

Using sensitivity analysis for efficient quantification of a belief network*

Veerle M.H. Coupé¹, Niels Peek², Jaap Ottenkamp³, J. Dik F. Habbema¹

¹ Center for Clinical Decision Sciences,
Department of Public Health, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
e-mail: coupe@mgz.fgg.eur.nl

² Department of Computer Science, Utrecht University
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands

³ Department of Paediatric Cardiology,
Leiden University Medical Center
P.O. Box 9600, 2300 RC Leiden, The Netherlands

Abstract

Sensitivity analysis is a method to investigate the effects of varying a model's parameters on its predictions. It was recently suggested as a suitable means to facilitate quantifying the joint probability distribution of a Bayesian belief network. This article presents practical experience with performing sensitivity analyses on a belief network in the field of medical prognosis and treatment planning. Three network quantifications with different levels of informedness were constructed. Two poorly-informed quantifications were improved by replacing the most influential parameters with the corresponding parameter estimates from the well-informed network quantification; these influential parameters were found by performing one-way sensitivity analyses. Subsequently, the results of the replacements were investigated by comparing network predictions. It was found that it may be sufficient to gather a limited number of highly-informed network parameters to obtain a satisfying network quantification. It is therefore concluded that sensitivity analysis can be used to improve the efficiency of quantifying a belief network.

Keywords: Belief networks; Quantification; Sensitivity analysis; Refinement

*Appeared in: *Artificial Intelligence in Medicine* 17:223–247, 1999.

1 Introduction

The framework of *Bayesian belief networks* was introduced in the late 1980s [14] for reasoning with uncertainty in a mathematically correct manner. It owes much of its popularity to the use of probability theory combined with an appealing graphical representation of conditional independence relations. As such, belief networks allow for explicit and declarative modelling of a problem domain, capturing domain knowledge that is relevant for solving knowledge-intensive problems [1, 7].

In building a belief network, two closely related tasks can be discerned; the construction of the graphical part of the network, and its subsequent quantification. Building the graphical part of a network consists of identifying variables in the domain under study, and assessing the conditional independence relations that exist between these variables; these relations are represented by an acyclic, directed graph. The quantification of a belief network amounts to assessing a local conditional probability distribution for each variable in the network. These local distributions uniquely define a joint probability distribution on the variables discerned, that respects the independence relations portrayed by the graph. A variable's distribution is conditioned on its parents in the graph; the number of parameters (conditional probabilities) that need to be assessed for the distribution grows exponentially in the number of parents. The total number of parameters that is needed to quantify a belief network may therefore be considerable. Furthermore, a substantial number of network variables may be hidden from direct observation; it is then very difficult, if not impossible, to collect quantitative data on these variables. For these reasons, quantification of a belief network is often considered a harsh task.

To facilitate the quantification of belief networks several methods have been proposed in the literature [8, 11, 4]. Recently, Coupé et al. [2] have described how *sensitivity analysis* can be used to reduce the quantification effort. Sensitivity analysis is a method to investigate the effects of varying a model's parameters on its predictions. For a belief network, it can reveal which parameters have a large effect on posterior probabilities, and, therefore, on which parameters the quantification effort should be focused. In [2], the authors propose a procedure of iteratively performing sensitivity analyses of an initially roughly quantified network, in order to stepwise refine the quantification.

This article presents an empirical investigation regarding the viability of this procedure. As a case study, we selected a belief network that describes the pathophysiology of ventricular septal defect (VSD), a frequently occurring congenital cardiac anomaly. It was developed as part of a larger decision-theoretic application for treatment planning and prognosis in the field of paediatric cardiology [16]. For the quantification, we have obtained subjective probability estimates. The use of subjective probabilities is indispensable in domains where there is a shortage of clinical data and many variables cannot be measured. Unfortunately, this may require a massive amount of probabilities to be estimated by field experts, a difficult and time-consuming task. The main objective of our research is to establish whether it is possible to reduce the number of parameters that have to be estimated by field experts. That is, we want to obtain a network quantification that gives predictions comparable to an expert-quantified network, without having to elicit all the network's parameters from the expert.

The following procedure was used in the investigation. Three network quantifications were obtained, differing in the level of *informedness* of the estimates. The term informedness refers to the knowledge about the problem domain of the person supplying the estimates; we assume that accuracy of network predictions increases with the level of informedness.

Extensive sensitivity analyses were performed on all three network quantifications, yielding a set of most influential parameters. In the two least informed quantifications, the estimates for these influential parameters were replaced, stepwise, with the estimates in the quantification of the field expert. The predictions of the resulting improved quantifications are compared with the predictions of a network that was completely quantified by the field expert.

Our results show that the procedure contributes to efficient quantification of a belief network: if, in a poorly-informed quantification, a limited number of highly influential parameters are replaced by more precise estimates, then the network gives predictions that are comparable to the network that is completely quantified with precise estimates. This means that we can avoid lengthy elicitation procedures, and focus the quantification effort on parameters to which network predictions are found to be most sensitive when varied.

The article is organised as follows. In Section 2, we briefly discuss the problem of treatment planning for patients with a ventricular septal defect, and present the qualitative part of the VSD network. Section 3 gives formal backgrounds of sensitivity analysis in belief networks, and describes the method of investigation to test the refinement procedure. Then, in Section 4, the results of the sensitivity analyses and subsequent refinements of the network quantification are presented. Discussion and conclusions are given in Section 5.

2 The VSD network

Ventricular septum defect (VSD) is the most frequently occurring congenital heart disease; approximately 2 to 3 out of each 1000 infants is born with this cardiac anomaly. It is a relatively well-understood disorder with many clinical features that are characteristic for congenital heart disease in general. We are currently developing an application to support the management of VSD patients in clinical practice, based on recently developed techniques from uncertainty reasoning and decision theory. Our aim is to deliver a ‘white-box’ system, in which the user can perceive what is going on, and can interact by proposing alternatives or adjust admissible treatment plans [15]. The core of the system is formed by a sequence of Bayesian belief networks that model VSD pathophysiology and its clinical findings at different stages of infant development. Heart, lungs and vessels are subject to a number of changes during the first years of life, which make it impractical to use the same belief network for each development stage. Instead, we chose to employ different belief networks for different stages; there is, of course, substantial overlap between these networks.

In this article, we focus on the belief network that models VSD pathophysiology for patients aged 3 to 6 months. After a VSD has been diagnosed (usually in the first weeks of life), the patient is monitored during the subsequent months. The age of 3 to 6 months is crucial from a clinical perspective, as it provides the first opportunity to establish the severity of disease. It is therefore the point in time where the clinician will want to assess a preliminary prognosis of the patient’s further development, and may already want to decide upon the treatment plan that is to be followed.

This section describes the *qualitative part* of the belief network for VSD patients aged 3–6 months. We take this part to comprise the directed graph that models conditional independence relations between domain variables, and various types of constraint on the probability distribution modelled by the network. Before we elaborate on the network itself, we first briefly review the domain under consideration.

2.1 VSD

A VSD is an abnormal opening in the ventricular septum, the fibromuscular wall that separates the heart's two ventricles [5]. The main pathophysiological consequence of the presence of a VSD is blood flow ("shunt") from the left to the right ventricle due to ventricular pressure differences. Left-to-right ventricular shunting renders typical murmurs that can be heard by auscultation of the heart, and abnormal vibrations of the heart (called a "thrill") that can be felt at the chest. The *shunt size*, i.e., the amount of blood flowing through the defect, depends primarily on the size of the defect and the pulmonary vascular resistance. The consequence of shunting is that oxygenous blood is recirculated through the lungs. As a result, pulmonary vascular pressure will rise, and systemic cardiac output will decrease. With large defects, the large shunt size and high pulmonary arterial pressure may lead to *heart failure*: the heart is unable to adequately fulfil its primary function, the circulation of blood through the body. Heart failure accounts for most of the typical symptoms associated with VSDs, such as shortness of breath, feeding problems, oedema, and growth arrearage. Severe heart failure may result in cardiomegaly (enlarged heart), hepatomegaly (enlarged liver), and pulmonary infections.

About 70% of all VSDs close spontaneously by normal tissue growth, [9], where small defects are more likely to close spontaneously than large ones. This development may take several months or even years, but it precludes the need for surgical intervention. Unfortunately, continual pulmonary overflow and hypertension may cause severe, irreversible damage to the pulmonary arterioles; this is termed *Eisenmenger's reaction*, and represents the primary risk to VSD patients. Eisenmenger's reaction is detected at early stages by considering the ratio of pulmonary and systemic vascular resistances; increasing pulmonary vascular resistance is indicative for the reaction. However, there exist no means to measure this resistance in clinical practice; it can only be estimated from related signs. For this reason, early surgical intervention is recommended for patients with large VSDs that are unlikely to close spontaneously. The majority of patients with timely repair of uncomplicated VSDs in infancy or early childhood have an excellent result with no clinical signs or symptoms and apparently normal life-expectancy [10].

For the clinician, the main problem is to decide if and when to submit a patient to surgery. Usually, the patient's condition is monitored without surgical intervention during the first year of life. During this period, non-invasive diagnostic tests such as auscultation and echocardiography are conducted repeatedly to gain insight into the shunt size and the pulmonary vascular resistance. When necessary, medical treatment is given to reduce heart failure. After the first year of life, the risks associated with surgical intervention have dropped, and a decision whether surgery is necessary is made. In cases of doubt concerning the state of the pulmonary arterioles, cardiac catheterisation or pulmonary biopsy may be performed prior to that decision to obtain more information on the severity of disease. Therapy is considered completed after closure of the defect, either spontaneously or by surgical intervention.

2.2 Qualitative part of the VSD network

A Bayesian belief network represents a joint probability distribution over a set of discrete, stochastic variables. Each node in the graph represents one of the domain variables, and the absence of an arc between two nodes means that there exists no direct dependency between the variables represented by these nodes; the variables are only dependent via intermediate

nodes in the graph. In practice, the most commonly used heuristic to assess a network's structure is the concept of *causality*: an arc is drawn between two nodes if there is known to exist a direct, causal relation between the corresponding variables. This often provides a sound overall representation of probabilistic dependencies in the problem domain.

For the VSD application, a network structure was hand-crafted with the aid of a field expert, a senior paediatric cardiologist. First, a set of variables jointly describing the VSD domain was selected, and then the graphical part of the belief network was assessed, using causality as the principle modelling heuristic. For a more elaborate description of this development process, we refer to [16]. The resulting network is depicted in Figure 1.

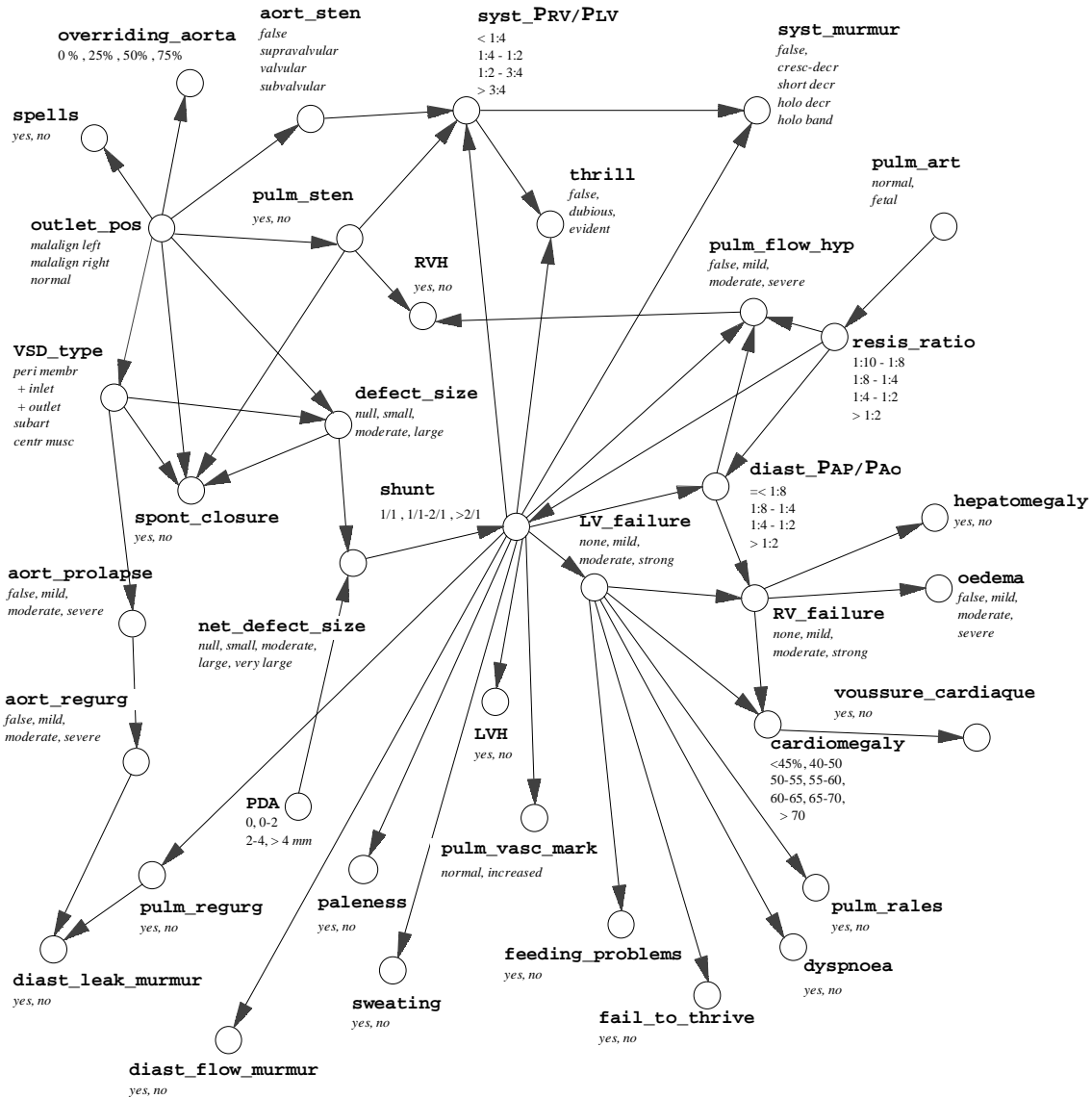


Figure 1: The VSD network.

Once the structure of a belief network has been established, the next task is to estimate the parameters for the quantitative part of the network; this part encodes a joint probability distribution over the network’s variables. For the VSD network, which consists of 38 variables, 1298 parameters are needed to fully quantify the probability distribution. A significant number of variables cannot be observed in clinical practice, making it very difficult to obtain estimates of the required parameters. In order to alleviate the quantification burden for the VSD network, we decided to collect three types of qualitative information on the probability distributions prior to the precise estimation of parameters:

1. *functional dependencies*, modelling deterministic relations between variables,
2. *consistency constraints*, describing which combinations of values cannot occur, and
3. *qualitative influences*, expressing the sign of probabilistic interactions.

Each of these information types can be interpreted as expressing a constraint on one of the local probability distributions in the network [4]. Below, we elaborate on each of them.

The first type of qualitative information consists of *functional dependencies*. These dependencies effectively model deterministic relations, as they express that a variable will necessarily take a certain value when its conditioning variables (i.e., its parent nodes in the graph) have a particular configuration of values. For example, in the VSD domain, we know that if there is severe aortic regurgitation (i.e., the aortic valve is very leaky), then cardiac auscultation will definitely reveal a typical kind of murmur (called a *leak murmur*) during the diastole, the phase of relaxation in the cardiac cycle. Formally, we have that

$$p(\text{diast_leak_murmur} = \textit{yes} \mid \text{aort_regurg} = \textit{severe}, \text{pulm_regurg} = x) = 1,$$

for any legal value of x for `pulm_regurg`, the second conditioning variable of `diast_leak_murmur`. Furthermore, the probabilities of other values for `diast_leak_murmur` (in this case only the value *no*) are 0 under these conditions. For the VSD network, 8 variables were completely described by functional dependencies; 10 additional variables were partially described by them. These dependencies, provided by the field expert, were all assumed to be reliable.

The second type of qualitative information on probability distributions consists of *consistency constraints*. These constraints exclude certain combinations of values that cannot possibly occur in practice. For instance, once the pulmonary arterioles have reached their normal state (6 to 12 weeks after birth), then the pulmonary vascular resistance cannot exceed 1/4 of the systemic vascular resistance. Formally:

$$p(\text{resis_ratio} > 1 : 4 \mid \text{pulm_art} = \textit{normal}) = 0. \tag{1}$$

For the VSD network, 102 consistency constraints were found, pertaining to 10 different variables. Overall, 560 parameters were determined by a functional dependency or a consistency constraint, reducing the number of parameters that need estimation to 738.

The third and last type of qualitative information on probability distributions consists of *qualitative influences* [19]. A qualitative influence is a symmetric property describing the sign of probabilistic interaction between two variables, building on orderings of these variables’ value domains. A positive (negative) qualitative influence, indicated by attaching the label ‘+’ (‘-’) to an arc in the graph, expresses that higher values of the one variable makes higher (lower) values of the other more likely, and vice versa. For instance, if the amount of

blood that flows through the VSD (the shunt size) grows, then increasing failure of the left ventricle becomes more likely. This is expressed by a positive qualitative influence between the variables `shunt` and `LV_failure`:

$$\text{shunt} \xrightarrow{+} \text{LV_failure} \quad (2)$$

which induces the following inequalities:

$$\begin{aligned} & \text{p}(\text{LV_failure} = \textit{none} \mid \text{shunt} \geq 2 : 1) \leq \\ & \quad \text{p}(\text{LV_failure} = \textit{none} \mid 1 : 1 < \text{shunt} < 2 : 1) \\ & \text{p}(\text{LV_failure} = \textit{moderate} \mid \text{shunt} \geq 2 : 1) \\ & + \text{p}(\text{LV_failure} = \textit{severe} \mid \text{shunt} \geq 2 : 1) \geq \\ & \quad \text{p}(\text{LV_failure} = \textit{moderate} \mid 1 : 1 < \text{shunt} < 2 : 1) \\ & \quad + \text{p}(\text{LV_failure} = \textit{severe} \mid 1 : 1 < \text{shunt} < 2 : 1) \\ & \text{p}(\text{LV_failure} = \textit{severe} \mid \text{shunt} \geq 2 : 1) \geq \\ & \quad \text{p}(\text{LV_failure} = \textit{severe} \mid 1 : 1 < \text{shunt} < 2 : 1) . \end{aligned} \quad (3)$$

These inequalities constrain the local probability distribution for the variable `LV_failure`. In co-operation with the field expert, a total of 24 positive and 5 negative qualitative influences was found for the VSD network.

3 Quantification and sensitivity analysis

When assessing the quantitative part of a belief network, numerous network parameters have to be estimated, either from frequencies found in statistical data, or subjectively by experts in the field of application. But often such statistical data are difficult, if not impossible to obtain, and gathering estimates from field experts is very time-consuming. In this article, we experimentally assess the viability of a procedure to facilitate the quantification task, proposed in [2]. This procedure, which is based on performing sensitivity analyses, is reviewed in Section 3.1. For the experimental investigation, where the VSD network is used as a case study, we have collected three quantifications that differ with respect to informedness of the estimates; these are described in Section 3.2. Furthermore, two variables in the network that are indicative for its performance, and five case profiles, describing typical patterns of observations, have been identified; these are described in Section 3.3.

3.1 One-way sensitivity analysis of a belief network

Sensitivity analysis is a technique to systematically study the effects of variations in the parameters of a mathematical model on this model's predictions. It is widely used in the fields of decision theory and mathematical modelling, [6, 12, 20]. For a Bayesian belief network, sensitivity analysis provides for studying the effects of variations in the estimates of the network's parameters on one or more posterior probabilities of interest. As such, sensitivity analysis allows for identifying network parameters that are highly influential, and should therefore be estimated with the highest accuracy. For less influential parameters, rough estimates may suffice. Sensitivity analysis can thus be used to increase the efficiency of quantifying a belief network, as it directs the quantification effort towards crucial parameters.

The simplest type of sensitivity analysis is a *one-way sensitivity analysis*. In a one-way sensitivity analysis of a belief network, the estimates of the network's parameters are varied

one at a time, keeping all others fixed. The analysis then reveals the separate effect of variation of a parameter estimate on posterior probabilities. In this investigation, we used the method for one-way sensitivity analysis proposed in [3].

Coupé et al. show that in a sensitivity analysis of a belief network, it is not necessary to vary all network parameters, given a particular posterior probability of interest. Only a subset of parameters will influence the posterior probability; this subset can be derived solely from the graphical structure of the network. We will refer to the *sensitivity set* as the set of variables whose parameters may be influential; the constitution of this set depends on the evidence entered into the network and the posterior probability one is interested in. For details concerning the identification of the sensitivity set, we refer to the paper concerned.

Furthermore, Coupé et al. show that there exist functional relationships between individual parameters and posterior probabilities in a belief network. Any posterior probability is a rational polynomial over the parameter under study:

$$\Pr^{\theta=x}(V = v \mid \xi) = \frac{a \cdot x + b}{x + c} \quad (4)$$

where θ is the parameter under study, and a , b , and c are real-valued constants. $\Pr^{\theta=x}(V = v \mid \xi)$ is the posterior probability of the value v for the variable V given evidence ξ . We refer to the right hand side of Eq. 4 as a *sensitivity function*. It is easily seen that systematic variation of the parameter under study is not necessary to determine its associated sensitivity function: if we compute the posterior $\Pr^{\theta=x}(V = v \mid \xi)$ with three different values for θ , the constants in the functional relationship can be determined. These constants are now used to calculate the first order derivative of the sensitivity function:

$$\frac{d}{dx} \Pr^{\theta=x}(V = v \mid \xi) = \frac{a \cdot c - b}{(x + c)^2} \quad (5)$$

If we apply this derivative function to the original parameter estimate, we obtain the gradient of the sensitivity function at that point. This quantity gives an impression of the influence of (small) variations in the estimate on the posterior $\Pr(V = v \mid \xi)$; see Figure 2 for illustrations.

As the influences of parameter estimates on posterior probabilities may vary with the evidence ξ , sensitivity analyses should be performed for several evidence sets. We will refer to these sets as *case profiles*; they preferably consist of realistic patterns of observations. Furthermore, it should be established which posterior probabilities are indicative for the performance of the belief network. The variables to which these probabilities pertain will be called the *variables of interest*. Both realistic case profiles and variables of interest depend on the envisioned application of the belief network under consideration.

To facilitate the quantification of a belief network, Coupé et al. now propose the following two-stage procedure. After the graphical part of the network has been assessed, a rough quantification is established. Such a rough quantification can be based, for instance, on a small collection of statistical data, or order-of-magnitude estimates derived from qualitative descriptions of the relations involved. The second step consists of performing one-way sensitivity analyses on the network; this requires the identification of several realistic case profiles, and one or more variables of interest. Finally, those parameter estimates that turn out to be highly influential are refined. Improved estimates are obtained, for instance, by gathering more statistical data on the variables involved, or by eliciting them from experts in the field. The effort of obtaining highly accurate parameter estimates is thus limited to a subset of network parameters.

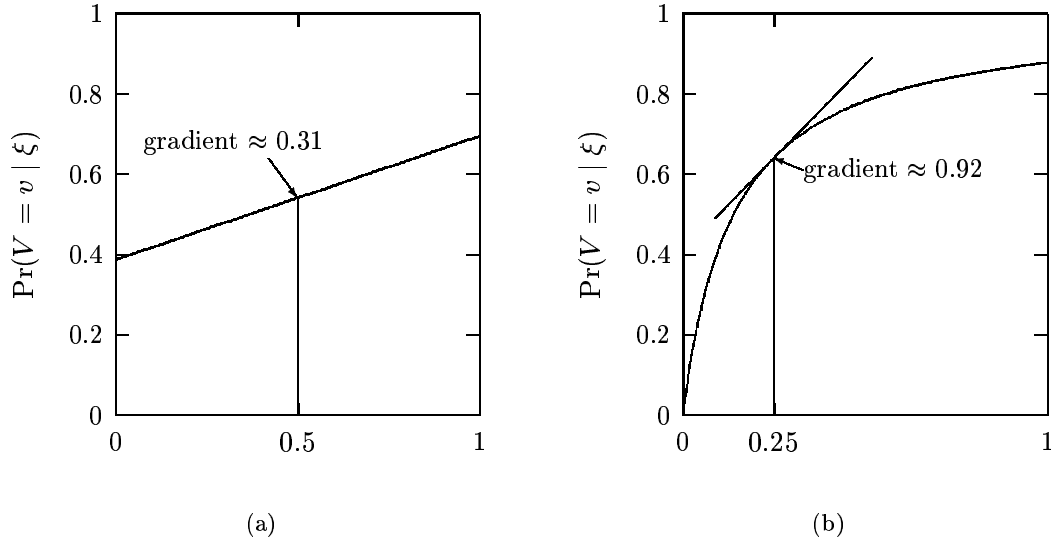


Figure 2: Two sample sensitivity functions. (a) A linear sensitivity function (i.e., $c = 0$). The gradient is independent of the actual point estimate. (b) A nonlinear sensitivity function. The gradient of the function is computed for the point estimate (0.25), providing an indication of the influence of small variations in the parameter estimate on the posterior probability.

We conclude this section by noting that the refinement of influential parameters in a given quantification will generally not reduce the sensitivity of posterior probabilities to parameter variation. A quantification solely comprising highly accurate parameter estimates will often contain just as much influential parameters as a completely random quantification: there exists no relation between sensitivity and ‘quality’ of a quantification.

3.2 Three quantifications of the VSD network

The main question now is: how compares a network quantification obtained by following the procedure described above to a network quantification that consists completely of accurate parameter estimates? Furthermore, what is the *efficiency gain*, in terms of quantification effort, yielded by the procedure? This article presents an experimental investigation into these issues, using the VSD network as a case study.

To be able to answer the questions above, we have acquired three different quantifications of the VSD network. A total of 738 parameters had to be established. The following network quantifications were obtained, listed in order of increasing informedness:

- Q1, consisting of completely uninformative parameter estimates,
- Q2, consisting of parameter estimates supplied by a non-expert, on the basis of qualitative characterisations of the uncertain relations, and
- Q3, consisting of parameter estimates supplied by a field expert.

In network quantification Q1, a uniform probability distribution was used for each variable in the network. The assessed deterministic relations, as described in Section 2.2, were however

preserved in this otherwise uninformative network quantification.

For quantification Q2, the parameter estimates were provided by a non-medical researcher who was involved in the construction of the graphical part of the network. For this quantification, the information on qualitative probabilistic influences between the variables in the network was used. Where possible, a linear model was assumed for the dependency between the parameter estimates for a variable and the values of the parents of that variable. That is, the difference in probability estimates for successive values of the conditioning parents is taken to be equal. In estimating the required prior distributions for variables without ascendants in the graph, both medical literature and qualitative statements of the field expert were used. For the estimated occurrence of different types of VSD, incidence figures found in the literature were used; for the various complications of disease, statements such as “common”, “rare”, “very rare”, given by the field expert during the elicitation of the qualitative part of the network, were translated to probability estimates. Furthermore, as in quantification 1, the deterministic relations were also ensured in quantification 2. In total, seven hours were spent on establishing this quantification of the VSD network.

For the network quantification Q3, parameter estimates were provided by a senior paediatric cardiologist. For each distribution, the cardiologist was asked to provide the expected number of patients out of a hundred with a specific value for the variable under consideration, given a configuration of its parents in the graph. Initially, the clinician felt reluctant to give such precise numbers; he was therefore asked to provide 95% and 50% confidence intervals in addition to the point estimates. As the confidence intervals allowed him to express his own uncertainty regarding the estimates, he felt more comfortable with this procedure. The total amount of time the cardiologist spent on the quantification of the network was approximately twenty-five hours.

Our objective now was to assess whether it is possible to improve quantifications Q1 and Q2 up to the level of Q3, where the improvements consist of selective revisions of influential parameters. These influential parameters are found by performing one-way sensitivity analyses of the network.

3.3 The variables of interest and case profiles under consideration

Two variables of interest were indicated by the field expert to be used in the sensitivity analyses: the variable `shunt` and the variable `resis_ratio`. These are the two most important variables in the network, in the sense that the patient’s prognosis largely depends on their values; a clinician usually bases his management decisions on estimates of these variables. We have therefore assumed that the performance of the VSD network can be measured by testing the accuracy of predictions for these variables. For the variable `shunt`, we have focused on the value `shunt` $\geq 2 : 1$, corresponding to a strongly increased pulmonary blood flow. Continual pulmonary overflow increases the risk of damage to the pulmonary circulation. For the variable `resis_ratio`, taking one of the four values 1:10–1:8, 1:8–1:4, 1:4–1:2, and $>1:2$, the value 1:10–1:8 was taken as the value of interest. This value reflects normal vascular resistance in the pulmonary circulation and therefore corresponds to a favourable situation; the remaining values for `resis_ratio` correspond to increasing pulmonary damage.

In analysing the network quantifications, we thus focused on the posterior probabilities

- $\Pr(\text{shunt} \geq 2 : 1 \mid \text{profile}_i)$ and
- $\Pr(1 : 10 \leq \text{resis_ratio} \leq 1 : 8 \mid \text{profile}_i)$,

and their respective sensitivity to parameter variations, where $profile_i$ stands for a particular case profile (i.e., a set of observations). Five case profiles were selected in cooperation with the field expert; they are shown in Table 1. Except for profile 1, each of these profiles reflects a realistic pattern of observations found in a particular type of VSD patient. Some profiles provide strong, unequivocal evidence towards certain predictions and pose no interpretation problems to clinicians in the field. For instance, profile 2 represents a patient with few, yet compelling findings, providing strong evidence for a small VSD. In profile 5, all symptoms related to a VSD are present, providing strong evidence for a large defect. For other profiles, the prognosis is more difficult to assess, and therefore more uncertain. Profiles 3 and 4 are less evident than profiles 2 and 5, and even somewhat contradictory. Typical signs of a VSD are absent, but still a holosystolic murmur is audible, which is symptomatic for the disease. In profile 4, moreover, a diastolic flow murmur is present, increasing the evidence for a large VSD. The profile that is listed first, finally, shows many symptoms common for a VSD patient, but necessary findings such as systolic murmur and thrill are absent; this profile corresponds to a patient not having a VSD, but some other, unknown disease.

4 Results

For each of the network quantifications Q1, Q2, and Q3, the sensitivity of the selected posterior probabilities to variations in the network parameters was analysed. Subsequently, quantifications Q1 and Q2 were refined by replacing the probability distributions of the most influential variables with distributions from quantification Q3. This section discusses the results of the sensitivity analyses, and the effects of refining network quantifications. First, in Section 4.1, we compare the predictions of the three network quantifications for the five case profiles. The results of the sensitivity analyses are presented in Section 4.2. In Section 4.3, we detail the refinement procedure, and compare the predictions of refined network quantifications to the predictions of quantification Q3.

<i>Evidence</i>	profile 1	profile 2	profile 3	profile 4	profile 5
syst_murmur	<i>no</i>	<i>short_decr</i>	<i>holo_band</i>	<i>holo_band</i>	<i>holo_band</i>
thrill	<i>none</i>	<i>none</i>	<i>none</i>	—	<i>evident</i>
diast_flow_murmur	<i>no</i>	<i>no</i>	<i>no</i>	<i>yes</i>	<i>yes</i>
paleness	<i>yes</i>	<i>no</i>	—	<i>no</i>	<i>yes</i>
sweating	<i>yes</i>	<i>no</i>	—	<i>no</i>	<i>yes</i>
hepatomegaly	<i>no</i>	<i>no</i>	—	—	<i>yes</i>
dyspnoea	<i>yes</i>	<i>no</i>	<i>no</i>	<i>no</i>	<i>yes</i>
feeding_problems	<i>yes</i>	<i>no</i>	<i>no</i>	<i>no</i>	<i>yes</i>
fail_to_thrive	<i>yes</i>	<i>no</i>	—	<i>no</i>	<i>yes</i>

Table 1: The case profiles for the VSD network. A maximum of nine observations is available for every case profile. The profiles are ordered according to the severity of disease. From left to right the VSD size increases and, therefore, also the likelihood of an unfavourable outcome.

4.1 Predictions of the three network quantifications

For each of the network quantifications and each of the case profiles, the predictions for the variables `shunt` and `resis_ratio` were computed; they are shown in Figures 3a and 3b, respectively.

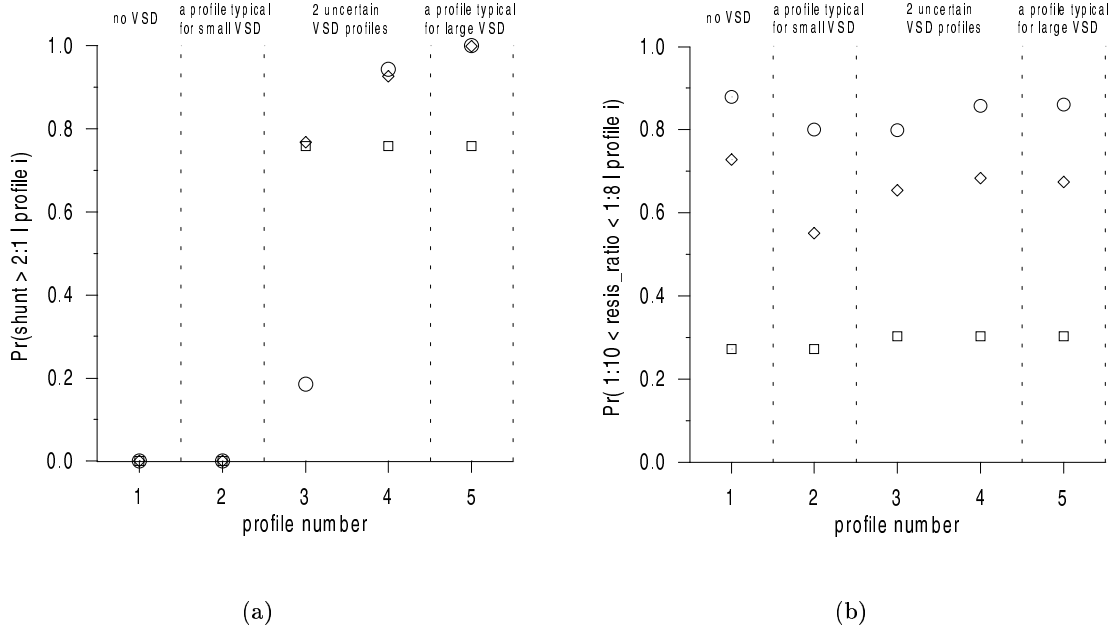


Figure 3: Predictions for the posterior probabilities (a) $\Pr(\text{shunt} \geq 2:1 \mid \text{profile}_i)$ and (b) $\Pr(1:10 \leq \text{resis_ratio} \leq 1:8 \mid \text{profile}_i)$ for each of the case profiles, computed from the three network quantifications (\square for Q1, \diamond for Q2 and \circ for Q3). The vertical dotted lines separate VSD profiles with typical characteristics.

First, consider Figure 3a, showing the results for the variable `shunt`. We see that all quantifications assign zero probability to a large shunt, given profiles 1 and 2. This is due to consistency constraints encoded in each of the quantifications: the lack of loud heart murmurs and thrill precludes existence of a large shunt. These profiles therefore provide no basis for comparison here. For the other profiles, we find that the quantifications Q1 through Q3 have increasingly more discriminative power; this is in line with the increasing level of informedness of the quantifications. Quantification Q1 gives the same prediction for each of these profiles, due to the uniform distributions used in this quantification. The predictions of quantifications Q2 and Q3 are more pronounced and they do provide the same ranking of profiles. For profiles 4 and 5, the predictions from Q2 and Q3 agree well, but for profile 3, however, a large difference is seen in the predictions. This corresponds well to the fact that profile 3 provides contradicting observations and is therefore hard to interpret. The field expert indicated, however, that quantification Q3’s prediction was best in line with his intuition for this profile.

Turning to the predictions for the variable `resis_ratio`, shown in Figure 3b, we see that for each of the quantifications the predictions remain more or less constant over the case profiles. However, the average level of the posteriors differs considerably per quantification:

	Profile	Q1	Q2	Q3
shunt = $\geq 2 : 1$	1	0	0	0
	2	0	0	0
	3	0.01172 (87%)	0.01685 (84%)	0.01446 (83%)
	4	0.01222 (83%)	0.00716 (83%)	0.00882 (85%)
	5	0.01256 (82%)	0.00071 (78%)	0.00000 (77%)
resis_ratio = $1 : 10 - 1 : 8$	1	0.00793 (70%)	0.01230 (80%)	0.02173 (80%)
	2	0.00638 (76%)	0.01058 (77%)	0.01292 (82%)
	3	0.00483 (88%)	0.00738 (85%)	0.00857 (85%)
	4	0.00485 (85%)	0.00726 (83%)	0.00704 (85%)
	5	0.00506 (79%)	0.00715 (79%)	0.00732 (78%)

Table 2: The sensitivities of the posterior probabilities $\Pr(\mathbf{shunt} \geq 2 : 1 \mid \mathit{profile}_i)$ and $\Pr(1 : 10 \leq \mathbf{resis_ratio} \leq 1 : 8 \mid \mathit{profile}_i)$ to parameter variations, averaged out over the network parameters considered in the analysis. The percentage of uninfluential parameters is given between parentheses.

whereas quantification Q1 assigns a low probability of normal pulmonary vascular resistance, quantification Q3 is fairly confident about this circumstance; quantification Q2 is located in between. These differences can be traced back to prior (unconditional) distributions for variables without ascendants in the graph, which are quite different for the three quantifications. In discussing these results, the field expert again confirmed the predictions of his own quantification (Q3): at the age of three to six months, VSD patients usually have a normal pulmonary vascular resistance, regardless of the severity of disease.

4.2 Results of the sensitivity analyses

A one-way sensitivity analysis of both posterior probabilities $\Pr(\mathbf{shunt} \geq 2 : 1 \mid \mathit{profile}_i)$ and $\Pr(1 : 10 \leq \mathbf{resis_ratio} \leq 1 : 8 \mid \mathit{profile}_i)$ was performed for every case profile, for each of the three quantifications. From the total of 1298 parameters in the network, 560 parameters that are determined by a functional relationship or a consistency constraint were excluded from the analysis. Table 2 shows the sensitivity of the posteriors given each case profile, averaged over the 738 parameters considered in the analysis. Between parentheses, the percentage of uninfluential network parameters is written. The high percentages illustrate that there is often a considerable number of network parameters inside the sensitivity set, that yet turn out to be uninfluential when varied. In Table 3 the maximum sensitivity found for each case profile and network quantification is listed.

First, consider the results for the variable **shunt**. We recall from the previous section that for profiles 1 and 2, the posterior $\Pr(\mathbf{shunt} \geq 2 : 1 \mid \mathit{profile}_i)$ is determined to be zero by consistency constraints. For this reason, varying parameters estimates will not affect the posterior in any of the network quantifications; it is completely insensitive, given these profiles. We therefore restrict the discussion to profiles 3, 4 and 5. We find that the predictions for **shunt** in quantifications Q2 and Q3 are significantly more sensitive to parameter variation for profiles 3 and 4 than for profile 5. An explanation for this pattern exists in the fact that profile 5 provides several independent pieces of evidence indicating a large shunt; varying individual parameters therefore hardly influences that prediction. In contrast, profiles 3 and

	Profile	Q1	Q2	Q3
shunt = $\geq 2 : 1$	1	0	0	0
	2	0	0	0
	3	-0.73372	-1.07992	1.51122
	4	-0.73372	-0.38576	-2.68748
	5	0.36686	-0.38789	-0.00020
resis_ratio = $1 : 10 - 1 : 8$	1	0.48170	-0.55913	-5.19101
	2	0.48170	0.53568	0.95637
	3	0.51299	0.49599	0.95812
	4	0.51299	0.48988	-0.82572
	5	0.51299	0.50188	-0.82463

Table 3: The maximum sensitivities of the posterior probabilities $\Pr(\mathbf{shunt} \geq 2 : 1 \mid \mathit{profile}_i)$ and $\Pr(1 : 10 \leq \mathbf{resis_ratio} \leq 1 : 8 \mid \mathit{profile}_i)$.

4 comprise contradicting observations: although heart murmurs indicating a large VSD are observed, none of the symptoms that would then be expected are present. In these cases, varying a single parameter can change the prediction for the **shunt** variable considerably. The result is not found for the uninformed quantification Q1: as this quantification gives the same prediction for each case profile, these predictions are almost equally sensitive to parameter variation.

For each case profile and each quantification, we identified the thirty parameters showing the largest influence on the posterior $\Pr(\mathbf{shunt} \geq 2 : 1 \mid \mathit{profile}_i)$. The variables to which they pertain are listed in Table 4, in order of decreasing maximum influence of their parameters. We note that there exists a substantial overlap in the variables to which highly influential parameters pertain, and therefore the number of variables is much smaller than thirty. Furthermore, the selections of variables per profile are roughly the same for all quantifications. This indicates that the structure of the belief network considerably affects the sensitivity of posteriors to parameter variation; the quantification that is used in the analysis is of secondary importance.

For quantification Q1, only the prediction variable itself (**shunt**), and observed, direct descendants of this variable are selected.¹ This is not surprising, as the uniform distributions used in this quantification eliminate all influences through longer pathways in the graph when only one parameter estimate is varied at a time. Therefore, only higher-order sensitivity analyses can reveal the propagation of influences through the graph for this quantification. The selections for quantifications Q2 and Q3 are supersets of the selection for quantification Q1. Notably, they also comprise variables at a greater distance of the **shunt** variable, and ascendant variables of **shunt** in the graph, e.g., **defect_size** and **resis_ratio**. The distributions of these ascendant variables represent prevalences of the disease and its complications, and are therefore influential on the posterior distribution of **shunt**. We conclude that the selections for quantifications Q2 and Q3 provide more realistic patterns of influential variables.

Consider now the results, in Tables 2 and 3, for the variable **resis_ratio**. For this variable, the differences in average and maximum sensitivity between the three quantifications and five case profiles are very small. This illustrates that there is no relation between

¹The variable **LV_failure** is functionally determined by its observed descendants and can therefore itself be regarded as observed.

	Q1	Q2	Q3
profile 3	LV_failure diast_flow_murmur thrill syst_murmur shunt	syst_murmur syst_P _{RV} /P _{LV} diast_flow_murmur LV_failure thrill shunt resis_ratio	LV_failure syst_P _{RV} /P _{LV} diast_flow_murmur syst_murmur pulm_sten shunt defect_size PDA outlet_pos thrill
profile 4	LV_failure paleness diast_flow_murmur sweating syst_murmur shunt	syst_P _{RV} /P _{LV} syst_murmur diast_flow_murmur paleness sweating LV_failure shunt outlet_pos pulm_sten resis_ratio	diast_flow_murmur LV_failure syst_P _{RV} /P _{LV} syst_murmur pulm_sten shunt paleness sweating defect_size PDA outlet_pos
profile 5	diast_flow_murmur paleness sweating thrill syst_murmur LV_failure shunt	diast_P _{AP} /P _{Ao} thrill syst_P _{RV} /P _{LV} diast_flow_murmur syst_murmur LV_failure paleness sweating shunt outlet_pos pulm_sten resis_ratio	diast_flow_murmur LV_failure sweating RV_failure thrill syst_murmur paleness VSD_type syst_P _{RV} /P _{LV} pulm_sten shunt

Table 4: The variables to which the thirty network parameters pertain that are most influential to the posterior $\Pr(\text{shunt} \geq 2 : 1 \mid \text{profile}_i)$, for profiles 3, 4, and 5. The variables are ordered with respect to the maximum influence of their parameters.

the level of informedness of a network quantification and the sensitivity for variations in parameter estimates. Furthermore, there exists no distinction between the ‘contradicting’ profiles 3 and 4 and other profiles; this seems to be correct as these contradictions mainly concern the size of the shunt. At the age of three to six months, only minor differences between patients with regard to the resistance ratio are to be expected. For the posterior $\Pr(1 : 10 \leq \text{resis_ratio} \leq 1 : 8 \mid \text{profile}_i)$, the variables pertaining to the thirty most influential parameters were also identified. Due to space limitations, these results are not shown here. However, similar observations hold as for the posterior $\Pr(\text{shunt} \geq 2 : 1 \mid \text{profile}_i)$. Again, the selections of influential variables per profile are roughly the same for each quantification. Every selection contains the variable `resis_ratio` itself, as well as its direct ascendant in the graph, the variable `pulm_art`. The direct descendants `diast_PAP/PAo` and `shunt` also turn out to be very influential. Furthermore, the selected variables at greater distance from `resis_ratio` partially overlap with those variables selected for `shunt`.

4.3 Predictions of the refined network quantifications

Using the results of the sensitivity analyses, quantifications Q1 and Q2 were stepwise refined with parameter estimates from quantification Q3. Although sensitivity analysis reveals the influence of individual parameters, we chose to substitute, at every refinement step, all parameter estimates pertaining to a network variable. The motivation for this approach is that parameter estimates often have little meaning in isolation: it is their relation with other parameter estimates from the same local distribution that matters. To select variables whose parameters are eligible for substitution, the sets of influential variables per profile (Table 4) were compiled to a single, ordered set; the order was determined by averaging the positions of the variables in the original sets.

With respect to the `shunt` variable, the following sets were thus compiled:

$$V_{Q_1}(\text{shunt}) = \{\text{diast_flow_murmur}, \text{LV_failure}, \text{syst_murmur}, \text{shunt}, \text{paleness}, \text{thrill}, \text{sweating}\} \quad (6)$$

$$V_{Q_2}(\text{shunt}) = \{\text{syst_P}_{RV}/\text{P}_{LV}, \text{syst_murmur}, \text{diast_flow_murmur}, \text{LV_failure}, \text{shunt}, \text{resis_ratio}, \text{thrill}, \text{paleness}, \text{sweating}, \text{outlet_pos}, \text{pulm_sten}, \text{diast_P}_{AP}/\text{P}_{Ao}\}. \quad (7)$$

With respect to the variable `resis_ratio`, we have

$$V_{Q_1}(\text{resis_ratio}) = \{\text{resis_ratio}, \text{shunt}, \text{pulm_art}, \text{diast_P}_{AP}/\text{P}_{Ao}, \text{LV_failure}, \text{PDA}\} \quad (8)$$

$$V_{Q_2}(\text{resis_ratio}) = \{\text{resis_ratio}, \text{shunt}, \text{pulm_art}, \text{LV_failure}, \text{diast_P}_{AP}/\text{P}_{Ao}, \text{syst_P}_{RV}/\text{P}_{LV}, \text{syst_murmur}, \text{RV_failure}, \text{diast_flow_murmur}, \text{paleness}, \text{sweating}, \text{thrill}\} \quad (9)$$

The variables are listed in decreasing order of influence on the respective variables.

The quantifications were stepwise refined by replacing, at step i , all parameter estimates pertaining to the i^{th} variable in the above sets by the corresponding estimates from quantification Q3. So, to improve the prediction for `shunt`, a total of seven and twelve refinement steps

were made for Q1 and Q2, respectively. For `resis_ratio`, a total of six and twelve refinement steps were made for Q1 and Q2. With each of the refined quantifications, the posteriors $\Pr(\text{shunt} \geq 2 : 1 \mid \text{profile}_i)$ and $\Pr(1 : 10 \leq \text{resis_ratio} \leq 1 : 8 \mid \text{profile}_i)$ were computed for the various case profile; these posteriors are plotted in Figures 4 and 5.

First consider the refinements for the posterior probability $\Pr(\text{shunt} > 2 : 1 \mid \text{profile}_i)$. The plots indicate that for both quantifications, the posteriors rapidly shift towards the posteriors of quantification Q3. For quantification Q1, for example, two refinement steps suffice to approach Q3’s predictions. For profile 5, even a single refinement of quantification Q1 (replacing the estimates of the variable `diast_flow_murmur`) is enough to nearly reach Q3’s prediction. However, the difference between posteriors does generally not decrease monotonically. This is most notable with the refinements of quantification Q2: the posteriors for profiles 3 and 4 show considerable fluctuations in the first seven steps. Thereafter, they quickly converge to the desired level. Unfortunately, this convergence is not obtained for quantification Q1: even after replacing the distributions of all seven variables with influential parameters, the posteriors still deviate from the posteriors of Q3.

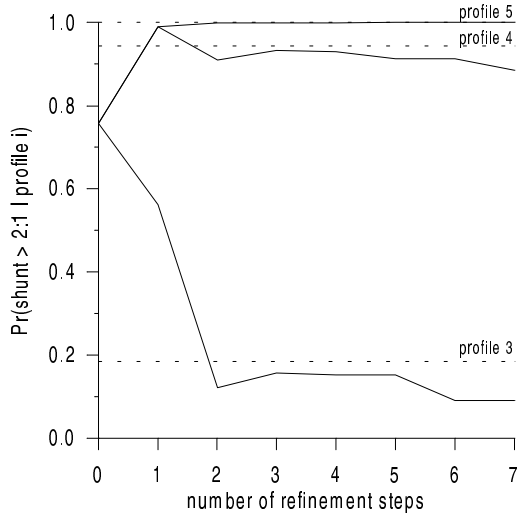
Now, we turn to the refinements for the posterior $\Pr(1 : 10 \leq \text{resis_ratio} \leq 1 : 8 \mid \text{profile}_i)$. As before, the posteriors quickly approach the posteriors of quantification Q3. For both quantification Q1 and Q2, three refinement steps suffice to reduce the difference with Q3’s predictions considerably. After the maximum of six refinement steps, however, still no convergence is reached for Q1. For quantification Q2, the refined posteriors lie very close to the predictions for Q3 after the maximum of eight steps.

So far, we have seen that the results of refinements are encouraging. However, these results pertain to the five profiles that were also used in the sensitivity analyses. In order to investigate whether the results generalise over more cases, the effects of refinements were also tested on clinical data. Thirty-six cases were selected from a database of VSD patients collected at the Leiden University Medical Centre in The Netherlands. These cases correspond to patients aged 3–6 months having VSD as their primary diagnosis. For each case, the predictions of quantifications Q1 and Q2 both before and after the refinements were compared with the predictions of quantification Q3. The average, maximum and minimum difference between the predictions from Q3 and both the original and refined predictions from Q1 and Q2 for the thirty-six cases were computed. In Tables 5a and 5b, these differences for refinement with two, four and six variables, respectively, are shown.

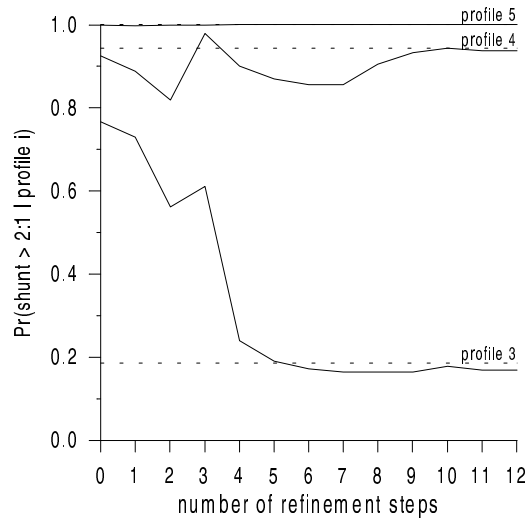
For quantification Q1, the results show that stepwise refining this quantification indeed steadily reduces the differences with the predictions from quantification Q3. Note, that we do not claim that the predictions for Q3 are reliable. Since no reliable outcome measurements were available for these thirty-six patients no validation of the various quantifications was performed. Therefore, we only compare the refined quantifications with Q3 and aim to obtain a network quantification giving similar predictions as Q3, without using all parameter estimates from Q3.

For quantification Q2, the results of the refinement procedure are less clear. For the variable `shunt`, the mean difference between Q3 and both Q2 and refinements of Q2 indeed show a decreasing trend. However, the original differences between the predictions from Q2 and Q3 are quite small, making drastic changes impossible. For `resis_ratio`, no significant effect of the refinements can be observed. This is an unsatisfying result; future research will have to uncover its causes.

As an example, in Figures 6a and b, the detailed results of the refinement procedure are given for two patients, one patient for whom Q3 predicts a low shunt and one patient for

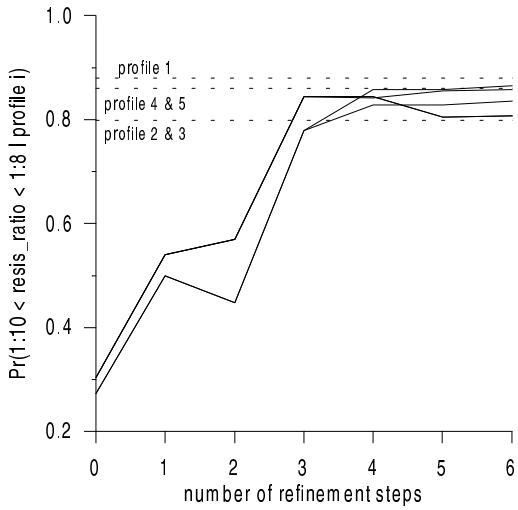


(a)

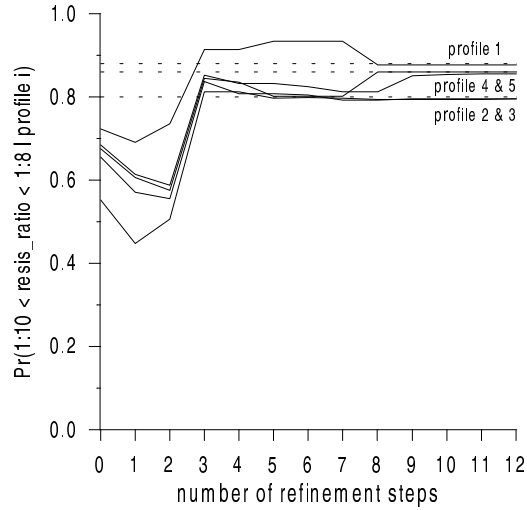


(b)

Figure 4: The predictions of (a) quantification Q1 and (b) quantification Q2 for the posterior probability $\Pr(\text{shunt} \geq 2 : 1 \mid \text{profile}_i)$, after successive refinement steps. The dotted horizontal lines correspond to the predictions of quantification Q3.



(a)



(b)

Figure 5: The predictions of (a) quantification Q1 and (b) quantification Q2 for the posterior probability $\Pr(1 : 10 \leq \text{resis_ratio} \leq 1 : 8 \mid \text{profile}_i)$, after successive refinement steps. The dotted horizontal lines correspond to the predictions of quantification Q3.

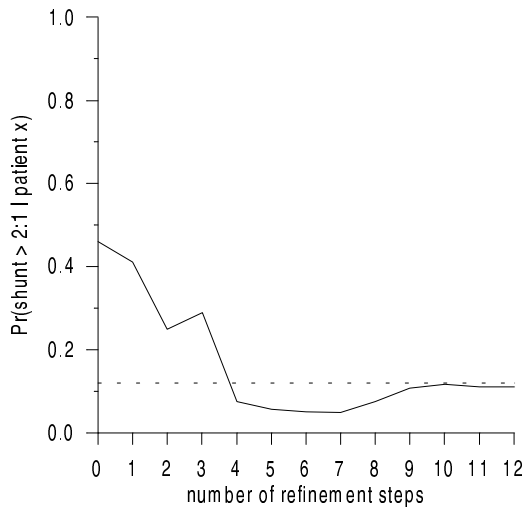
	step	mean	max	min
Q1	0	0.2920	0.6396	0.1178
	2	0.1999	0.8381	0.0002
	4	0.1905	0.8080	0.0000
	6	0.0927	0.3859	0.0001
Q2	0	0.1313	0.5501	0.0042
	2	0.1497	0.5501	0.0022
	4	0.0686	0.3285	0.0000
	6	0.1032	0.3859	0.0001

(a) **shunt**

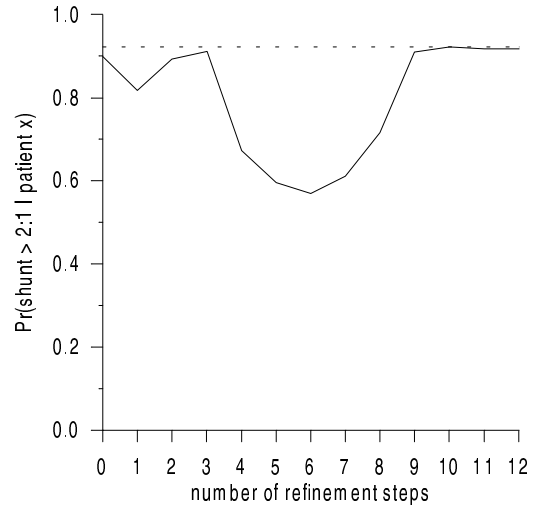
	step	mean	max	min
Q1	0	0.5186	0.6058	0.0616
	2	0.2999	0.3527	0.2252
	4	0.0412	0.0868	0.0083
	6	0.0303	0.0697	0.0002
Q2	0	0.3881	0.7789	0.0410
	2	0.3829	0.7664	0.0223
	4	0.4426	0.9113	0.0024
	6	0.4407	0.8992	0.0035

(b) **resis_ratio**

Table 5: The mean, maximum and minimum difference with Q3 of the refined predictions of Q1 and Q2, for all cases in the database whose predictions are not determined by a consistency constraint.



(a)



(b)

Figure 6: The effect of refining quantification Q2 for (a) a patient for whom the procedure is effective and (b) a patient for whom the procedure gives large fluctuations.

whom a high shunt is expected. Figures 6a illustrates that the refinement procedure may work very effectively. After four refinement steps, the predictions from Q2 have approached Q3 considerably. Convergence, however, is reached only after nine steps. Figures 6b shows a patient for whom the refinement procedure originally only worsens the predictions. This suggests that in applying the refinement procedure to quantify a network efficiently, it may be worthwhile to use real patient data for the sensitivity analyses.

5 Discussion and Conclusions

Quantifying a Bayesian belief network is a difficult and time-consuming task, precluding easy application of belief-network technology in practice. However, it has been claimed that, once the graphical part of the network correctly models the independence relations in the domain of application, then the behaviour of the network is insensitive to the quality of the majority of quantification parameters, [17]. If this is true, then a satisfactory network quantification can be obtained by only estimating a small set of highly influential parameters from well-informed sources, and taking rough estimates for the others. We have presented an empirical investigation into this claim, by comparing the predictions of a well-informed quantification with poorly-informed quantifications, in which only influential parameters were reconsidered. These influential parameters were identified by performing one-way sensitivity analyses.

The results of our investigation suggest that using a two-step procedure of only reconsidering influential parameters works: partially refined, poorly-informed quantifications give predictions that are comparable to a well-informed quantification. The procedure could be repeated several times depending on the level of informedness of the quantification at hand. However, the procedure shows better results for the case profiles that were used to identify influential parameters, than for cases from a real-world, clinical database. We conclude that it is preferable to use a large set of cases when refining a network quantification. With respect to sensitivity analysis, it was found that the structure of the belief network considerably affects the influence of parameter variation to posterior probabilities; the quantification that is used in the analysis is of secondary importance. Furthermore, our results confirm that there is no relation between sensitivity of posteriors to parameter variation and the level of informedness of the quantification at hand.

Throughout the investigation, we have assumed that the structure of the belief network, as elicited from the field expert, is correct; the same assumption was made for functional relations and consistency constraints on network variables that were identified prior to quantification. We believe that these assumptions have had little or no impact on the results that were found. When building a real-world application, however, critical evaluation of these parts of the network model is also necessary: sensitivity analyses should not be restricted to numerical information. Furthermore, the re-quantifications concerned influential parameters that were found by performing one-way sensitivity analyses. These analyses measure the effects of individual parameter variations. Therefore, synergetic effects of varying multiple parameters are not detected, although they may have an important effect on the network's predictions. To reveal such effects, higher-order sensitivity analyses are required.

The method proposed here is not limited to elicitation of network parameters from field experts; it is also applicable to parameter elicitation from other sources such as clinical data sets or frequencies reported in the literature. In fact, we believe that the usage of objective statistical sources is indispensable to obtain a network quantification of sufficient quality.

Subjective probability estimates are known to suffer from several forms of calibration and bias, [18], and their reliability is therefore not beyond dispute. Moreover, in the medical field it is often possible to collect datasets of reasonable size. The basic procedure investigated, however, applies equally well when a combination of quantification sources is employed.

To conclude, we believe that sensitivity analysis provides a promising addition to the methods that exist to facilitate belief network quantification. In future research, we plan to investigate more sophisticated procedures than the one described here. For instance, instead of making a final network quantification on the basis of a single sensitivity analysis, it is probably better to have a few alternating steps of sensitivity analyses and improvements of the quantification. This approach takes into consideration that by each refinement, the set of highly influential parameters is changed. Furthermore, it may be worthwhile to additionally perform higher-order sensitivity analyses, or even *uncertainty analyses*, which investigate the joint effect of varying all network parameters simultaneously. And finally, the expected accuracy of parameter estimates, as expressed by confidence intervals, can be involved in the analysis. This is accomplished by taking the variation of a sensitivity function over a confidence interval instead of over the whole range as a measure of the parameter's influence. Then, parameter estimates with high expected accuracy (i.e., having small confidence intervals) will only be reconsidered when the sensitivity function is extremely steep over the confidence interval.

Acknowledgements

The investigations were (partly) supported by the Netherlands Computer Science Research Foundation with financial support from the Netherlands Organisation for Scientific Research (NWO). The authors wish to thank Robert Cannon for valuable discussions, and Linda van der Gaag, Peter Lucas and Silja Renooij for their useful comments on the manuscript.

References

- [1] S. Andreassen, M. Woldbye, B. Falck, and S.K. Andersen, MUNIN – A causal probabilistic network for interpretation of electromyographic findings, in: *Proceedings of the Tenth International Joint Conference on Artificial Intelligence* (Morgan Kaufmann, Palo Alto, 1987) 366–372.
- [2] V.M.H. Coupé, L.C. van der Gaag, and J.D.F. Habbema, Sensitivity analysis: an aid for belief-network quantification, *Technical Report UU-CS-1999-13*, Dept. of Computer Science, Utrecht University, 1999.
- [3] V.M.H. Coupé and L.C. van der Gaag, Practicable sensitivity analysis of Bayesian belief networks, *Technical Report UU-CS-1998-10*, Dept. of Computer Science, Utrecht University, 1998.
- [4] M.J. Druzdzel and L.C. van der Gaag, Elicitation of probabilities for belief networks: combining qualitative and quantitative information, in: *Eleventh Conference on Uncertainty in Artificial Intelligence* (Morgan Kaufmann, Palo Alto, 1995) 141–148.
- [5] T.P. Graham and H.P. Gutgesell, Ventricular septal defects, in: A.J. Moss and F.H. Adams et al., eds., *Heart Disease in Infants, Children, and Adolescents* (Williams & Wilkins, Baltimore, 1995) 724–746.

- [6] J.D.F. Habbema, P.M.M. Bossuyt, and D.W.J. Dippel, Analysing clinical decision analyses, *Statistics in Medicine* 9 (1990) 1229–1242.
- [7] D.E. Heckerman, E.J. Horvitz, and B.N. Nathwani, Toward normative expert systems, Part 1: The Pathfinder project, *Methods of Information in Medicine* 31 (1992) 90–105.
- [8] A.L. Jensen, Quantification experience of a DSS for mildew management in winter wheat, in: M.J. Druzdzel, L.C. van der Gaag, M. Henrion, and F.V. Jensen, eds., *Working Notes of the Workshop on Building Probabilistic Networks: Where Do the Numbers Come From?* (1995) 23–31.
- [9] L.J. Krovetz, Spontaneous closure of ventricular septal defect, *American Journal of Cardiology* 81 (1998) 100–101.
- [10] J.H. Moller, C. Patton, R.L. Varco, and C.W. Lillehei, Late results (30 to 35 years) after operative closure of isolated ventricular septal defect from 1954 to 1960, *American Journal of Cardiology* 68 (1991) 1491–1497.
- [11] S. Monti and G. Carenini, Dealing with expert inconsistencies: the sooner the better, in: M.J. Druzdzel, L.C. van der Gaag, M. Henrion, and F.V. Jensen, eds., *Working Notes of the Workshop on Building Probabilistic Networks: Where Do the Numbers Come From?* (1995) 33–40.
- [12] M. Morgan and M. Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (Cambridge University Press, Cambridge, 1990).
- [13] K.G. Olesen, U. Kjaerulff, F. Jensen, F.V. Jensen, B. Falck, S. Andreassen, and S.K. Andersen, A MUNIN network for the median nerve – a case study on loops, *Applied Artificial Intelligence* 3 (1989) 385–404.
- [14] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, Palo Alto, 1988).
- [15] N. Peek, Explicit temporal models for decision-theoretic planning of clinical management, *Artificial Intelligence in Medicine* 15(2) (1999) 135–154.
- [16] N. Peek and J. Ottenkamp, Developing a decision-theoretic network for a congenital heart disease, in: E. Keravnou, C. Garbay, R. Baud, and J. Wyatt, eds., *AIME '97: Proceedings of the Sixth Conference on Artificial Intelligence in Medicine Europe* (Springer Verlag, Berlin, 1997) 157–168.
- [17] M. Pradhan, M. Henrion, G. Provan, B. Del Favero, and K. Huang, The sensitivity of belief networks to imprecise probabilities: an experimental investigation, Technical Report SMI-95-0595, Stanford University, Section on Medical Informatics, 1995.
- [18] D. Kahneman, P. Slovic, and A. Tversky, *Judgement under Uncertainty: Heuristics and Biases* (Cambridge University Press, Cambridge, 1982).
- [19] M.P. Wellman, Fundamental concepts of qualitative probabilistic networks, *Artificial Intelligence* 44(3) (1990) 257–303.
- [20] D. von Winterfeldt and W. Edwards, *Decision Analysis and Behavioral Research* (Cambridge University Press, Cambridge, 1986).