anchors. As a result, the accuracy of assessed probabilities is determined by the anchors provided or rough estimates made by the elicitor who interprets the given answer. Checking for the laws of probability is easy in case one state is asked with a binary variable. However, with tripartite variables precise interpretations of the scales to check for the laws of probability is not found plausible during the elicitation, because of the difficulty in interpreting and summing up results if not on an anchor.

Evaluation of PE method 2

The use of open answers with the opportunity of dividing 100 points is found to be an easy, understandable method. In the case of a limited number of different states, experts found this method to have a close representation of their initial thoughts. The experts are able to indicate that they are working with a matrix, checking their given answers for consistency. In case 27 different conditional probabilities are asked, experts indicate that this is not manageable anymore and an overview between the questions is lost. Important is that experts interpreted every question separately, creating a Stop-Start situation at the end of each question. Physically the same happens. With every question experts visualize the hypothetical situation stated and then answer this question. Hypothetical situations which the experts found very unrealistic are, as a result, found very difficult to answer. The consistency between the questions is checked at the end of each elicited node, or is forgotten.
Evaluation of PE method 3
Experts’ first reaction when confronted with a table is one of misunderstanding. After a short explanation on how to read the table it is found to be straightforward and easy for both the elicitor and the respondent. After experiencing the tables, experts indicated to be glad not to have done this by making use of the open answer format, although their opinion was influenced by the overview presented with the table. The table provoked the experts to be consistent, by relating their estimated probabilities with already estimated probabilities for that node. The table made experts structure their opinion by erasing estimated probabilities until they were satisfied about all the estimated probabilities in the table. The overview of the table worked as an anchor, because difficult conditional probabilities (which the experts had never thought of) are related to situations with which they are more familiar/confident with. This resulted in experts being able to estimate probabilities of less familiar hypothetical situations.

Evaluation of all three PE methods
Where method 3 was evaluated on the basis of method 2, a difference in estimated probabilities appeared. This difference appeared be due to a different interpretation of open answer format compared to the table format (i.e. the questions were interpreted as individual hypothetical situations, whereas the questions with respect to the table format have been presented in the context of the other hypothetical situations).

In the three methods tested, it is found that the elicitation above all must comply with issue 7; unambiguosness. The BN’s used during this case study were not fully unambiguous. As a consequence, many variables and states were interpreted differently and even after explanation by the elicitor, some answers could still not be given. This can be caused by lack of knowledge on this subject. This resulted in experts missing the opportunity to use a specific PE method. Future research should perhaps try to test the PE methods separately from the expertise of the experts questioned, but on a more general topic. On the other hand, the case study made the BN more unambiguous, which served the higher cause, the implementation of the BN for early reliability prediction.

Answer to the first research sub question
In this paragraph the first research sub question will be answered, by making use of the evaluations from the previous paragraphs. The first research sub question is:

*Which probability elicitation method is found most adequate to use in this situation, according to experts?*

All experts are able to cope with all three PE methods to reflect on their opinions about relationships within the BN. All three methods were found straightforward, easy to handle and not difficult to learn, complying with Renooij (2001a) (issue 6: familiarizing expert). However, experts are unable to reproduce their answers with any of the researched methods. At moment 1, few elicitation questions per node, experts indicated that the open answers format gives a good representation of their initial feeling. At moment 2, many elicitation questions per node, the use of the table format has been preferred. Although influenced by the overview presented in a table format, experts expect a higher level of internal consistency with the use of method 3 than with methods 1 and 2. Even though reproducibility was not guaranteed, they still expected a high degree of internal consistency. In spite of an expected variance between the given answers between elicitations, experts indicate that the variance within a table will be less with method 3 than with method 1 and 2, as a result of the structured reasoning used when using the table format.
The BN’s used in this research are already limited to their minimum size. As a result, many conditional probabilities have to be assessed. This results in the selection of a table format for the PE method. Important is that this conclusion cannot be made without some critical notes from both experts and the researcher to prevent biases. The next chapter will investigate the generalization and consistency of the results found during this chapter.

The experts’ comments:
- The use of a table made them reconsider their initial estimates.
- Internal consistency is more important than initially thought.
- Expert felt to be systematic when filling in probabilities.
- To check tables for internal consistency is intensive (checking for trends).

Besides the issues mentioned in chapter 4, the use of a table should also take into account:
(recommendations of researcher)
- Elicitation should be supported by the structure that is used, in this case a table. A table is able to support the expert to check for consistencies and to find out what his opinion is with unlikely situations. However, the expert should not use the table to presume linearity in the answers. Caution has to be taken that it is not the other way around, where the structure determines the elicitation.
- Each business environment needs a customized PE method.
- Tables are formed by putting the variables in a certain order. In case an expert indicates that he orders the states differently than stated, caution should be taken. The structure of the table should not prevent the expert being able to indicate his opinion.
- Experts are concentrated for a longer period of timed, than with the other two methods.
- It is found that the elicitation above all must comply with issue 7; unambiguousness
6 Implementation of the selected probability elicitation method

In this chapter case study 2 is described. This chapter contains 5 sections. In section 6.1 what’s and why’s of case study 2 are described. The next section (6.2) describes the PE processes conducted between elicitor and experts. This is followed by the data gathering in section 6.3. In section 6.4 an analysis is described whether the PE method can be used for all three PCP phases. In section (6.5) the analysis is described of the consistency of the probabilities obtained. In the final section (6.6) research sub questions two and three are answered.

No previous research has been found that tried to obtain all the conditional probabilities for a BN by making use of a table format for the elicitation procedure. In case study 2 the generalization of method 3 is explored by conducting elicitations over all three phases of the PCP (research sub question 2, 3.2). With the obtained results, the consistency of the data set is investigated (research sub question 3, 3.2).

Using some data obtained during case study 1, case study 2 can be seen as the result of the snowball effect as described by Verschuren & Doorewaard (1990), where case study 2 continues with findings about the consistency of the resulting PE method. A table format (method 3) for specifying conditional probabilities is the result of case study 1. Case study 2, discussed in this chapter will explore the results of case study 1.

6.1 The what’s and why’s of case study 2

In this section the what’s and why’s of case study 1 are described keynote wise. Table 7 shows in the left column the what’s (actions) that have been undertaken and a reference to a section. In the right column the why’s (reasons) are shown as well as a reference to a section.

Table 7. The what's and why's of case study 2.

<table>
<thead>
<tr>
<th>WHAT’S (actions)</th>
<th>WHY’S (reasons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Contact the participating experts (4.6)</td>
<td>- To participate in case study 2.and to check their willingness to cooperate (issue 5, 4.2)</td>
</tr>
<tr>
<td>- To prepare for the PE process (6.2):</td>
<td></td>
</tr>
<tr>
<td>- Prepare a training</td>
<td>- To limit the influence of biases (6.2.2)</td>
</tr>
<tr>
<td>- Prepare a PE document</td>
<td>- To create the same base of PE for each expert (6.2.2)</td>
</tr>
<tr>
<td>- Prepare an evaluation</td>
<td>- To approach each expert in the same way to obtain natural behaviour (6.2)</td>
</tr>
<tr>
<td></td>
<td>- To limit the influence of the elicitor (6.2.1)</td>
</tr>
<tr>
<td></td>
<td>- To familiarize experts with method (issue 6, 4.2)</td>
</tr>
<tr>
<td></td>
<td>- Check the obtained data (evaluation of answers section 6.5)</td>
</tr>
</tbody>
</table>
### 6.2 Case study 2: probability elicitation process

The PE process in case study 2 is described in the following section.

The initial steps, such as background, identification and recruitment of experts (Jenkinson, 2005), have already been discussed (4.6) and will not be elaborated on. The PE process in case study 2 is divided as follows (Spezterl & von Holstein, 1975; Garthwaite et al., 2005; Jenkinson, 2005, Bedford et al., 2006):

- An introduction (set up, motivation, structuring, conditioning, probability and assessment training).
- The probability elicitation (encoding).
- Assessing the adequacy (verifying).

The aim of the PE process is (among others) to obtain natural behaviour of the participants. To obtain natural behaviour, the elicitor is as uninvolved as possible. Especially in case studies, it is found that it is difficult to control the reactivity (the many possible subtle influences which influence participant and/or elicitor) and biases of the elicitor (Graziano & Raulin, 2004). As a result, three major actions are taken. This list is kept within arms reach during the elicitation process (Renooij, 2001a). Firstly, a predetermined presentation (training) was given at the start of the elicitation process. Secondly, a PE document has been constructed with which the experts are elicited (appendix H). Finally, the model is made as unambiguous (Issue 7) as possible by providing a list of nodes, states and their definitions (see appendix K).

The duration of the full elicitation process in case study 2 is constrained by a maximum of ninety minutes (4.3 issue 13, duration). Preferably all phases of the elicitation process should be executed within this time span. It is likely within a business environment that an expert is not able to spend ninety minutes in one session. A BN might be too large to be elicited in one go. In both cases, the PE process is split into two slots of sixty minutes. These slots are in total 30 minutes longer (2x60=120 minutes), because the start up (approximately 2x 15 minutes introduction time) by the expert has to be conducted twice.

The elicitation process used in case study 2 addresses problems/issues, identified in literature (among others, chapter 4) and from case study 1 (Chapter 5). As in any case study, the scope of case study 2 is limited.

Jenkinson (2005) identified that cultural difference between countries influences the results of PE. As a consequence, cultural difference due to the expert’s background may have already been introduced. Because experts from the qualitative stage are used, cultural differences may already influence the BN’s. Cultural differences are not taken into account.
Moreover, the influences of individual payoffs (i.e. the individual interest an expert has with certain outcomes of the research) that may be associated with the PE (Spetzler & Staël von Holstein, 1975) are not taken into account either.

This section has identified the elicitation process, the use of training, a document and feedback. Before discussing what the training, the used document and the feedback contain, the role of the people involved in the elicitation process is discussed in the next section.

### 6.2.1 Roles during elicitation

The roles of all participants during the PE process are not always clear (Bedford et al., 2006). The following distinctive roles are found in literature:

- **The participant**: Is the ‘domain/substantive expert’ and has relevant knowledge about the variables and contributes his own assessment.

- **The elicitor**: Is the ‘analyst/statistician’ responsible for writing the assessment, giving probabilistic training, ‘validating’ results and providing feedback.

One person can take up more roles than one (Jenkinson, 2005; Bedford et al, 2006). In case study 2 there are the elicitor (researcher) and the participant (expert).

### 6.2.2 Training

A presentation has been used during the set up phase. Three aspects of training are mentioned by Jenkinson (2005):

1. Probability and probability distributions
2. Information about the most common judgement heuristics and biases, including advice about how to overcome them
3. Practice elicitations, particularly using examples where the “true answer” is known, but unlikely to be known by any of the experts.”

Jenkinson (2005) has described two types of normative understanding. In this research it is assumed that all the experts participating in this research have at least a normative understanding of general numeracy, since they are engineers. The second normative understanding, understanding of probabilities, is not assumed and accordingly incorporated in the training. Because the knowledge of experts on these topics is not known, each expert receives the same training. Training is found to be crucial for engineers, because it shows that PE is scientific and gives them an understanding of the biases involved (Jenkinson, 2005). As a result an example and an exercise are given to educate (if necessary) and inform them of biases. Due to the time constraints not all the biases and preventions are discussed, but the elicitor takes note of what happens and informs the respondent in case biases are noticed. A practical example and exercise have been conducted to comply with the third aspect. However instead of a “true answer” the experts were asked to explicitly elicit a custom made elicitation to gain insight in how experts thought and what they are taking into account.

The elicitation has been conducted on the basis of a presentation and exercise. The presentation contained the following topics:

1. General knowledge on probabilities
2. Explanation on how to read the BN
3. Attention to the definitions of each node
4. Explanation on how to read the tables
5. An example on how to use the table and what kind of biases play a role
6. An elicitation exercise (in which insight was tested and more biases were explained)

Appendix L contains the slides used for the presentation. Although not all topics are mentioned in the slides, topics 3 and 4 are verbally explained in the consecutive order as presented. The next section will discuss the elicitation document that is made.

6.2.3 Probability elicitation document

Three PE documents are used (one for every PCP phase and its BN). The documents used contain two parts. The first is a list of all the nodes, states and their definitions (appendix K). The second part contains an introduction and the actual PE document (appendix H). During the introduction the expert is informed on the importance and purpose of the research (Spetzler & Staël von Holstein, 1975). The PE consists of a step by step procedure for the elicitation of the BN. Elicitation takes place per variable that influenced the ‘reliability node’, from outside (parent nodes) inwards (child nodes). In this way experts’ reason from cause to effect, this causes less inconvenience (Nadkarni & Shenoy, 2004). As argued in chapter 5, each section contains a picture of the BN and a sample question of the tables. The first table elicited also had an explicit explanation on how to interpret the table, to answer the need for explanation from practise. The final section looks at the influence on reliability during the particular phase of all elicited nodes.

At least one condition that has to be met is that the document is found to be unambiguous, both confirmed in literature (issue 7) and practise. As a result, unambiguousness is tested by making use of a couple of iterations of the elicitation process with a test panel (issue 4; selection of experts) (section 4.2). An unambiguousness document serves both the unambiguous interpretation of the experts and limits the influence of the elicitor (Graziano & Raulin, 2004)

Each expert receives his own elicitation document and scratch paper, making it possible for experts to do whatever they feel is needed to become familiar with the document and the assessed probabilities. Occurrences are, for example, making notes, erasing, doodling, scratching and comparing pages side-by-side (Renooij, 2001a). Experts also indicated that this made it easier for them to check the tables for consistencies. One thing that helped was to leave the document unstapled.

6.2.4 Evaluation of estimated probabilities

After the document has been used for PE purposes, an evaluation takes place. The evaluation takes into account evaluation of obtained probabilities (issue 15) and evaluation of the method used (issue 16). These consistency issues (15,16) are therefore described under section 6.5.

6.3 Case study 2: data gathering

In this section the PE process is executed for all three BN’s (i.e. for each PCP phase, CD, DD, and MI). The data gathering has been performed under 13 experts for the second case study. The number of experts per PCP phase is as follows:
- 5 experts for the CD phase
- 3 experts for the DD phase
- 5 experts for the MI phase
The PE process has been conducted with each of the experts. Some experts were able to finish the entire PE document, whereas other experts could only finish a part of the PE document. In table 8 an overview of the number of experts for each PCP phase is presented. The left column shows the PCP phases (CD, DD, MI). The centre column shows if BN were fully or partly elicited. The right column represents the number of experts for each row.

<table>
<thead>
<tr>
<th>BN</th>
<th>Fully/partly</th>
<th>Nr. of experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>Fully</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Partly</td>
<td>0</td>
</tr>
<tr>
<td>DD</td>
<td>Fully</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Partly</td>
<td>1</td>
</tr>
<tr>
<td>MI</td>
<td>Fully</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Partly</td>
<td>2</td>
</tr>
</tbody>
</table>

The obtained probabilities for each BN are available for further research, but have not been added to the appendix, due to the size of the documents and the fact that the analyzing quantitative data is not the objective of the research sub questions.

### 6.4 Case study 2: generalization

The second sub question of this thesis is discussed. At stake is whether the PE method can be used for three PCP phases in terms of acceptance of the method by the participating experts.

Evaluation with the participating experts produces the following results. Two experts were only familiar with a small part of the model, because they were specialized in a single node of the BN involved. One expert filled in one table and the parent nodes and refused to fill in the rest, because he thought that it can all be calculated on the basis of the input given. So he disagreed with the PE method used. On the other hand, ten experts were able to fill out the complete PE document, making use of the provided information and PE method.

To conclude, 12 out of 13 experts are able to fill out the PE document without encountering any major difficulties. Even after evaluation of the obtained probabilities, experts indicated satisfaction with their own result. As a consequence, it can be stated that the PE procedure is accepted by almost all (12/13) of the participating experts. The experts were able to cope with the used PE method. As a result, the PE method can be generalized for all three phases of the PCP.

### 6.5 Case study 2: consistency

The third sub question of this thesis is discussed in this section. Objects of the discussion are three types of consistency:

- Internal consistency.
- Consistency over time.
- Mutual consistency.

#### 6.5.1 Internal consistency

Internal consistency is defined as the precision as well as the accuracy of the obtained probabilities.
**Precision and accuracy**

The evaluation of the obtained probabilities serves two purposes. The first purpose is to determine the precision of the obtained probabilities. The second purpose is to determine the accuracy (Cooke (1991); Apostolakis (1990); Cooke & Goossens, 2008). Apostolakis (1990) refers to these two purposes as respectively ‘relative frequency’, i.e. looking at probabilities on the basis of the number of identical trials, and ‘degree of belief’, i.e. act ‘f’ is preferred over act ‘g’. The latter can be seen as if the probabilities fall within the boundaries set by the expert.

In table 9 the dots are representing the estimated probabilities by the experts. On the two axes two terms are shown: precision and accuracy. Precision can be seen as the variance (i.e. distance) between two dots. The accuracy can be seen as the distance between each dot and the centre of each circle. The smaller the variance between the dots represents a higher precision of the expert’s estimated probabilities. The more the dots are near the centre the higher the accuracy of the expert’s estimated probabilities is.

### Table 9: Matrix comparing Accuracy & Precision

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Precision can be seen from two perspectives. The first perspective is by obtaining two values of the same expert over time. The second perspective is by obtaining one value of the expert, and comparing it to a reference point to determine the precision and accuracy. Due to the lack of objective data it is not possible to make use of a reference point, a so called seed variable (Cooke, 1991). In the case of available seed variables, these could be used to weigh each expert’s estimated probabilities compared with (objective) seed variables. As a result, it would be possible to determine how precise the expert’s estimated probabilities are.

The precision of experts estimated probabilities cannot be determined in case study 2, because there are no available seed variables. Therefore the focus in this section is on the accuracy of the estimated probabilities.

Accuracy can be described by calibrating the difference between the initial estimated probability by the expert and the centre of the circle (table 9) over time. The circle represents the belief of the expert. To relate this to table 9: if the obtained probability (dot) is within or outside the boundaries (circles).

**Trying to determine accuracy by using ranking**

The use of experts indicates a level of expertise on a topic other people do not have (Issue 1; who’s the expert). From an expert it is assumed that if this expert is presented with multiple variables the ranking of these variables, from important to less important, according to the belief he holds, is accurate.

After the probabilities have been elicited from the experts, sensitivity analyses rank the influence of the parent nodes on the selected sensitivity nodes (i.e. of the investigated child nodes) (Renooij, 2001a).
The outcomes (i.e. the ranked order of parent nodes) of a sensitivity analyses can be represented in tornado graphs\(^2\). Figure 155 shows two tornado graphs that belong to the node ‘Quality of tests during product assembly’, of one expert.

![Tornado Graphs](image)

**Figure 15. Two tornado graphs for the node ‘Quality of tests during product assembly’ in respectively the states ‘low’ (left) and ‘nominal’ (right).**

As figure 15 shows one node in two states (‘low’ left and ‘nominal’ in the right graph), the tornado graphs have two different rankings (from top to bottom). To determine the accuracy by making use of ranking, each conditional probability should be checked, because different rankings are possible for each state of the child node. To determine the accuracy for the whole BN by ranking would be extremely time consuming.

**Determining the accuracy through process feedback**

The initial probability as stated by the expert is used as reference point. Kynn (2008) proposes to offer process feedback, for example by making use of different representations of probabilities (say graphical). This could be done to give a summary and allow experts to reconsider probability estimates or confirm the stated probability estimates. This assumption is used to evaluate the answers given by experts on accuracy, assuming that an expert can confirm or disapprove with a ranking of variables. It is advisable to give process feedback during the same elicitation session to prevent the results being influenced by the heuristic ‘availability’ (section 4.1).

The use of note probability tables (NPT’s) serves as an advantage. An NPT shows an overview of all estimated probabilities per node per expert during the PE process. Thus the expert is facilitated to check his own estimated probabilities for consistency during the PE process.

Experts make use of different ways to fill in the NPT (observations 5.5). Experts are able to interpret the NPT in the way it is stated (filling out the highest state of each variable ranked at the top). They might fill out first the maximum and minimal values. Both ways lead to biases. First overconfidence can occur, another bias is linearity. Experts may falsely assume linearity (e.g. from highest to lowest state) (issue 16), because conditional variables change in a predictive way in the table. It is possible that experts do not check their estimated probabilities. Thus the biases of linearity cannot totally be prevented. Experts are therefore

---

asked to explicitly give reasons for their assuming of linearity and/or for their overconfidence (i.e. process feedback for linearity or overconfidence).

<table>
<thead>
<tr>
<th>Elicitation by expert</th>
<th>Processing of obtained estimated probabilities by elicitor</th>
<th>Constructing of bar charts by elicitor</th>
<th>Giving process feedback to expert by elicitor</th>
</tr>
</thead>
</table>

Figure 16. Process feedback flowchart.

The flowchart from figure 16 shows the actions the elicitor has to undertake to give process feedback to the expert in case study 2. The obtained estimated probabilities, represented in the NPT’s, are processed from the PE document by entering the obtained data in a computer programme (Microsoft Office Excel 2003©). The correctness of the estimated probabilities as well as the correctness of the actual entering of data is double checked, by means of the laws of probability (discussed in 5.1.3) by the computer programme. The digital data set has been used to convert NPT’s of all nodes for each BN, into bar charts (figure 17) for each individual relationship represented by the BN in approximately ten minutes. The bar charts are used by the elicitor to give process feedback to the expert.

Process feedback is used, in case study 2, during a small time interval. Process feedback with numbers would lead to testing the expert’s memories on estimated probabilities (availability, section 4.1). The goal is to test the accuracy of the estimated probabilities. Therefore bar charts are used which contain no numbers. The mutual ranking and order of magnitude of the estimated probabilities are shown in the bar charts. By this graphical means, experts are facilitated to reason again for their estimated probabilities, whilst the use of numbers is avoided.

Each individual relationship has to be represented, because it is difficult for experts to reason the effect (of multiple variables changing at the same time) on their estimated probabilities. The bar charts represent what the effect is on the child node in case all parent nodes are fixed, except one (i.e. in figure 17 the supplier quality is fixed in the state “partly insight ...”). The non fixed parent node (i.e. “managing supply chain”) varies over all its states, showing the impact of the estimated probabilities for one child node in a bar chart (Figure 17).

Figure 17: Situation with part of the BN and bar chart that is evaluated
During process feedback each node in the BN is addressed. From each node the most unlikely situation(s) is/are checked with the expert. Unlikely situations can be marked as situations that are not expected by the elicitor. An example is given in figure 18, where the two highest states are rated the same. In this way not each conditional probability has to be checked, but checks are performed at random to check whether the expert agrees with earlier assessed probabilities. Compared with the sensitivity analyses (tornado diagrams) this saves time and is systematic/understandable for the expert.

A critical note on the use of the computer programme for this kind of process feedback is that all the assessed probabilities had to be entered manually. With future research it is suggested to use a computer programme for the PE process instead of a hardcopy PE document.

![Figure 18: Unlikely conditional probability for a bar chart, during evaluation](image)

Experts gave as argumentation on Figure 18 that it is not about doing things and documenting, or doing things and not documenting. Little difference can be made between these two. The main concern is if problems are dealt with. In case of problems, reliability is automatically taken into account. As a consequence, this condition is high.

By making use of evaluation through the use of bar charts, the experts are found to be internally consistent with the probabilities assessed. Process feedback is used to evaluate the ranking of conditional probabilities. Although a couple of experts indicated some changes in ranking, the majority agrees with the probabilities assessed by making use of the PE process. The next section will investigate if the results are also consistent over time.

6.5.2 Consistency over time

The consistency over time is approached by looking at the repeatability, reproducibility and stability as described (in section 4.4 issue 16). To be able to make any comments on this part, the elicitation has to be performed at least two times with each expert to be able to compare obtained probabilities. Complicating factors that can influence the results should be taken
into account. Firstly, during section 5.6 experts mentioned that they are unable to reproduce their answers with any of the researched PE methods. Secondly, when the experts are approached for repeatability and reproducibility experts should be in the same position as the first time, in other words a ‘blank opinion’, to be able to compare results. In the case PE’s are conducted within a small time interval, there is the possibility that an expert’s memory is tested instead of the opinion he holds on a conditional probability. This would make the results vulnerable for the heuristic ‘availability’ (Section 4.1). What is sought for is that the expert reasons the conditional probabilities again in exactly the same state of mind he did the first time. Expert’s knowledge is based on experience, which is dependent on time. As time moves on, the experience of the expert changes, which results in a possible change of opinion. Thus repeatability, reproducibility and stability are not very likely.

To conclude, consistency over time has not been researched in case study 2 for obvious reasons, mentioned in the paragraph above.

### 6.5.3 Mutual consistency

Comparison between expert’s estimated probabilities is an indication for how much experts agree with each other. This section shows that the mutual consistency, found in case study 2, seems to be low. The reasons why will be discussed.

Figure 19 (on the next page) shows examples of estimated probabilities for all three PCP phases. The Y-axis represents a probability scale. The X-axis shows three states of a child node. The left side of the next page shows mutual consistent estimated probabilities in three states of three child nodes (top to bottom PCP phases CD, DD, MI). The right side of the next page shows mutually inconsistent estimated probabilities in three states of three child nodes. Each range of estimated probability is represented by a vertical line between two points: the maximum and the minimum value for the estimated probability. A large range suggests low mutual consistency and a small range suggests a higher mutual consistency.

Figure 19 shows that there are large differences (i.e. the variance between the answers indicated with a red line) between the experts over all three phases of the PCP. Within one child node both mutually consistent as well as mutually inconsistent answers can be found. Jenkinson (2005) states that large mutual inconsistencies point out that these estimated conditional probabilities are subject to uncertainty. A cause for these differences has not been researched in case study 2. Due to the large variety in mutual consistency throughout the NPT’s nothing can be said about the mutual consistency of the PE method used.

### 6.6 Case study 2: Findings

This research has investigated the following two research questions:

2. *Can the probability elicitation method from sub-question 1 be generalized?*

Twelve out of thirteen participating experts are able to conduct the probability elicitation process without encountering any major difficulties. Even after evaluation of the obtained probabilities, experts indicated satisfaction with their own result. It can be stated that the probability elicitation procedure is accepted by the majority (12/13) of the participating experts. The experts are able to cope with the used probability elicitation method developed. As a result, the probability elicitation method can be generalized over all three phases of the PCP.
Figure 19: Ideal situations with mutual consistency (left) against mutual inconsistencies (right). CD at top, then DD, MI bottom.
3. Is the probability elicitation method from sub-question 1 internally consistent, consistent over time and mutually consistent?

The conclusion from case study 2 is that no proof has been obtained for the overall consistency of direct PE methods with table formats (method 3).

- Internal consistency: By making use of evaluation through the use of bar charts, the experts are found to be internally consistent (in terms of accuracy) with the probabilities assessed. Process feedback is used to evaluate the ranking of conditional probabilities. Although a couple of experts indicated some changes in ranking, the majority agree with the probabilities assessed by making use of the probability elicitation process.

- Consistency over time: Case study 2 contains no research on this issue. The complicating factors of stability in the answers that play a role with the repeatability and reproducibility of elicitations is the reason. Instead the elicitation has been performed with multiple experts to gather a data set which can be used as reference point for future research.

- Mutual consistency: Case study 2 shows (6.5.3) that there are large differences between the experts over all three phases of the PCP. Within one child node both mutually consistent as well as mutually inconsistent answers can be found.

The next chapter will answer the main research question.
7 Conclusions, nuances and recommendations

In this chapter the conclusions of this research are drawn and discussed. This chapter contains four sections. In section 7.1 the answers for each research question are given with a brief underpinning. Given answers are nuanced by three statements in section 7.2. In the next section (7.3) recommendations are presented. In the final section 7.4 the contributions of this thesis are discussed.

7.1 Conclusions

In this section the main- and sub research questions are answered. First an answer is given for the main research question:

“Which probability elicitation method can be used in this research environment to quantify the node probability tables on the basis of experts’ opinions only?”

Direct Probability Elicitations (PE) method with a table format is accepted by the experts. This can be concluded from both case study 1 and case study 2.

1. Which probability elicitation method is found most adequate to use in this research environment, according to experts?

Direct PE methods with table formats are the most adequate to use in this research environment, according to experts (5.7). From the literature study can be concluded that direct PE methods are the most adequate in the given research environment (5.1). Case study 1 concludes that Direct PE methods with table formats are the most adequate to use.

2. Can the resulting probability elicitation method be generalized?

Direct PE methods with table formats can be generalized. With case study 2 the direct PE method is tested for all three PCP phases. 12 out of 13 experts were able to cope with both the PE process as well as the PE method (6.4). The table format is accepted over all three PCP phases. As a result, the table format of the direct PE method can be generalized.

3. Is the resulting probability elicitation method internally consistent, consistent over time and mutually consistent?

The conclusion from case study 2 is that no proof has been obtained for the overall consistency of direct PE methods with table formats (method 3).

7.2 Nuance

In this section the answers of the previous section are discussed and nuanced. To nuance the conclusions of this thesis, the following remarks can be made:

1. Due to the ambiguity of the terms “most adequate”, “generalized”, “consistency”, the conclusions could be disputed.
2. Acceptability by the experts of the direct PE method chosen could be better measured.
3. The conclusion that consistency has not been reached by the resulting direct PE method could be more elaborately underpinned.

7.3 Limitations & Recommendations

The performed research is limited by a very specialized area of implementing a Bayesian network in practise. As a result a number of concessions had to be made during this research to be able to say something about a small part, instead of nothing about a large area. A list of
The most important limitations is given as well as recommendations for future research. In table 10 each limitation (mentioned in the left column) is linked to possible future research (mentioned in the right column).

**Table 10. Limitations (left) and recommendations (right).**

<table>
<thead>
<tr>
<th>Limitations</th>
<th>Future Research/Recommendations</th>
</tr>
</thead>
</table>
| 1. The applicability of the PE method has been tested on a limited scale. No seed variables were available and no research over time was possible. However a data set has been obtained which can be further analysed. | 1. The obtained data set is made available for future research.  
- The data set could be tested on consistency over time as well as on mutual consistency. |
| 2. This research made use of individual elicitations.                                           | 2. Jenkinson (2005) proposes three elicitation methods based on individual elicitations. Two of the group elicitations can still be performed based on the obtained data to obtain consensus (obtaining mutual consistencies). |
| 3. No historical data/ seed variables have been identified that could be used (section 6.5).   | 3. Internal consistency (in terms of precision) could be investigated by finding an alternative and/or obtaining historical data/seed variables to check estimated probabilities. |
| 4. A limited number of issues has been identified in chapter 4 and taken into account.        | 4. More issues can be identified, and/or other solutions may be sought for issues identified during this research. |
| 5. Only direct elicitation methods have been selected in this research for investigation. Even though this is not recommended by literature. | 5. Using indirect/parametric elicitation methods currently are very labour-intensive. If time and/or the work involved can be minimized these methods can also provide possible solutions. |
| 6. The selection of experts has been based on the function descriptions within PH.  
- During the research it was discovered that the function in relation to other colleagues and seniority are also of importance. | 6. Future research can focus on the influence of selecting experts on their influence within a project team.  
E.g. figure 20, where the CD-phase and the relation of the experts amongst each other are shown. |
7.4 Contributions

**Contribution to scientific research**
The main contribution of this thesis to scientific research is the investigation towards the use of a direct PE method with table format for the elicitation of NPT’s on the basis of experts’ opinions. The implementation of the direct PE method resulted in an extensive data set, for all three PCP phases, which can be used for future research. The internal consistency (in terms of accuracy) can be checked by using bar charts (section 6.5).

**Contribution to Philips Healthcare**
By analysing the extensive quantitative data set obtained in the case studies 1 and 2, an initial invalidated prior insight could be gathered. This prior insight could give an indication of how variables (nodes) influence the reliability for each BN.
References


List of appendices

The appendices can be found in a separate document with the title: “Appendices to Master thesis “An adequate method to quantify probabilities “”

Appendix A: Different approaches for constructing a Bayesian network
Appendix B: Philips Allura Xper
Appendix C: Bayesian networks of the product creation process
Appendix D: Participating experts at Philips Healthcare
Appendix E: Probability elicitation document for probability elicitation method selection
Appendix F: Definitions used during the probability elicitation method selection
Appendix G: Answers from case study 1
Appendix H: Probability elicitation document for the probability elicitation process
  Appendix H.1: Concept and definition phase
  Appendix H.2: Design and development phase
  Appendix H.3: Manufacturing and installation phase
Appendix K: Definitions used during the probability elicitation process
Appendix L: Training --slides--