Bayesian Networks for Evidence Based Clinical Decision Support.

Yet, Barbaros

QUEEN MARY, UNIVERSITY OF LONDON

Bayesian Networks for Evidence Based Clinical Decision Support

Barbaros Yet

2013

Submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Declaration

I, Barbaros Yet, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged and my contribution indicated. Previously published material is also acknowledged.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Barbaros Yet

Date: 12/12/2013

Acknowledgement

I would like to thank İdil for her endless support, patience and love, which made this thesis possible.

Dr William Marsh's supervision and guidance has been invaluable. I am grateful for the countless things I have learned from him.

I would like to thank Professor Norman Fenton and Professor Martin Neil for their valuable contributions and brilliant ideas.

I have been very lucky to collaborate with Mr Zane Perkins and Mr Nigel Tai. It has been a pleasure working with them.

I would like to thank Yiğit and Ebru. They have always helped me make the most important decisions in my career.

Finally, I am grateful to 'annem', Asuman, for always helping me set the right objectives and encouraging me to follow them, and to 'babam', Orhun, for teaching me the essential skills to reach to those objectives.

Abstract

Evidence based medicine (EBM) is defined as the use of best available evidence for decision making, and it has been the predominant paradigm in clinical decision making for the last 20 years. EBM requires evidence from multiple sources to be combined, as published results may not be directly applicable to individual patients. For example, randomised controlled trials (RCT) often exclude patients with comorbidities, so a clinician has to combine the results of the RCT with evidence about comorbidities using his clinical knowledge of how disease, treatment and comorbidities interact with each other. Bayesian networks (BN) are well suited for assisting clinicians making evidence-based decisions as they can combine knowledge, data and other sources of evidence. The graphical structure of BN is suitable for representing knowledge about the mechanisms linking diseases, treatments and comorbidities and the strength of relations in this structure can be learned from data and published results. However, there is still a lack of techniques that systematically use knowledge, data and published results together to build BNs.

This thesis advances techniques for using knowledge, data and published results to develop and refine BNs for assisting clinical decision-making. In particular, the thesis presents four novel contributions. First, it proposes a method of combining knowledge and data to build BNs that reason in a way that is consistent with knowledge and data by allowing the BN model to include variables that cannot be measured directly. Second, it proposes techniques to build BNs that provide decision support by combining the evidence from meta-analysis of published studies with clinical knowledge and data. Third, it presents an evidence framework that supplements clinical BNs by representing the description and source of medical evidence supporting each element of a BN. Fourth, it proposes a knowledge engineering method for abstracting a BN structure by showing how each abstraction operation changes knowledge encoded in the structure. These novel techniques are illustrated by a clinical case-study in trauma-care. The aim of the case-study is to provide decision support in treatment of mangled extremities by using clinical expertise, data and published evidence about the subject. The case study is done in collaboration with the trauma unit of the Royal London Hospital.

Table of Contents

Glossary of Abbreviations

AI	Artificial Intelligence
AIS	Abbreviated Injury Score
APTTR	Active Partial Thromboplastin Time Ratio
AR	Arterial Repair
AS	Anatomical Site
ATC	Acute Traumatic Coagulopathy
AUROC	Area Under the ROC Curve
BE	Base Excess
BN	Bayesian Network
BS	Brier Score
BSS	Brier Skill Score
CI	Conditional Independence
COAST	Coagulopathy of Severe Trauma Score
CPD	Conditional Probability Table
DCS	Damage Control Surgery
EM	Expectation-Maximisation
EBM	Evidence Based Medicine
FAST	Focussed Assessment with Sonography for Trauma
GCS	Glasgow Coma Score
GS	Grow-shrink Algorithm
HC	Hill Climbing Algorithm
HFS	Hannover Fracture Score
HR	Heart Rate
HT	Haemothorax
IC	Inductive Causation
IDE	Interactive Development Environment
INR	International Normalised Ratio
ISR	United States Army Institute of Surgical Research
ISS	Injury Severity Score
LB	Long Bone Injury
LEVT	Lower Extremity Vascular Trauma
LSI	Limb Salvage Index
MAI	Arterial Injury at Multiple Levels
MCMC	Markov Chain Monte Carlo
MESI	Mangled Extremity Syndrome Index
MESS	Mangled Extremity Severity Score
ML	Machine Learning
MMHC	Max-Min Hill Climbing Algorithm
NE	Nonviable Extremity

NISSA	Nerve, Ischemia, Soft tissue, Skeletal, Shock, Age Score
NPT	Node Probability Table
NT	Mr Nigel Tai
OOBN	Object Oriented Bayesian Networks
OWL	Web Ontology Language
PMID	PubMed Identification Number
PSI	Predictive Salvage Index
PTR	Prothrombin Time Ratio
RCT	Randomised Controlled Trial
RF	Repair Failure
RLH	Royal London Hospital
ROTEM	Rotational Thromboelastometry
RR	Respiratory Rate
SBP	Systolic Blood Pressure
SF-36	Short Form 36 Health Survey
SMFA	Short Musculoskeletal Function Assessment
TRISS	Trauma and Injury Severity Score
UMLS	Unified Medical Language System
UP	Unstable Pelvis
ZP	Mr Zane Perkins

List of Figures

List of Tables

Chapter 1

Introduction

Evidence based medicine (EBM) has been the predominant paradigm in medical decision making for the last 20 years (Evidence-Based Medicine Working Group, 1992; Sackett et al., 1996; Straus et al., 2005). The underlying idea of EBM is to search and employ the best available evidence to make clinical decisions. Several ranking systems have been proposed to weigh the evidence when multiple sources of evidence is available (Guyatt et al., 2008; Hadorn et al., 1996; Harbour and Miller, 2001). These systems rank the evidence according to the way it is collected: the evidence from randomised controlled trials (RCT) has higher ranks than the evidence from datasets, and the evidence from expert opinion have the lowest ranks. However, a study with lower rank must never be ignored unless a higher rank study targets exactly the same population with exactly the same inclusion criteria (Marshall, 2006; Rawlins, 2008). RCTs provide the highest ranked evidence as they are powerful tools for understanding treatment effects by eliminating confounding and biases. The absence of RCTs, however, is not the same as the absence of evidence (Sackett et al., 1996; Smith and Pell, 2003). Evidence from clinical expertise and datasets should be used even when RCTs are available since:

1. It is not possible to conduct RCTs for many clinical problems because of ethical or practical difficulties (Horton, 2000; Rawlins, 2008; Sackett et al., 1996). For example, although a prosthesis following a limb amputation can be beneficial for patients with painful or non-functioning limbs, it is ethically impossible to conduct an RCT for studying the benefits and disadvantages of this intervention. Horton (2000) describes the time and cost challenges of trialing the clinical use of coronary stents, with the results that several types of these stents are commonly used without any RCTs supporting their use.

Clinicians evaluated the benefits and risks of the coronary stents based on their expertise and the information from medical datasets.

2. The results from RCTs often cannot be generalised to the individual patients treated by clinicians (Marshall, 2006). Bradford Hill, the architect of RCTs, notes the issues about generalisability: "it is wise to limit the questions strictly to a few and to be absolutely precise upon those few. The loss in so doing lies, of course, in the fact that the answers are limited to very specific questions and clearly cannot be generalised upon outside their fields." (Hill, 1951). RCT studies are designed with strict inclusion criteria to observe the effects in less time, and to decrease the already high costs (Rawlins, 2008). For example, as comorbidities can interact with treatment effects, patients with comorbidities are often excluded from RCTs to observe the effects in less time and with fewer patients. Individual patients, however, can have comorbidities and thus the result of such RCTs may not be valid for them. Moreover, RCTs about the same subject can have conflicting results because of the differences in their inclusion criteria (Marshall, 2006; Rawlins, 2008).

In order to make evidence-based decisions for individual patients, all of the relevant evidence about the disease, treatment options, and background factors of the patient must be taken into account, and combined, whether they are RCT or not. Clinicians are an essential part of this: the evidence can be combined for individual patients only by using 1) the clinical opinion about the similarities and differences between an individual patient and the available evidence, 2) the clinical expertise about the disease mechanisms of how different evidence relates to each other (Guyatt et al., 2004; Haynes et al., 2002; Marshall, 2006; Sackett et al., 1996).

There are, however, significant challenges for clinicians to apply EBM in daily practice. The time that clinicians can spare for reviewing evidence keeps getting smaller as their workload continues to increase (Royal College of Nursing, 2012; Smith, 2013). Even though technologies, such as PubMed and MeSH, has made it easier to access publication, identifying the relevant evidence is often time consuming. The ever increasing number of medical journals and publications makes this even more challenging (Alper et al., 2004; Haynes, 1993). Moreover, combining

the results of the related studies can be both mathematically challenging and time consuming.

Evidence can potentially be encoded in quantitative models, which can then make mathematically correct predictions for individual patients. For example, a model that combines the separate pieces of evidence regarding treatment and comorbidity outcomes can make predictions for the individual patients who have both of these factors. In order to combine such evidence, a quantitative model must be capable of modelling the clinical knowledge about the mechanisms between the treatment, comorbidity and outcome. Most traditional modelling approaches, however, cannot represent the clinical knowledge about disease mechanisms especially when the mechanisms are complicated containing multiple and interrelated pathways (Buchan et al., 2009). For example, statistical tools, such as meta-analysis, can effectively combine the evidence about simple relations but they are not well suited for integrating knowledge about the complicated mechanistic relations from clinicians.

A Bayesian network (BN) is a probabilistic graphical model that is composed of a graphical structure that represents the relations between the variables, and a set of parameters that defines the strength of these relations. A BN can be used to make probabilistic inferences given the information observed. BNs offer a convenient and powerful approach for providing decision support based on knowledge and data. The graphical structure of the BN is well suited for representing knowledge about the disease mechanisms and clinical pathways. Evidence, from RCTs, data and clinical opinion, can be combined to learn the strength of relations in this structure. As a result of these unique features, BNs offer a powerful way of providing evidence-based decision support for individual patients. Moreover, the reasoning mechanism and predictions of BNs can be presented to clinicians as they have a graphical structure suited for representing knowledge.

BNs for EBM, however, cannot be built automatically from data. Clinicians must be closely involved in various stages of the modelling to identify the related evidence and to provide clinical knowledge about the mechanistic relations. Although several knowledge engineering methodologies exists (Cano et al., 2011; Flores et al., 2011; Helsper and van Der Gaag, 2007; Laskey and Mahoney, 2000; Neil et al., 2000),

there are still many issues in BN development that need to be addressed with methods that systematically combine knowledge and evidence.

Another challenge in applying BNs for EBM is to present the evidence behind the BNs. Many publications do not give a thorough description of the BN structure even when the BN is based on extensive clinical knowledge (for examples of inadequately described knowledge-based BNs see Ahmed et al., 2009; Burnside et al., 2006; Onisko et al., 1998; Wasyluk et al., 2001). This makes it difficult, if not impossible, to understand the evidence supporting the BN and its derivation steps.

1.1 Research Objectives

The primary objective of this thesis is to provide practical tools that combine evidence to provide decision support for EBM. The secondary research objectives that contribute to the primary objective are:

1. To show that it is possible to build decision support models that are consistent with the best available evidence by combining clinical knowledge with data. The observed data is more useful when it is analysed consistent with clinical knowledge.

2. To show that it is possible to provide clinical decision support by combining the evidence from systematic reviews that is pooled by meta-analysis with clinical knowledge and data about the domain.

3. To show how a practical decision support model can be derived through a series of simplifications without losing the link between the simplified model and underlying domain knowledge.

4. To represent both supporting and conflicting clinical evidence about the important clinical factors and relations involved in decision making whether or not they are included in the decision support model.

The novel contributions are illustrated by a case study about the treatment of mangled extremities. In an attempt to provide decision support for this treatment, we propose two BNs that are developed by combining clinical knowledge, previous

research and data about the domain. We examine systematic ways of using different sources of evidence in BN development, and presenting the evidence to the user.

The case-study was done in collaboration with the trauma sciences unit of the Royal London Hospital (RLH). The RLH provided the clinical expertise and patient datasets that are used throughout the thesis. The AgenaRisk software was used for building and calculating the BNs models presented in this thesis (Agena Ltd, 2013).

1.2 Structure of the Thesis

Chapter 2 presents an introduction to BNs and their conditional independence (CI) properties. The introduction is followed by a review of the existing methods for building BNs from knowledge and data. The BN properties presented in this chapter are necessary to follow the methodologies presented in Chapters 5 – 8.

Chapter 3 examines the potential benefits of quantitative models in medical and surgical decision making. It reviews the existing approaches for developing medical decision support models, and investigates why some models are not being adopted by clinicians.

Chapter 4 introduces the trauma case study and reviews the previous models that have been developed for this domain. This chapter examines the decision making in mangled extremity treatment, and discusses the challenges of building useful decision support models for the domain. Finally, this chapter gives an overview of Chapters 5 – 8, with a brief discussion of how these chapters address the challenges.

Chapter 5 proposes a methodology of combining knowledge and data to build BNs that reason in a way that is consistent with knowledge and data by allowing the BN model to include variables that cannot be measured directly. The methodology is illustrated by a BN that is used to provide decision in mangled extremity treatment by predicting a potentially fatal physiological disorder in early stages of the treatment. Several variables in this BN, including the variable indicating the state of the physiological disorder, cannot be directly measured and thus they are not present in the dataset.

Chapter 6 proposes a methodology of building decision support BNs by combining the results of systematic reviews and meta-analyses with knowledge and data. The methodology is illustrated by a BN that predicts the short-term – viability – outcomes of the treatment of mangled extremities. A systematic review and meta-analysis have been conducted to collect information about the factors affecting this treatment.

Chapter 7 proposes a knowledge engineering methodology to derive a BN structure through a series of simplifications. The proposed methodology shows how each simplification step affects the knowledge encoded in the BN.

Chapter 8 proposes a framework to represent clinical evidence behind BNs. The evidence framework is able to organise and present both conflicting and supporting evidence related to fragments, variables and relations in a BN.

Chapter 9 summarises the novel contributions of the thesis, and discusses the future directions of research.

1.3 Publications and Awards

This section shows a list of the publications, conference presentations and awards that are based on this thesis.

Publications

1. Yet B, Marsh DWR (2014) "Compatible and Incompatible Abstractions in Bayesian Networks" *Knowledge-Based Systems*. DOI: 10.1016/j.knosys. 2014.02.020

2. Yet B, Perkins ZB, Fenton NE, Tai N, Marsh DWR (2013) "Not Just Data: A Method for Improving Prediction with Knowledge" *Journal of Biomedical Informatics*. DOI: 10.1016/j.jbi.2013.10.012

3. Yet B, Marsh DWR, Perkins ZB, Tai N, Fenton NE (2013) "Predicting Latent Variables with Knowledge and Data: A Case Study in Trauma Care" *29th Conference on Uncertainty in Artificial Intelligence (UAI-13)*

Applications Workshop: Part-1 Big Data Meets Complex Models, Bellevue Washington, USA, 11-15 July, p.49

4. Perkins ZB, Yet B, Glasgow S, Brohi K, Marsh DWR, Tai N (2013) "Early Prediction of Acute Traumatic Coagulopathy Using Admission Clinical Variables", In Proceedings of the 15th Congress of the European Shock Society, September 12-14 Vienna, *Shock*, 40 (Supp-1) , p. 25, DOI: 10.1097/ SHK.0b013e3182a590b8

5. Yet B, Perkins ZB, Marsh DWR, Fenton NE (2011) "Towards a Method of Building Causal Bayesian Networks for Prognostic Decision Support" *ProBioMed-11 - Probabilistic Problem Solving in BioMedicine 2011*, Bled, Slovenia , 2-6 July, pp. 107-120

6. Yet B, Perkins ZB, Rasmussen TE, Tai N, Marsh DWR "Combining Data and Meta-analysis to Develop Bayesian networks for Clinical Decision Support" submitted to Medical Decision Making.

7. Yet B, Perkins ZB, Tai N, Marsh DWR "Explicit Evidence for Clinical Bayesian Networks" submitted to Artificial Intelligence in Medicine.

Conference Presentations (Abstract Submission)

Yet B, Perkins ZB, Kokuer M, Tai N, Marsh DWR (2012) "Decision Support for Trauma Surgery: Causal Modelling Using Bayesian Networks" *World Trauma Congress 2012*, Rio de Janeiro, Brazil, 2012 22-25 August.

Awards

Our work "Early Prediction of Acute Traumatic Coagulopathy", presented by Mr Zane Perkins, received the Young Investigator Award at the 15th Congress of the European Shock Society. The presentation that received the award was focused on the description and results of the Acute Traumatic Coagulopathy (ATC) BN. The details of the development methodology and validation of the ATC BN are presented in Chapter 5.

Chapter 2

Bayesian Networks

This chapter provides an introduction to Bayes' theorem and BNs. We illustrate the reasoning mechanism and flow of evidence in BNs by a simple example. Next, we describe the mathematical properties of probability distributions and conditional independence in BNs. These properties are necessary to follow the novel methodologies described in Chapters 5 – 8. Finally, we summarise the steps of building BNs and review the existing methods for building BNs from knowledge and data.

2.1 Bayes' Theorem

Bayes' theorem is a simple equation that shows how a conditional probability depends on its inverse conditional probability. According to Bayes' theorem, the probability of an event A conditioned on an event B can be calculated as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' theorem expresses how a prior belief about a probability should change in the light of new evidence. For example, it can be used to update the probability of a diagnosis hypothesis given the observation of a symptom. Suppose that the prevalence of tuberculosis in a particular community is 1%, and 44% of the people in the same community suffers from shortness of breath. By considering the historical patient records, we know that 79% of the patients who had been diagnosed with tuberculosis also suffered from shortness of breath. Although this information tells nothing about the probability of having tuberculosis given that one suffers from shortness of breath, this probability can be calculated using Bayes' theorem.

Let T represent the event 'the patient has tuberculosis' and S represent the event 'the patient has shortness of breath'. The probability of having tuberculosis given that the patient has shortness of breath can be calculated as:

$$P(T|S) = \frac{P(S|T)P(T)}{P(S)} = \frac{0.79 * 0.01}{0.44} \cong 0.02$$

The probability of tuberculosis increased from 1% to 2% when we observe that the patient suffers from shortness of breath.

When we need to apply Bayes' theorem to complex problems with many variables, we can use graphical models called BNs to represent the problem and update the probabilities. The following section introduces the basics of BNs.

2.2 Introduction to Bayesian Networks

BNs are graphical probabilistic models that are composed of a graphical structure and a set of parameters. The graphical structure of a BN contains nodes representing variables and directed edges representing relations between those variables. If a directed edge connects variables A and B as in $A \rightarrow B$, A is called a parent variable and B is called a child variable. Figure 2.1 shows a BN model, known as the Asia BN, which has 8 nodes and 8 edges.

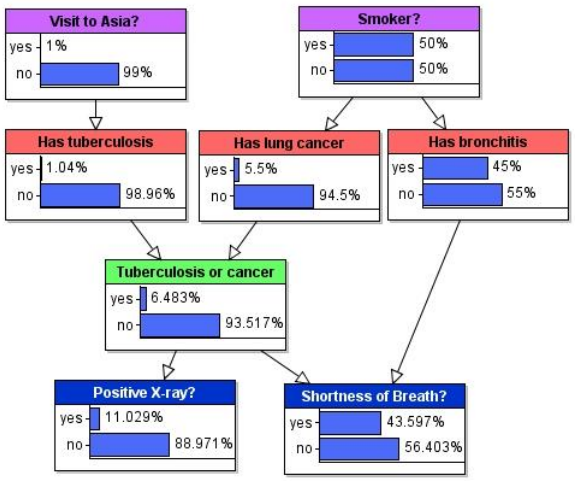


Figure 2.1 Asia BN

Variables in a BN can either be discrete or continuous. Discrete variables are defined by mutually exclusive and collectively exhaustive set of states. All of the variables in Asia BN are discrete variables that have 2 states.

Each variable in a BN has a set of parameters that defines its probabilistic relation with its parents, or its prior distribution if the variable does not have any parents. The parameters of discrete nodes are encoded by node probability tables (NPT). A NPT contains probability values for each state of the variable given every combination of the states of its parent variables. Table 2.1 shows the NPT of the 'Has tuberculosis' variable in the Asia BN. The NPT has 4 probability values since the variable has 1 parent, and both the variable and its parent have 2 states each.

Table 2.1 NPT of the 'Has tuberculosis' variable

		Visit to Asia?	
		Yes	**No**
Has tuberculosis	**Yes**	0.05	0.01
	No	0.95	0.99

The probability distributions of continuous variables can be defined by using statistical distributions or functions of their parent variables (see Fenton and Neil (2012a; 2012b) for a thorough introduction to modelling with discrete and continuous variables in BNs). In the following chapter, we illustrate how BNs reason by an example about the Asia BN.

2.3 Reasoning with Bayesian Networks

Mr John Doe has been suffering from an unusual shortness of breath lately. He cannot stop worrying about the possibility of having cancer even though he tries to reassure himself by thinking of more common causes of this condition such as bronchitis. Eventually, he decides to visit a clinician to get a diagnosis. The clinician uses the Asia BN (see Figure 2.1) as a decision support tool to diagnose Mr Doe's condition. The clinician initially considers 3 disease hypotheses: cancer, tuberculosis and bronchitis. The BN model has a variable representing each of these hypotheses ('Has tuberculosis', 'Has lung cancer', 'Has bronchitis'), and it can make

probabilistic calculations about the hypotheses based on information entered to the model.

First, the clinician asks about Mr Doe's symptoms, and recalculates the probabilities as he enters the symptom about shortness of breath. Bronchitis is the most probable hypothesis at this stage (see Figure 2.2a). The clinician demands a chest x-ray as he does not want to misdiagnose a life-threatening disease such as tuberculosis or cancer. The result of the x-ray turns out to be positive which makes the clinician more worried about the cancer hypothesis (see Figure 2.2b).

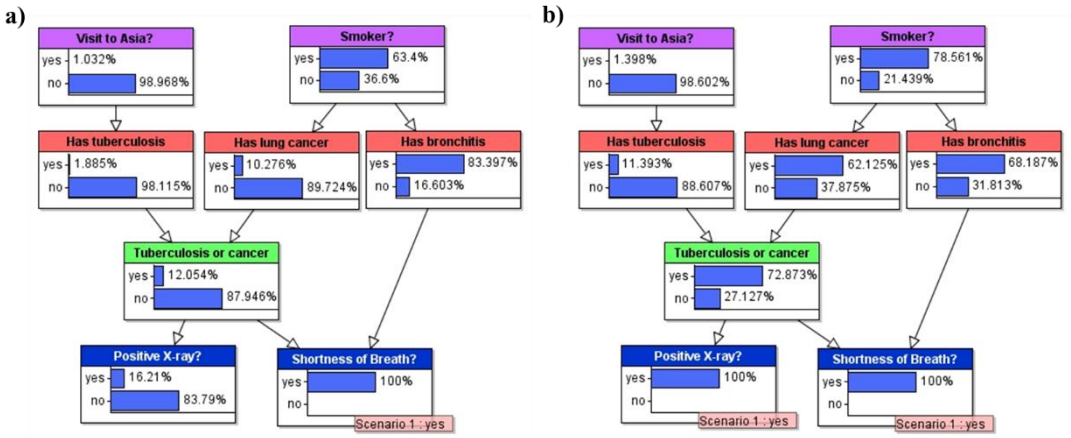


Figure 2.2 Probabilities updated after observing a) symptoms and b) X-ray

In order to collect more information, the clinician asks questions about the cancer cases in Mr Doe's family and his smoking habit. Mr Doe says that he does not smoke regularly but he smoked a few cigarettes at his recent holiday in Cambodia. The trip to Cambodia may be an important piece of information for the tuberculosis hypotheses at this stage of treatment as tuberculosis is more prevalent in this country (World Health Organization, 2012). The probability of cancer is much lower after the information about smoking and visit to Asia is entered, and tuberculosis, which initially had a low prior probability, is now a more convincing diagnosis (see Figure 2.3a and Figure 2.3b).

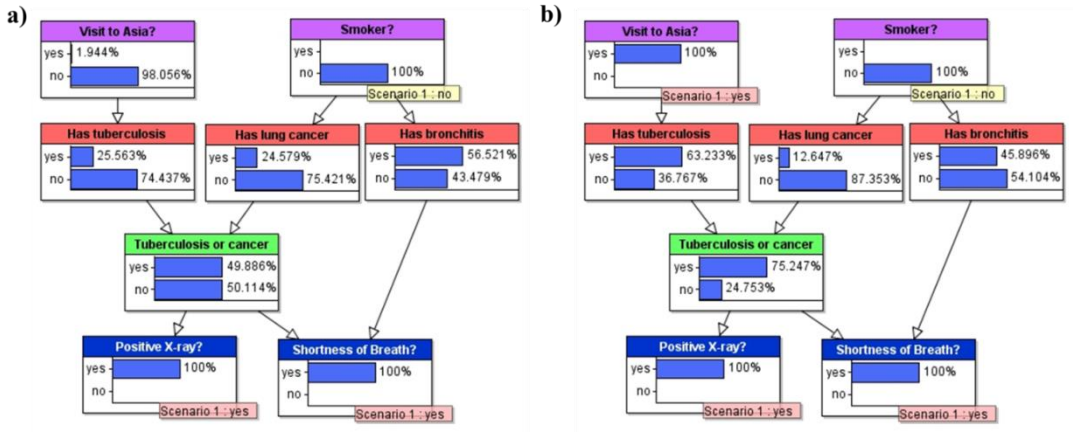


Figure 2.3 Probabilities updated after observing a) smoking history and b) visit to Asia

This simple example illustrates the 3 ways that BNs propagate information to update probabilities:

1. *Causal reasoning:* Entering an observation to a 'cause' node will update the probabilities in its 'effect' nodes. In Asia BN, knowing the patient's visit to Asia increased the probability of tuberculosis

2. *Diagnostic reasoning:* Entering an observation to an 'effect' node will update the probabilities in its 'cause' nodes. For example, observing the patient's shortness of breath increased the probability of bronchitis

3. *Explaining away:* If any of the 'effect' nodes or their descendants is observed, entering an observation to a 'cause' node will update the probabilities of the other 'cause' nodes. For example, after knowing the results of the x-ray and the presence of shortness of breath, knowing the visit to Asia increases the probability of tuberculosis and decreases the probability of cancer, which is the other cause of a positive X-ray and shortness of breath. In other words, a higher probability of tuberculosis, resulting from the trip to Asia, explained away the other causes of the positive X-ray and shortness of breath. Such flow of information would not happen if the state of the X-ray result and shortness of breath were not known.

The following section presents a formal definition of BNs and their conditional independence (CI) properties.

2.4 Condition Independence and Bayesian Networks

A BN can represent a joint probability distribution compactly in a factorised way. The graphical structure of a BN is a directed acyclic graph that encodes a set of CI assertions about its variables. Every node in a BN is independent of its non-descendants given that the state of its parents is known. Therefore, each node has a conditional probability distribution (CPD) that defines its probabilistic relation with its parents. A probability distribution P_X factorises over a BN structure G_X if P_X can be decomposed into the product of factors $P_X = P(X_1, \dots, X_n) = \prod_{i=1}^{n} P\left(X_i \middle| PA_{X_i}^{G_X}\right)$ where $X_1, \dots, X_n$ are a set of variables, $PA_{X_i}^{G_X}$ is the set of parents of X_i in G_X.

The CIs that can be encoded in a BN can be shown by the relation between three variables.

1. If two variables, A and B, are directly connected by an edge, as shown in Figure 2.4a, a BN does not assert any CI conditions between these variables.

2. If there is a serial relation between three variables A, V and B, as shown in Figure 2.4b, then A and B becomes independent given that the state of V is known.

3. If there is a diverging relation between A, V and B, as shown in Figure 2.4c, A and B becomes independent given that the state of V is known.

4. If there is a converging relation between A, V and B, as shown in Figure 2.4d, A and B are independent. However, this independence disappears if the state of V or one of its descendants is known.

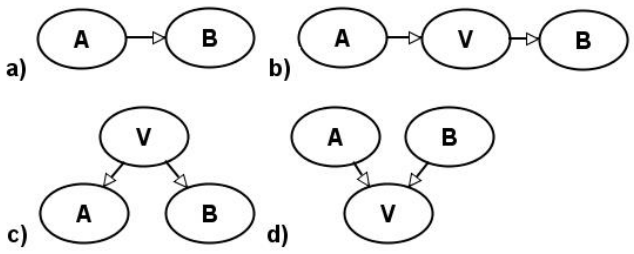


Figure 2.4 (a) Direct Connection (b) Serial Relation (c) Diverging Relation (d) Converging Relation

In general, CI assertions of a BN can be determined by d-separation (Pearl, 1988):

d-separation: A trail $X_1 \leftrightharpoons \cdots \leftrightharpoons X_n$ is a consecutive sequence of edges that can be in any direction. Let G be a BN structure, A, B and V be a three disjoint sets of nodes in G. A and B are d-separated by V, $dsep_G(A; B|V)$, if and only if there is no *active trail* between A and B given that V is observed. An active trail requires the following conditions:

1. For every converging relation $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ in the trail, the node X_i or one of its descendants is a member of V.
2. The other nodes in the trail are not members of V.

If A and B are d-separated given V in the BN structure G, then A and B are conditionally independent given V in any probability distribution that factorises over the BN. A and B are called d-connected if they are not d-separated. It follows from the definition of d-separation that adding an edge to a BN increases the number of trails and therefore does not increase the number of CI conditions.

A BN structure G asserts a set of conditional independencies $I(G)$. P can factorise on G if $I(G)$ is a subset of $I(P)$, i.e. the set of conditional independencies in P. Such G is called an I-map of P.

$$G \text{ is an I-map of } P \text{ if and only if } I(G) \subseteq I(P)$$

Any CI that holds on the BN structure G must also hold on the probability distribution P, if P factorises over G. On the other hand, P can have additional CI conditions that are not reflected in G. Therefore, a probability distribution can factorise over various BN structures.

An example of this situation can be seen by the two BNs in Figure 2.5. Some probability distributions can factorise on both of these BNs even though their graphical structure is different. In the BN in Figure 2.5a, as well as in the probability distribution P that factorises over this BN, A and B are conditionally independent given that the state of C is not known. This CI is not represented in the graphical structure of the BN in Figure 2.5b. However, the CI condition can still be present in the probability distribution that factorises over this BN structure. In other words, the

CI between A and B can be encoded in the parameters of this BN rather than its structure. The BN on the left is preferable since an edge between A and B is unnecessary for this probability distribution, and additional edges increase the computational burden of a BN. The obvious conclusion is to choose a BN structure that encodes all of the independencies of the probability distribution in its graphical structure. Unfortunately, this is not possible in general. Symmetric variable-level CIs or some regularities in the parameters do not have a BN structure that represents all of the CIs (Pearl, 1988).

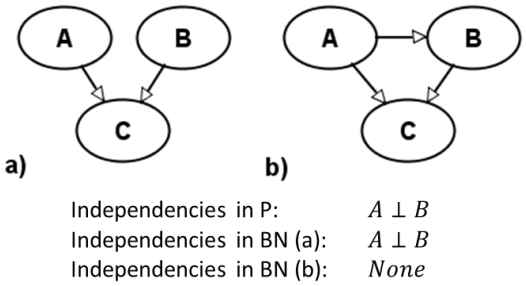


Figure 2.5 Same Probability Distribution Factorised over Two Different BN Structures

2.5 Features of Bayesian Networks

Algorithmic breakthroughs in the 1980s (Lauritzen and Spiegelhalter, 1988), and more recent advances for using continuous variables (Neil et al., 2007), have made it possible to calculate inferences on a large number of continuous and discrete variables in BNs. The strengths of BNs in knowledge representation and probabilistic reasoning made them an attractive tool for providing decision support in a wide variety of domains including medicine (Lucas et al., 2004), finance (Neil et al., 2009), law (Fenton, 2011; Fenton et al., 2013), sports (Constantinou et al., 2013, 2012), reliability (Marquez et al., 2010) and safety (Bearfield and Marsh, 2005; Marsh and Bearfield, 2004). The benefits that BNs offer include:

- **Knowledge Representation:** BNs have a graphical structure that is well-suited for representing causal relations. This makes it possible to encode domain knowledge about the causal and associational relations in the BN structure. Unlike the statistical – black box – approaches, the reasoning and predictions of a BN can be explained as its graphical structure can be built in

a way that make sense to domain experts.

- **Causal Structure:** The structure of knowledge based BNs are often built based on causal relations since 1) this is a natural way of expressing knowledge by domain experts (Fenton and Neil, 2012c; Lucas, 1995) 2) probability distributions can be represented more sparsely this way (Koller and Friedman, 2009a). Moreover, variables that are important in a domain but not available in data can be modelled using causal relations elicited from domain experts. Causal BNs also makes it possible to distinguish between observations and interventions allowing analysis of interventions and counterfactuals (Pearl, 2000).

- **Missing Observations:** Statistical models, such as regression models, require values for all of the independent variables in the model to calculate the value of the dependent variable. The predictions cannot be generated when some of the values are missing. BNs, on the other hand, have no specific set of variables that must necessarily be observed. A BN can calculate the posterior probability distribution of its unknown variables whenever an observation is entered to any of its variables. When additional observations are entered, the BN updates the probability distribution based on the new information.

- **Flow of Information:** When a variable is observed in a BN, it can update the probability distribution of its both 'cause' and 'effect' variables. Information can flow both forwards and backwards in BNs allowing both causal and diagnostic reasoning as shown in the Asia BN example (see Section 2.1). Moreover, when the state of an 'effect' variable is known, observing the state of its causes can be used to update the probability of the other – unobserved – causes. This type of reasoning is crucial for making what-if analysis and cannot be done by statistical models such as multivariate regression.

- **Probability Distribution:** BN can represent probability distributions compactly in a factorised way as every variable is conditioned on its parents (see Section 2.4). In other words, probability distributions can be defined more sparsely in BNs therefore requiring less data and expert resources for

their parameters.

2.6 Building Bayesian Networks

A BN can be built in two steps:

1. **Structure:** The structure of the BN is defined in the first step. This involves identifying the set of variables that are important in the problem domain, and defining the set of states for each of these variables. Afterwards, the relations between the variables, and the directions of those relations are defined.

2. **Parameters:** The parameters, representing the strength of the relations in the BN structure, are defined in the second step. If a variable and its parents are discrete, a probability is defined for each probability value in the NPT of the variable. If continuous variables are present, a statistical distribution and the necessary parameters are defined.

Both BN structure and parameters can be learned from data, elicited from experts or estimated by a combination of them. In the remainder of this section, we review the existing methods for defining the structure and parameters of a BN.

2.6.1 Knowledge Engineering Methods

Structure

The recommended way of modelling the correct probability distribution and CI is to model the causal relations in a BN structure (Fenton and Neil, 2012c; Koller and Friedman, 2009a). However, eliciting a causal structure can be challenging, and assistance can be required, especially when a large number of variables and complex relations need to be modelled.

Neil et al. (2000) use specific BN fragments called idioms for representing common types of uncertain reasoning. Knowledge engineers and domain experts select the most appropriate idioms for their modelling problems and use these idioms as building blocks for their BN structure. Idioms are reused for the similar modelling tasks in order to develop BNs efficiently and consistently.

Koller and Pfeffer (1997) describe object-oriented Bayesian networks (OOBN) language that represents BNs with inter-related objects. OOBN are particularly useful for complex models that contain repeated fragments, where objects can be reused to decrease the modelling effort. Laskey and Mahoney (1997) also use object-oriented concepts to construct a BN by using semantically meaningful fragments as basic building blocks.

Nadkarni and Shenoy (2004) use a managerial tool, called causal maps, to capture causal information from domain experts. The causal map is transformed into a BN by assuming that it represents the dependency-map of the probability distribution.

Laskey and Mahoney (2000) propose a system engineering approach that uses a spiral lifecycle model for BN development. Their approach starts by defining objectives and building initial prototypes with simple features. These prototypes are evaluated and rebuild. This process helps a knowledge engineer understand the domain and a domain expert understand the principles of BN modelling. The systems engineering approach uses network fragments (Laskey and Mahoney, 1997) as basic elements of BN development.

Heckerman (1990) describes similarity networks that can be used for diagnosing a single hypothesis that has mutually exclusive and exhaustive states. In this approach, each pair of similar hypotheses is connected in a similarity network. A separate BN network structure is elicited for each pair of these similar hypotheses. Then, the separate BN structures are merged to form the final BN structure. This approach divides the task of network building into pieces that are easier to manage. However, it can only be applied when the hypotheses are mutually exclusive and exhaustive, and the hypothesis variable has no parents.

Abstraction methods for simplifying an expert elicited BN has also been proposed. However, most of the methods have been designed for a specific problem and are not generalisable to a wider range of problems. Srinivas (1994) proposes a hierarchical BN approach for fault diagnosis in engineering systems. In this approach, functional schematics are defined in multiple levels of abstraction between the inputs and outputs of the system. Shachter's topological operations (1986) are used to reach to

higher level schematics. The different abstraction levels of schematics must have the same inputs and outputs.

Wu and Poh (2000) propose a set of operations that changes the abstraction level of a knowledge-based influence diagram. They propose the 'extend' and 'retract' operations to add and remove the parents of a variable. The 'abstract' operation merges a set of variables that shares a single parent and child. The 'refine' operation is the opposite of the 'abstract' operation. These operations can be applied to a limited variety of modelling tasks. For example, Wu and Poh (2000) do not discuss how to apply the 'abstract' operation to variables that do not share the same parent or that have multiple parents.

Parameters

The parameters of a BN can be elicited from domain experts without using any data. Several direct and indirect methods have been proposed to elicit probabilities including the use of probability scales and lotteries (Korb and Nicholson, 2004a; Renooij, 2001; Van der Gaag et al., 2002). Probability elicitation is a challenging task as domain experts display various kinds of biases while expressing probabilities (see Tversky and Kahneman (1974) and O'Hagan et al. (2006) for a detailed discussion of these issues). Methods to overcome these biases can take too much time and make it infeasible to elicit a large number of parameters from domain experts (Renooij, 2001).

Parameters elicited from experts can be refined by sensitivity analysis methods (Coupé et al., 2000, 1999). In this approach, a knowledge engineer selects a target variable and examines the changes in the marginal probability distribution of this variable by systematically changing other parameters. This can be computationally expensive especially when the parameters of multiple variables are changed simultaneously (Coupé et al., 2000). Other types of sensitivity analysis exists for analysing the effects of observations and edge removals (Korb and Nicholson, 2004b; Renooij, 2010).

There are several techniques for decreasing the number of parameters in a BN. These techniques can be used to reduce the number of parameters that needs to be elicited

from experts. The parameter space of a variable grows rapidly as it has more parent variables. Adding an intermediate variable between the variable and its parents can reduce the size of its parameter space. This approach is known as 'parent divorcing' (Nielsen and Jensen, 2007). Canonical models, such as Noisy-OR and Noisy-Max gates, are also used for simplifying the elicitation task (Diez and Druzdzel, 2006; Henrion, 1987; Pearl, 1988; Pradhan et al., 1994). These models decrease the number of parameters in a NPT by assuming that the effect of each parent variable is independent from other parents. For example, Noisy-OR (Pearl, 1988) assumes that the presence of any of the causes is enough for the presence of the effect but there is a possibility that some of the causes may fail to produce the effect as indicated by the term 'Noisy'. Parent divorcing and canonical models can be used together with parameter learning approaches when data is not large enough.

Ranked nodes can simplify parameter elicitation for variables with ordinal scale (Fenton et al., 2007). A ranked node is an approximation of the truncated normal distribution to the multinomial distribution with ordinal scale. Fenton et al. (2007) provides a framework for using ranked nodes for parameter elicitation in BNs. In this approach, parameters are defined by 1) selecting a suitable ranked node function for modelling the relation between the variable and its parents, 2) eliciting the weights required for the ranked node function from domain experts, 3) eliciting the expert's degree of confidence in these weights. The ranked nodes offer the possibility of modelling a wide range of relations for the variables with ordinal scale. Moreover, a ranked node requires fewer parameters compared to a complete NPT therefore the elicitation task requires significantly less effort (Fenton et al., 2007). However, selecting a suitable function for the elicited relation can be challenging as it demands thorough understanding of the behaviour of different ranked node functions.

2.6.2 Data Based Methods

Structure

Structure learning algorithms for BNs can be divided into two categories: constraint-based algorithms and score-based algorithms. Constraint-based algorithms aim to determine CIs in the dataset, and build a structure satisfying these CIs. The tests required for determining CIs may become computationally infeasible as the BN gets

larger. Therefore, these algorithms make several simplification assumptions such as limiting the maximum number of parents that a variable can have.

A common statistical test for identifying CIs in data is the χ^2 test. This test calculates the false-rejection probability of a CI hypothesis. The mutual information measure, which is mathematically related to the χ^2 test, is also used for testing the same hypothesis. A more recent CI test for constraint-based learning is developed by Dash and Druzdzel (2003). A non-parametric test is proposed by Margaritis (2004).

Constraint-based algorithms such as IC (Pearl and Verma, 1991) can learn a part of the causal relations from data. However, the true – complete – causal structure is not identifiable from the data. Even if a learning algorithm identifies all of the CIs in the probability distribution, it may not find the true causal structure as multiple BN structures can represent the same probability distribution (see Section 2.2). Moreover, since data is noisy, we may never be sure about the CIs identified by the learning algorithm. Notable constraint-based structure learning algorithms include IC (Pearl and Verma, 1991), LCD (Cooper, 1997), PC (Spirtes et al., 2001), Grow-shrink (Margaritis, 2003) and TDPA (Cheng et al., 2002).

Scored-based algorithms aim to find the BN structure that maximises a likelihood score. Adding edges to a BN increases the likelihood of representing the probability distribution but it can also reduce the quality of parameter estimation by dividing the data. Therefore, the scoring functions for these algorithms are often a combination of the goodness of fit and penalty for additional edges. Commonly used scoring functions include the Bayesian information criterion (Cruz-Ramírez et al., 2006; Schwarz, 1978), minimum description length (Lam and Bacchus, 1994), minimum message length (Wallace et al., 1996; Wallace and Korb, 1999) and BDe score (Heckerman et al., 1995).

Based on the selected scoring function, a score-based algorithm searches the space of possible BN structures to find the structure with the maximum score. The search is done by removing, adding or reversing edges between the variables available in data. The algorithms can either search the space of singular BN structures or the space of equivalent structure classes. Notable search algorithms include Cooper and

Herskovits (1992), Glover and Laguna (1997), Chickering (2003; 1996), Chickering and Meek (2002), and Castelo and Kocka, (2003).

Tsamardinos et al. (2006) proposed a combination of score-based and constraint-based methods for structure learning. Their algorithm, called max min hill climbing (MMHC), defines a skeleton for the BN structure based on a constraint-based method, and orients the edges in the skeleton by maximising a scoring function.

Structure learning is more complicated when missing values exist in the data. Calculation of the scoring functions becomes more difficult as these functions do not decompose when missing values exist. Daly et al. (2011) and Koller and Friedman (2009b) provide a thorough review of structure learning methods for complete and incomplete data.

Parameters

A popular approach for parameter learning is to find the parameters that maximises the likelihood of the model given the data. For discrete variables, the maximum likelihood estimates can be found by calculating the related conditional probabilities in the data. Replacing zero probabilities with small values can increase the performance of the model in other datasets. Parameters can also be estimated by a Bayesian approach, which uses a prior distribution, representing the background knowledge, for the parameters and updates the prior based on data. Bayesian approach can provide better results especially for small datasets as it includes expert knowledge into parameter learning.

Parameter learning becomes more difficult when data contains missing values. A simple way to deal with missing values is to complete the data by assigning values to them. The values can be assigned randomly, sampled from a distribution or estimated from the data. This approach is called imputation in statistics. After the missing values are assigned, standard parameter learning methods can be used.

Expectation-maximisation (EM) is an iterative algorithm that uses the BN structure to deal with missing values (Lauritzen, 1995). EM starts with assigning initial values either to the BN parameters or to the missing values. In each iteration, EM calculates the parameters based on expected values of the missing values, and it updates the

expected values based on the new parameters. EM is guaranteed to converge to a local maximum. EM has also been applied to learn the parameters of canonical models such as noisy-OR (Meek and Heckerman, 1997).

Bayesian learning can also be used for datasets with missing values. While calculating the posteriors in Bayesian learning is often trivial for complete datasets, it becomes computationally expensive, and sometimes infeasible, when missing values are present. In complete datasets, the parameters of different CPDs are independent of each other, and the posterior often has a compact form that can be solved analytically. However, the parameters become correlated when missing values exists. A thorough introduction to Bayesian parameter learning with complete and incomplete data is presented by Koller and Friedman (2009c; 2009d).

2.6.3 Hybrid Methods that Combine Knowledge and Data

Previous sections discussed several limitations of purely data and knowledge driven techniques. Methodologies that combine data and expert knowledge seek to overcome these limitations by using all available information in BN development. However, research in this area is still in early steps, and there are many challenges that need to be addressed.

Structure

Flores et al. (2011) proposes a method that integrates expert's opinion about the presence and direction of the arcs into structure learning. In this method, experts can define the type of the relations between variables and assign a prior probability representing their confidence. For example, an expert can say that he is 80% confident that there is a direct relation between two variables but he is not sure about the causal direction of this relation. The expert can also define other types of relations including direct causal connection, causal dependence, temporal order, and correlation. Afterwards, the BN structure is learned based on these expert priors using a score-based method.

Cano et al.'s method (2011) uses expert judgement during the learning process instead of using it as priors. A Bayesian score is used for the learning algorithm. The

arcs that have the most uncertainty, according to the learning algorithm, are shown to experts. Afterwards, the experts make the final decision about the presence and direction of these arcs. This approach can decrease the time spent by experts since their opinion is only used for the most uncertain BN elements.

Velikova et al. (2013) uses structure learning methods as a complementary approach to evaluate and refine the BN structure built with experts. Antal et al. (2004) proposed a method for combining data and textual information from the medical literature to build BNs. They use information retrieval techniques to assist structure learning based on the textual information in medical literature.

Parameters

Bayesian learning methods can integrate expert knowledge into parameter learning by using informative priors. However, eliciting numbers for informative distributions can be difficult as experts often feel less confident in expressing quantitative statements (Druzdzel and Van Der Gaag, 2000). Therefore, using qualitative constraints, such as "value of A is greater than value of B", can be more convenient. Zhou et al. (2013a, 2013b) proposed a technique for integrating expert knowledge as constraints when learning multinomial parameters from data. Similar approaches are also proposed by Feelders and Van der Gaag (2006) for binomial parameters, by Tong and Ji (2008) for a limited amount of constraints, and by Khan et al. (2011) for diagnostic BNs.

2.6.4 Knowledge Gap

The knowledge engineering and machine learning communities has focussed less on hybrid methodologies compared to purely knowledge or data driven approaches. Although the number of studies about hybrid methodologies has been increasing in recent years, many of these studies have addressed similar challenges. From the reviewed studies, the hybrid structure learning methods mainly focus on using knowledge to assist a data-based structure learning algorithms. The hybrid parameter learning studies mainly focus on using knowledge as constraints for parameter learning. Combination of knowledge and data also has potential benefits in other challenges of BN modelling that need to be addressed. For example, BNs that reason

consistent with knowledge often contains variables that cannot be directly measured and thus not available in the dataset. Hybrid methodologies that combine knowledge and data are required to deal with this task. In the following chapter, we discuss the application of knowledge and data driven techniques for medical models. In Chapter 4, we introduce the medical case study and, by using the case study, we illustrate several modelling challenges that can be dealt with novel hybrid methodologies.

Chapter 3

Clinical Decision Support

The nature of medical decision making is inherently uncertain as the true state of a patient can almost never be observed (Sox and Higgins, 1988). A clinician can interview the patient, examine his physical conditions and conduct laboratory tests but these may not necessarily reveal the true state of the patient. The findings of these tests are often related to more than one disease therefore they can only help in ruling out some diseases and decreasing the uncertainty regarding the diagnosis. Selecting the best treatment is also not simple because of the uncertainty. The treatment options often have different benefits and disadvantages; an optimal decision, that is better in all aspects, may not exist. According to Sox and Higgins (1988), clinicians generate hypotheses about the patient's problem often in early stages of patient care. They compare only a few hypotheses and gather more information to confirm or falsify them. In this regard, medical decision making follows the principles of Bayesian reasoning.

There are systematic errors that clinicians, and other experts, make when they reason with uncertainty (Tversky and Kahneman, 1974). The classic study by Casscells et al. (1978) and the cases by Gigerenzer (2003) show examples of striking errors that clinicians make when calculating probabilities. Quantitative models, such as BNs, can be helpful in preventing these errors.

In this chapter, we examine the potential benefits of quantitative models for providing clinical decision support. We review the pitfalls of the existing approaches to examine why many of these models have not been employed by clinicians.

3.1 Surgical Decision Making

Most surgical decisions are made in challenging conditions. Surgeons often face uncertainty and strict time constraints when making decisions with critical and irreversible outcomes. A branch of decision science, called naturalistic decision making, studies decision making under such conditions (Klein, 2008; Lipshitz et al., 2001). Other domains of naturalistic decision making include piloting and firefighting (Klein et al., 1986; Orsanu and Fischer, 1997).

Flin et al. (2007) examined surgical decision making using concepts from other domains of naturalistic decision making. They defined surgical decision making in two stages. In the first stage, a surgeon focuses on situation awareness. He observes the overall situation of the patient, diagnoses the anatomical and physiological disorders, and evaluates the expected outcomes and risks. In the second stage, the surgeon adopts one of the following decision making strategies to select a treatment that maximises the expected outcomes:

1. **Recognition:** The surgeon recognises the problem and recalls a treatment used in a similar problem.
2. **Rule-based:** The surgeon selects the treatment recommended by the guidelines.
3. **Analytical:** The surgeon evaluates multiple treatment options simultaneously and selects the one with maximum benefit.
4. **Creative:** The surgeon uses a novel course of action for an unfamiliar situation.

Surveys of experienced surgeons showed that they adopt the recognition and analytical decision making strategies more commonly than the rule-based and creative strategies (Pauley et al., 2011). The recognition strategy is preferred when the operation is familiar to the surgeon or when there is only one plausible option (Jacklin et al., 2008; Pauley et al., 2011). The analytical strategy is preferred when there are multiple treatment options with similar risks and benefits. The analytical strategy is more common in surgery than in other domains of naturalistic decision

making (Cristancho et al., 2013; Klein et al., 1993, 1986). The rule-based strategy is more commonly preferred by novice surgeons (Pauley et al., 2011); experienced surgeons occasionally adopt this strategy mainly for routine surgical operations (Jacklin et al., 2008). The creative strategy is not considered suitable for most surgical operations because of the risks and time constraints involved, and thus it is rarely adopted (Flin et al., 2007; Pauley et al., 2011).

Quantitative models offer the potential to improve two areas of surgical decision making. First of all, the situation awareness stage can be improved by models that calculate risks and probabilities. Way et al. (2003) showed that misperception of risks leads to poor outcomes in many surgical operations even when the surgical skill and judgement is adequate. Experts make errors when reasoning with uncertainty (Casscells et al., 1978; Gigerenzer, 2003; Kahneman, 2011; Tversky and Kahneman, 1974), and uncertainty is often abundant at the situation awareness stage (Flin et al., 2007; Way et al., 2003). Quantitative models can be used at the situation awareness stage to quantify uncertainty when evaluating probabilities and risks. Secondly, the analytical decision making strategy can be assisted by quantitative models. Experts who adopt this strategy need to calculate the expected outcomes of the available hypotheses. The calculations have to be made iteratively as the observed state of the patient changes. Experts have to deal with the uncertainty regarding these calculations. Moreover, decision science research showed that experts can consider only a few hypotheses at a time (Kahneman, 2011; Sox and Higgins, 1988). Quantitative models can be used to assist analytical decision making strategy by enabling calculation of a larger number of hypotheses.

3.2 Statistical Modelling Approaches in Medicine

Models that predict the course of a disease or a medical condition, based on a single or multiple variables, are called prognostic models in the medical literature. Typically, the relation of the predictors to the model outcome is analysed by multivariate statistical models or similar approaches (Abu-Hanna and Lucas, 2001). The accepted way of selecting predictors is to adjust the variables and check their effects on the outcome in data. If an adjustment of a variable is connected to the outcome with statistical significance, the variable can be called an independent

predictor (Royston et al., 2009). The danger is that correlation is confused with causation. For example, grey hair is an independent risk factor for heart disease, however, if two men of the same age but different hair colours are considered, grey hair does not probably increase the heart disease risk (Brotman et al., 2005). Therefore, the independent predictors are not necessarily causal factors; they are the factors that are correlated with causal factors according to the available data and selected variables. The number and identity of the included variables is sometimes considered to be not important (Katz, 2003; Moons et al., 2009b). Consequently, the independent predictors and their relations to outcome can be completely different between studies. Jenks and Volkers (1992) shows more extreme examples about variable selection where electric-razors or owning refrigerators have been identified as risk factors for cancer.

Although, a large number of prognostic models are developed and published, the majority are not adopted into clinical practice (Altman et al., 2009). The predominant reason for this are concerns regarding model accuracy. Accuracy alone, however, does not ensure use of a model. Predictors with different sets of variables can be statistically accurate but statistical accuracy of a model does not ensure its clinical acceptance (Moons et al., 2009a) and there are now widely accepted arguments against the use of statistical significance tests and their associated p-values (Goodman, 1999; Ziliak and McCloskey, 2008). On the other hand, some models with mediocre performance are widely used in clinical practice (Moons et al., 2009a).

Clinicians tend to reject a prognostic model if they are not convinced that the model's performance, for their patients, will be similar to its published performance in validation studies (Moons et al., 2009a; Wyatt and Altman, 1995). The clinical evidence supporting the model and its reasoning mechanism must be understood for clinicians to evaluate its prospective performance in their practice (Akhtar and Forse, 2010; Wyatt and Altman, 1995). It may be necessary to modify an existing prognostic model because of the changes in clinical knowledge and practice. In this case, the model is often retrained from scratch with the new data. This approach discards all of the information in the previous model even though a part of the model may still be relevant (Moons et al., 2009a). This can be avoided by identifying the

obsolete parts of the model and refining only those parts. The link between clinical knowledge and the reasoning mechanism of the model must be clear for clinicians to identify and modify the obsolete parts. However, most statistical models are linear equations that represent the correlations in their training dataset. The structure of these models is often not intuitive to clinicians as they do not reflect causal relations. For example, it is difficult, if not impossible, to modify a regression model to include a new clinical factor that is known to affect the outcome and some of the independent variables through multiple causal pathways.

Wyatt and Altman (1995) argue that useful prognostic models have 4 properties in common: clinical credibility, accuracy, generalisability, and ability to provide useful decision support. Traditional – purely data-driven – modelling approaches often target only one of these qualities: statistical accuracy. They disregard domain knowledge about the clinically relevant variables and their causal relations which is often available in abundance. As a result, evidence supporting the model becomes limited to its training data. Using knowledge from other sources, as well as data, can be an important step to achieve all four properties. In the following section, we review knowledge and data driven artificial intelligence (AI) approaches for developing clinical models.

3.3 Artificial Intelligence and Bayesian Networks in Medicine

Medicine has been a popular domain for applying AI techniques. The early applications tried to imitate decision making of clinicians using a set of rules defined in their system. MYCIN system was arguably the first successful application in medicine (Buchanan and Shortliffe, 1984; Shortliffe, 1976). It was developed in the 1970s to recommend treatments for infectious blood diseases based on a knowledge base of about 500 rules. MYCIN performed better than most clinicians who were not specialised in infectious blood diseases (Yu et al., 1979). Other notable expert systems include Internist (Miller et al., 1982), Casnet (Weiss et al., 1978) and ABEL (Patil et al., 1981). The early expert systems had two major disadvantages. First, by imitating the decision making of experts, these systems also imitated its undesirable

flaws (Druzdzel and Flynn, 2009). Second, their rule-based reasoning mechanism could not represent the uncertain nature of medicine. As a result, most expert systems have not been widely used in clinical practice.

Machine learning (ML) is a branch of AI that focuses on learning systems purely from data. Following the advances in computing technology, ML applications have become increasingly popular. They have had successful results in some medical areas where data is available in large amounts. These include analysis of radiography images (Savage, 2012) and identification of the patterns between genes and diseases (Shipp et al., 2002). However, just like statistical models, ML methods have not always provided useful models for complicated clinical decisions even when a large amount of data is available.

According to Buchan et al. (2009), this is because the complexity of clinical mechanisms is not taken into account by purely data-based approaches. Knowledge, from clinicians and published evidence, should be used to uncover the clinical mechanisms related with data, and the data should be used on top of this to reflect the complexity. Patel et al. (2009) confirmed this in a recent panel between the leading researchers of AI in medicine: combining knowledge and data offer the potential to be useful in areas that purely knowledge-based or data-based approaches fail (Holmes and Peek, 2007; Patel et al., 2009; Zupan et al., 2006).

BNs are well-suited for combining knowledge and data (see Chapter 2 for a review of knowledge and data driven approaches for BNs). Complex clinical mechanisms with multiple pathways can be represented in the structure of BNs (Fenton and Neil, 2010). Moreover, the probabilistic reasoning of BNs can effectively deal with uncertainty and unobserved variables, both of which are not rare in medical decision making.

BNs have been a popular approach in medicine for more than 20 years (see Table 3.1) (Abu-Hanna and Lucas, 2001; Lucas et al., 2004). Many of the early applications of BNs were built purely by knowledge. One of the first large-scale applications in medicine, the Pathfinder project, was aimed to diagnose 60 diseases of lymph nodes (Heckerman and Nathwani, 1992; Heckerman et al., 1989). Both the structure and parameters of Pathfinder were elicited from experts. It was

commercialised as a training tool, called IntelliPath, for junior pathologists (Nathwani et al., 1990).

Alarm BN is one of the earliest applications in emergency medicine (Beinlich et al., 1989). Its aim is to diagnose patient disorders and generate alarm messages based on the inputs from the patient monitoring devices. The BN structure and CPDs were elicited from experts. Data was used for learning the probability distributions of variables without parents.

Table 3.1 Some Knowledge and Data Driven Applications of BN in Medicine

Purely Data-Based	Expert Knowledge Used in Modelling
Arteriosclerosis (McGeachie et al., 2009)	Anticoagulant Treatment (Yet et al., 2013a)
Breast Cancer (Cruz-Ramírez et al., 2007)	Breast Cancer (Wang et al., 1999)
Cardiac Surgery (Verduijn et al., 2007)	Chronic Obstructive Pulmonary Disease (van der Heijden et al., 2013)
Chronic Obstructive Pulmonary Disease (Himes et al., 2009)	Echocardiography (Díez et al., 1997)
Clinical Therapeutics (Nordmann and Berdeaux, 2007)	Electromyography (Andreassen et al., 1989)
Colorectal Cancer Surgery (Nissan et al., 2010)	Gastric Lymphoma (Lucas et al., 1998)
Head Injuries (Sakellaropoulos and Nikiforidis, 1999)	Intensive Care (Beinlich et al., 1989)
Intensive Care (Celi et al., 2008; Crump et al., 2011)	Infectious Diseases (Charitos et al., 2009; Lucas et al., 2000; Schurink et al., 2007, 2005; Visscher et al., 2008)
Malignant Skin Melanoma (Sierra and Larrañaga, 1998)	Liver Disorders (Onisko et al., 1998; Wasyluk et al., 2001)
Mortality Prediction (Celi et al., 2008)	Multidisciplinary Team Meetings (Ogunsanya, 2012)
Radiography Interpretation (Maskery et al., 2008; Neumann et al., 2010)	Multiple morbidities (Lappenschaar et al., 2013)
Venous Thromboembolism (Kline et al., 2005)	Nasogastric Feeding Tube Insertion (Hanna et al., 2010)
	Oesophageal Cancer (Helsper and van Der Gaag, 2007; Van der Gaag et al., 2002)
	Pathologic Disorders (Heckerman and Nathwani, 1992)
	Prostate Cancer (Lacave and Díez, 2003)
	Pyloric Stenosis – Pediatric Surgery (Alvarez et al., 2006)
	Radiography Interpretation (Burnside et al., 2006; Velikova et al., 2013, 2009)
	Trauma Diagnosis (Ahmed et al., 2009)

Causality has been an important principle for representing expert knowledge in BN models as it is a natural way of expressing knowledge (Lucas, 1995) and it enables

development of less complex models (Koller and Friedman, 2009a). Even before the discovery of BNs, causality was employed by many AI approaches to model complex pathways between treatments and diseases (Patil et al., 1981; Weiss et al., 1978). Many large scale applications of BNs, including MUNIN (Andreassen et al., 1989) and Hepar II (Onisko et al., 1998), have causal structures based on domain knowledge.

Many BN applications use domain knowledge in development of the BN (see Table 3.1). Knowledge behind the BN, however, is often unclear to anyone else except the developers of the BN. The derivation of the BN structure, and supporting medical evidence, is often explained in a few paragraphs due to the space limitations in publications (for examples of inadequately described knowledge-based BNs see Ahmed et al., 2009; Burnside et al., 2006; Onisko et al., 1998; Wasyluk et al., 2001). The BN structure is often presented but the structure may not be descriptive enough as variable names often contain a limited amount of characters. As a result, the knowledge behind the BN may not be disseminated even when it is based on strong evidence.

More recently, BNs have been used to synthesise published evidence for clinical guidelines (Ni et al., 2011). Hanna et al. (2010) used BNs to develop clinical guidelines for inserting nasogastric feeding tubes. They modelled the reliability of different tests of verifying the position of a nasogastric feeding tube using a BN model. The variables and states of the BN were selected based on expert opinion. Published statistics from relevant studies were used to define the parameters. The BN model was used to compare the performance of different measurements, and to prepare a clinical guideline showing the optimal measurements. Hanna et al's study illustrates the power of BNs for combining evidence from different sources.

BNs can also be used as a purely data-based ML approach by using the learning techniques described in Section 2.6.2. Such BNs, as well as other ML approaches, are well suited for pattern recognition and knowledge discovery from large datasets. For example, data-driven BNs have had successful results in revealing biological relations in genome datasets (Friedman, 2004; Needham et al., 2007). These BNs, however, share the same disadvantages with other purely data based methods when they are applied to complicated decision making problems.

Fenton (2012) shows a simple example where purely data based approaches fail even when data is available in large amounts. In this example, the sample size of the dataset is large but there is no data about some rare combination of events. Therefore, a purely data-driven approach cannot gather information about these events even though the dataset is large in overall. Yet, experts are able to provide knowledge about the rare events based on other sources of information. Fenton's example (2012) has been encountered in real medical problems: in Chapter 6 we describe a similar challenge about modelling of rare events in the trauma case-study. We address this challenge by combining the information from previous publications with domain knowledge and data.

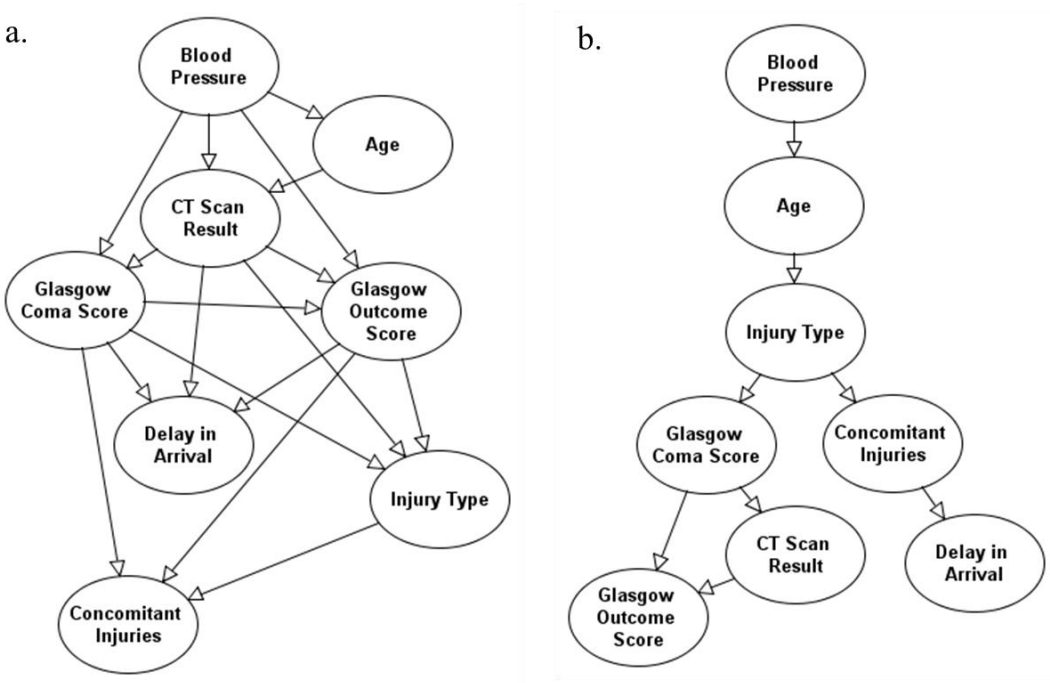


Figure 3.1 BNs for Head Injury Prognosis by Sakellaropoulos and Nikiforidis (1999)

Despite these problems, many clinical BNs continue to ignore clinical knowledge regardless of its abundance. For example, Sakellaropoulos and Nikiforidis (1999) built two, purely data-based, BNs to predict prognosis of head injuries by using two different learning methods on the same dataset (see Figure 3.1). Some the arcs between the same variables are in the opposite direction in these BNs, and some contradict with clinical understanding of the subject and common sense. For example, blood pressure is the parent of age in both models but the direction of this relation should be the opposite from a causal perspective as it makes sense to think

that old age causes increased blood pressure; not the other way around. Similarly, the result of the CT scan is the parent of delay in hospital admission in both models but this is confusing as CT scan is a measurement that is done after a patient is admitted. It may be more reasonable to think that delays in arrival to hospital made the patient's condition worse therefore led to a worse result from the CT scan. Another possibility is to include a clinically relevant latent variable to make the BN more consistent with clinical knowledge. For example, a latent variable about the severity of injury can be included as the parent of 'CT scan' and child of 'delays in arrival' based on an expert statement such as: "the severity of injury is measured by CT scan, and delays in arrival may worsen the state of overall injury".

In summary, it is difficult to explain how these BNs reason to a clinician apart from saying that they predict the previous cases in their data. Moreover, the data may not produce consistent results alone, as in this example, the BN structure may change significantly depending on the learning method applied. The results of the BNs may be statistically accurate but, like many statistical models, they fail to satisfy the other three properties that useful prognostic models should have: clinical credibility, generalisability and ability to provide useful decision support (Wyatt and Altman, 1995). The following chapter illustrates the challenges of building useful decision support models in the trauma case study.

Chapter 4

Case Study: Trauma Care

This chapter introduces the clinical case study which covers the treatment of mangled extremities in trauma care. We illustrate the challenges of building useful decision support models for the case study from two aspects. Firstly, we review the existing models for trauma care and discuss their limitations. Secondly, we describe the available datasets for the case study and discuss the need for expert knowledge to analyse the datasets. The chapter finishes with an overview of the novel methodologies proposed in the following chapters of the thesis.

4.1 Overview of the Case Study

One of the most difficult decisions for a clinician to make is whether to amputate or salvage a mangled extremity. This decision, with irreversible consequences for the patient, revolves around three possible adverse outcomes, which change prominence as the treatment progresses.

1. **Death.** Many trauma patients arrive at hospital with severely deranged physiology. Their risk of death is high and most prominent during the early stage of treatment. To reduce the risk of death, clinicians should resuscitate these patients, and allow their physiology to recover, before embarking on definitive limb reconstruction operations. Therefore, in the early stages of treatment, it is crucial to evaluate the physiological status and predict the risk of death before deciding to undertake a treatment.

2. **Limb tissue viability.** If the limb loses its blood supply for too long, its tissues cease to become viable and surgical removal of these tissues is

inevitable. The viability of the limb tissues is evaluated as the extent of the injury is assessed. The limb may become unsalvageable if a large amount of its tissues become unviable and are removed.

3. **Non-functional limb.** A salvaged limb may be more or less functional due to anatomical problems such as loss of muscle compartments or transected nerves. For some patients a prosthetic limb may be preferable to a non-functional or painful limb.

The clinician's concerns about these three treatment outcomes changes as the treatment progresses. The probabilities of the adverse outcomes are both positively and negatively related with each other so it may not be possible to find a decision that minimises all of them. For example, lengthy reconstruction surgery can salvage the patient's limb, but it can also put the patient's life in danger when the patient is physiologically unwell. In later stages of the treatment, following correction of the initial physiology, infections of the damaged limb tissues may again threaten the patient's life. Finally, the clinicians may decide to amputate the limb if it is not likely to be functional in the long run. Although the choice of treatment is the same, the underlying reasoning changes significantly through different stages of the treatment.

The activity diagram in Figure 4.1 illustrates the decision making stages and changing priorities in treatment of mangled extremities. In the remainder of this section, we describe the activity diagram by an example of a patient who survived a motorcycle accident. The patient is physiologically unstable since he had lost a large volume of blood before arriving to the hospital. One of his lower extremities has a bleeding traumatic injury. At the emergency room, surgeons assess his physiological state, risk of death, and injury. Since the patient's physiology is severely deranged, the surgeons decide to resuscitate the patient until his physiological state improves. His limb appears to be salvageable but the surgeons decide to delay any definitive reconstruction operations as such operations may put the patient's life in danger due to his physiological state. The surgeons make quick preventive operations to stop the bleeding in the lower extremity until the patient's physiology becomes stable enough for a definitive reconstruction operation. These preventive and lifesaving operations are also known as damage control surgery (DCS).

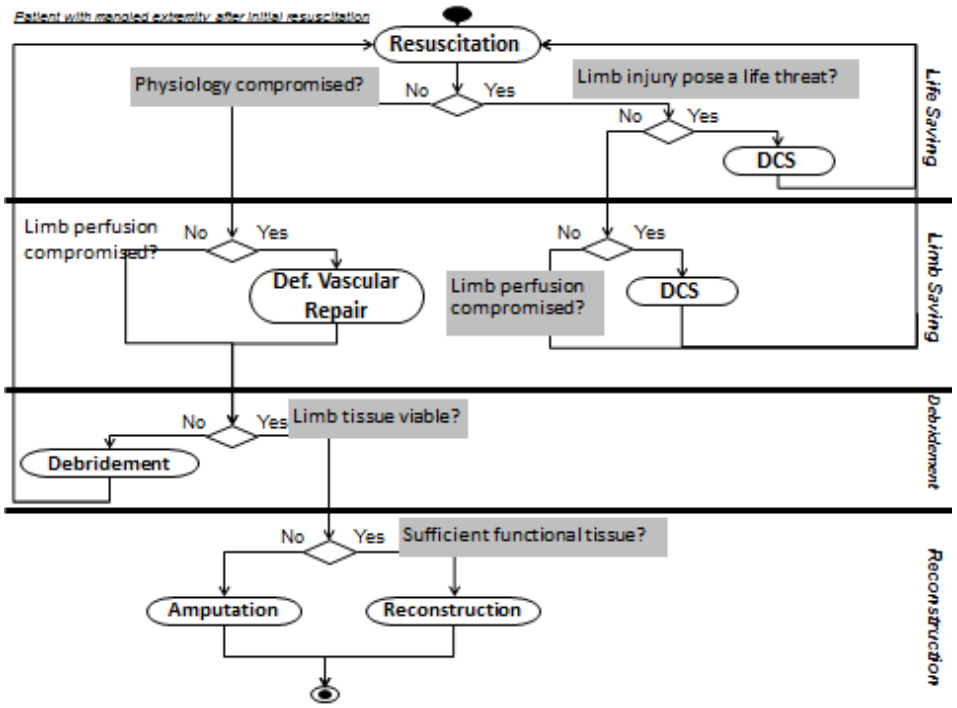


Figure 4.1 Activity Diagram for Treatment of Mangled Extremities

After the patient's physiology stabilises, the surgeons attempt to repair the vascular injuries on the lower extremity by a definitive reconstruction operation. The blood circulation in his lower extremity had been compromised as a result of the motorcycle accident and this lack of circulation caused a part of the soft tissue to become non-viable. The contamination and direct damage from the injury also caused a part of the soft tissue to become non-viable. At this stage, the surgeons assess the likelihood of successful limb salvage as they remove the non-viable tissue from the limb. Amputation may become inevitable if a large part of the soft tissue becomes non-viable since, without an adequate cover of soft tissue, his wounds may become infected, his limb may not function well and the vascular repair may fail.

The surgeons also assess the projected functional outcome throughout the care. They evaluate whether an amputation followed by a prosthetic limb may lead to better outcomes than the reconstruction of his limb.

4.1.1 Medical Collaborations

This section describes the medical collaborations that provided clinical knowledge and datasets used in the trauma case study.

4.1.1.1 Domain Experts

A trauma registrar at RLH, Mr Zane Perkins (ZP), provided clinical knowledge for analysing the case study and developing the decision support models presented in this thesis. ZP was closely involved in development of the BN models. His contribution included providing clinical knowledge, making systematic reviews of clinical literature and clinical verification of the developed models.

A consultant trauma surgeon at RLH, Mr Nigel Tai (NT), was involved in clinical verification of the developed models. NT was ZP's primary research supervisor.

4.1.1.2 Trauma Unit at the Royal London Hospital

The case study was done in collaboration with the trauma unit at RLH. RLH is an internationally recognised leader in trauma care and trauma research. The trauma unit is the busiest in the United Kingdom treating over 2000 injured patients in a year, a quarter of whom were severely injured. The hospital is also the lead for a network of trauma hospitals, the London Trauma System, which provides specialist trauma care for the millions of people living in London and the South-East of England. This trauma system is believed to be the largest of its kind in the world. As a major trauma centre, the hospital provides access to the latest technology, treatments and expert trauma clinicians around the clock. Evidence has shown that people who suffer serious injuries need the highest quality specialist care to give them the best chances of survival and recovery. The most common causes of injury seen at RLH are road traffic collisions, followed by stabbings and falls from a height. Nearly half of the trauma patients have an injury to an extremity or the pelvic girdle, and 1% of these patients end up having lower limb amputations. A large multidiscipline team manages those with severe limb injuries. These devastating injuries carry a high mortality and morbidity in a predominantly young population. The multidiscipline approach ensures the best possible outcome for these patients. The clinical datasets used for developing the ATC BN were provided by the RLH trauma unit and by their national and international collaborations with other hospitals (see Section 4.3.1 for a description of the ATC datasets).

4.1.1.3 United States Army Institute of Surgical Research

United States Army Institute of Surgical Research (ISR) in San Antonio, Texas provided the dataset used for developing the lower extremity vascular trauma (LEVT) BN (see Section 4.3.2 for a description of the LEVT dataset). The author and ZP visited ISR to extract and refine the LEVT dataset, and to develop the LEVT BN model.

ISR is one of the 6 research institutes within the United States Army Medical Research and Materiel Command. It is the leading research institute of combat casualty care for the United States Army. ISR focuses on a wide variety of research areas including extremity trauma, burn treatment, emergency medical monitoring and casualty care engineering. A diverse workforce of over 250 military and civilian personnel works at ISR to accomplish these research objectives.

4.1.2 Decision Support Requirements

Since the treatment of mangled extremities involves multiple decisions that affect multiple outcomes in multiple stages, it is crucial to define the scope of the decision support models according to the requirements of the clinicians. We conducted a series of interviews with the domain expert (ZP) in order to have a shared understanding of the areas where probabilistic models can provide useful decision support. The description of the case study and the activity diagram (see Figure 4.1) in Section 4.1 were also produced as a result of these interviews. The domain expert suggested that predictions of the following outcomes can potentially assist the decision making in treatment of traumatic lower extremity injuries:

1. **Death and ATC**: One of the most critical physiological problems at the early stage of treatments is acute traumatic coagulopathy (ATC). ATC is the failure of the body's protective mechanisms to limit bleeding. The patients with ATC have a considerably higher risk of bleeding and death. A model that accurately predicts ATC and the related risk of death can be used as the basis of a risk-benefit analysis for limb reconstruction operations during the early stages of treatment.

2. **Limb Tissue Viability:** The extremity may become unsalvageable if large amount of tissue becomes unviable and removed. Success of the vascular reconstruction is essential for the tissues to have adequate blood supply and remain viable. Predicting the outcome of a reconstruction operation and projected soft tissue viability would be useful in early decision making. Such prediction would allow informed treatment decisions and be helpful in assessing the risk of failure of a salvage attempt.

3. **Non-functional Limb:** Since amputations followed by prostheses can sometimes lead to better functional outcomes than salvaged extremities, predicting the long-term function outcomes of a salvaged extremity would assist the decision making.

We developed two BN models that aim to provide decision support for the first and second outcomes above. The first BN model predicts ATC and mortality using the observations that are available in the first 15 minutes of the treatment. The development methodology and results of this BN are described in Chapter 5. The second BN model predicts the short term outcomes of a vascular reconstruction operation by estimating the soft tissue viability. The development methodology and results of this BN are presented in Chapter 6. The third – non-functional limb – outcome was considered to be out of the scope of this thesis due to the issues discussed in Section 4.3.3.

In the following section, we review the existing models that were built to provide decision support for the death, limb tissue viability and non-functional limb outcomes.

4.2 Review of Existing Models in Trauma Care

Most decision support models in trauma care are designed as scoring systems: they calculate a score for the situation of a patient using several inputs. Some scoring systems aim to summarise clinical conditions, leaving decisions to a clinician. For example, the Glasgow coma scale (GCS) summarises the level of consciousness after head injury (Teasdale and Jennett, 1974). The abbreviated injury scale (AIS) summarises the severity of anatomical injury in different parts of the body (Civil and

Schwab, 1988; Gennarelli and Wodzin, 2008), and the injury severity score (ISS) (Baker et al., 1974) summarises the severity of all injuries combined by using AIS scores. Other scoring systems aim to recommend a treatment by setting a threshold to the calculated score. For example, the mangled extremity scoring system (MESS) recommends amputation if the score is over a certain threshold value (Johansen et al., 1990). However, this adds little to an experienced clinician's judgement especially when the score is close to the threshold. If there is a discrepancy between the model's recommendations and clinician's decisions, the model does not provide any useful decision support apart from implying that the recommended decision was the decision made in a similar circumstance in the model's training data.

Both kinds of scoring systems have been developed for trauma care. In the remainder of this section, we review the scoring systems related to the 3 main treatment outcomes in our case study: death, limb tissue viability and non-functional limb.

4.2.1 Death

The revised trauma score (RTS) is one of the earliest scoring systems about patient physiology (Champion et al., 1989, 1981). RTS was originally developed as a triage tool that assigns patients to trauma care if they score less than a predefined threshold value. However, RTS has been mainly used to predict mortality as it is found to be correlated with the rate of survival. RTS is calculated from three inputs: blood pressure, respiratory rate and GCS. Several studies indicate that RTS is overly simple and lacks important factors, such as those about anatomy, for predicting mortality (Gabbe et al., 2003; Russell et al., 2010).

The trauma and injury severity score (TRISS) calculates the probability of death by combining the scores from RTS and ISS, and also adjusting for patient's age and mechanism of injury (Boyd et al., 1987). TRISS cannot be used for decision support in early stages of care since acquiring the necessary injury descriptions for ISS may take several weeks. It has been mainly used for auditing and performance assessment (Russell et al., 2011).

In a recent study, Perel et al. (2012) proposed a regression model, called the CRASH-2 prognostic model, to predict mortality specifically for bleeding trauma

patients. The independent variables in the model include injury characteristics, physiological variables and demographic factors such as the level of income. The variables were selected according to their p-values in a large international dataset. Limitations of using p-values for building clinical models are discussed in Section 3.2.

ATC is one of the most critical physiological problems in trauma care. The coagulopathy of severe trauma score (COAST) (Mitra et al., 2011) has been developed for predicting this condition but its performance were found inadequate for clinical use (Brohi, 2011). The COAST score had several erroneous assumptions about modelling of the latent variables; the limitations of the COAST score are examined further in Chapter 5.

4.2.2 Limb Tissue Viability

Many scoring systems have been developed to provide decision support for the treatment of traumatic lower extremities at the initial evaluation of a patient (see Table 4.1). These models calculate a score for the patient based on several physical and physiological factors, and recommend amputation if the score is above a certain threshold value. Table 4.2 shows the variables used for calculating scores in each of the scoring systems.

Table 4.1 Scoring Systems for Traumatic Limb Injuries

Scoring Systems for Traumatic Limb Injuries
Mangled Extremity Syndrome Index (MESI) (Gregory et al., 1985)
Predictive Salvage Index (PSI) (Howe et al., 1987)
Hannover Fracture Scale (HFS) (Südkamp et al., 1989)
Mangled Extremity Severity Score (MESS) (Johansen et al., 1990)
Limb Salvage Index (LSI) (Russell et al., 1991)
Nerve, Ischemia, Soft tissue, Skeletal, Shock, Age Score (NISSA) (McNamara et al., 1994)

The scoring systems were developed based on the historical decisions in their training datasets. For example, the threshold for MESS, which is one of the most popular limb scoring systems, was defined with the score that discriminated all of the amputations and salvages in the training dataset of MESS (Johansen et al., 1990). MESS is likely to be overfitted to its training dataset as the dataset was small,

containing 39 patients. The initial performance of MESS was never repeated in external validations (see Table 4.3).

Bosse et al. (2001) conducted a prospective multi-centre study that validated 5 of the scoring systems on 556 patients. This study concluded that the scoring systems are not good predictors of short term outcome and they should not be used for decision making in clinical practice. The sensitivity and specificity values in Bosse et al.'s study (2001) were substantially lower than the values reported by the models' developers (see Table 4.4).

Table 4.2 Variables Used in Scoring Systems for Traumatic Limb Injuries

	MESI	PSI	HFS	MESS	LSI	NISSSA
Patient Age	▪			▪		▪
Bone Fracture	▪	▪	▪	▪	▪	▪
Comorbidities	▪		▪			
Nerve Injury	▪	▪	▪		▪	▪
Physiology/Shock	▪			▪		▪
Skin/soft tissue Injury	▪	▪	▪	▪	▪	▪
Time until Treatment	▪	▪	▪			
Vascular Ischemia	▪	▪	▪	▪	▪	▪

The scoring systems recommend a decision based on the historical decisions in their training dataset. In other words, they try to imitate the decisions that are made in similar circumstances in the training dataset without relating them to objective patient outcomes. This kind of approach is fundamentally flawed since some of the decisions, which were correct at the time, may become incorrect due to the changes in medical knowledge and practice.

Table 4.3 Validations of MESS

Study	Participants	Sensitivity	Specificity
Johansen et al. (1990)	26	100%	100%
Robertson (1991)	154	43%	100%
Bonanni et al. (1993)	89	22%	53%
Durham et al. (1996)	51	79%	83%
Bosse et al. (2001)	556	46%	91%
Korompilias et al. (2009)	63	87%	71%

Studies about the sensation in feet and amputation decisions illustrate the change of clinical practice and decision making in time. A limb cannot function without a

functional nerve therefore a permanent nerve dysfunction can make amputation inevitable. A nerve dysfunction in a lower limb can be diagnosed with the lack of sensation in the foot. A survey of orthopaedic surgeons showed that they considered insensate feet as an important factor for making amputation decisions (Swiontkowski et al., 2002). Several scoring systems also used insensate feet as a predictor of amputations (Johansen et al., 1990; McNamara et al., 1994). However, insensate foot is only an indicator of a more important decision factor: permanent nerve dysfunction. A few years after the Swiontkowski et al.'s survey, Bosse et al. (2005) showed that although insensate foot can indicate permanent dysfunction of nerves, it can also be related to long-term but temporary problems. They observed that some patients regain their sensation after several years. Bosse et al. (2005) concluded that surgeons should avoid making amputation decisions based on sensation in feet. Using sensation in feet, as a predictor, possibly increased the scoring systems' accuracy for predicting historical amputations but it would have affected clinical outcomes negatively if these models were used for decision making.

Swiontkowski et al.'s survey shows that some of the historical amputation decisions were possibly based on the sensation in the foot. If so, a data-driven approach would find correlation between the sensation and amputation, and would continue to build erroneous models. Such errors can be avoided if causal relations between sensation in the foot and permanent nerve function are modelled with domain experts. In this case, the experts would be able to indicate evidence from Bosse et al's study and include other – temporary – causes of the loss of sensation in the model.

Table 4.4 Internal and External Validation Results of Scoring Systems

Scoring System	**Sensitivity**		**Specificity**	
	Internal Validation*	External Validation**	Internal Validation*	External Validation**
PSI	0.78	0.46	1	0.87
MESS	1	0.46	1	0.91
LSI	1	0.46	1	0.97
NISSA	0.81	0.33	0.92	0.98
HFS	0.82	0.37	0.99	0.98

***Results reported by authors, **Results from Bosse et al. (2001)**

4.2.3 Non Functional Limb

Following Bosse et al.'s study (2001), Ly et al. (2008) examined the performance of the scoring systems for predicting limb function. They showed that none of the scores reviewed by Bosse et al. (2001) are usefully correlated with function outcomes.

Several survey based scores are available for summarising function outcomes. The short musculoskeletal function assessment (SMFA) score calculates multiple scores about function and emotional status based on 46 questions (Swiontkowski et al., 1999). Similarly, the short form – 36 – health survey (SF-36) calculates several outcome scores including a function score based on 36 questions (Ware and Sherbourne, 1992).

4.3 Available Datasets

This section describes the datasets used for the trauma case-study. The datasets were used to develop and validate the BN models presented in Chapter 5 and 6.

4.3.1 ATC Datasets

We used three datasets to develop and validate a decision support model that predicts physiological derangements and mortality at early stages of trauma care.

The first of these datasets, called the training dataset, were used to build the ATC BN described in Chapter 5. The training dataset contains detailed information about 600 trauma patients who were treated at RLH. Table 4.5 summarises the available information in the dataset in different categories.

The second dataset, called the test dataset, contains 300 patients who were treated at RLH at a later date than the first 600 patients in the training dataset. The test dataset was used for the temporal validation of the ATC BN. A temporal validation is the validation of a predictive model using data collected from the same population after the model was developed.

Table 4.5 Information in the ATC Datasets

Data Section	Available Information
Patient Characteristics	Age, Gender and Comorbidities of the patient
Injury Characteristics	Mechanism and Energy of Injury, Injury Descriptions, ISS and AIS scores, result of the FAST* scan, presence of haemothorax, pelvic fractures and long bone injuries.
Initial Observations	Heart rate, systolic blood pressure, Glasgow coma score and body temperature of the patient measured shortly after admission to the hospital
Initial Point of Care Results	Blood pH, lactate and base excess values from the arterial blood gas test, ROTEM* test results including EXTEM A5* and A30* values measured shortly after admission to the hospital.
Initial Laboratory Results	INR*, PTR* and APTTR* values measured shortly after admission.
Fluid Transfusions	The amount of blood product and other fluid transfusions before and after admission.
Later Point of Care and Laboratory Results	Blood pH, lactate and base excess values, ROTEM EXTEM A5* and A30* values, INR*, PTR* and APTTR* values measured after 4^{th}, 8^{th} and 12^{th} unit of blood is transfused to the patient.
Outcome	Survival outcome

FAST: Focused Assessment with Sonography for Trauma, ROTEM EXTEM A5 and A30: Amplitude of Rotational Thromboelastometry **Extem tests at 5^{th} and 30^{th} minutes, INR: International Normalised Ratio, PTR: Prothrombin Ratio, APTTR: Activated Partial Thromboplastin Time Ratio.*

The third dataset, called the external dataset, contains 122 patients: 92 patients who were treated at a hospital in Oxford, UK, and 30 patients who were treated at a hospital in Cologne, Germany. This dataset was used for the external validation of the ATC BN. The variables in this dataset are exactly the same as the other two datasets from RLH since all datasets were collected as a part of a research collaboration led by RLH. Penetrating injuries, such as stabbings, were less common, and blunt injuries, such as traffic accidents, were more common in the external dataset compared to the datasets from RLH. A summary of the injuries in the training and validation datasets are shown in Table 4.6.

All of the datasets contained missing variables mainly due to recording errors or missing laboratory tests. Apart from the missing values, two of the most important clinical variables about the physiological derangements, i.e. ATC and Hypoperfusion, were not available in any of the datasets. Lack of data for these

variables makes it challenging to build a model to predict physiological derangements and related mortality. In Chapter 5, we discuss these challenges and propose a methodology that systematically uses expert knowledge to overcome them. Using this methodology, a group of experts, including ZP, reviewed the categorical and free-text information in the data and extracted information about the Hypoperfusion and ATC variables.

Table 4.6 Training and Validation Datasets for the ATC BN

ATC BN Datasets	Training	Temporal Validation	External Validation
Patients	600	300	122
Deaths	71	27	23
Massive Blood Transfusion	37	21	7
High Energy Injuries	202	78	42
Low Energy Injuries	398	182	76
Unknown Energy	-	40	4
Blunt Injuries	475	239	121
Penetrating Injury	125	61	1

4.3.2 LEVT Dataset

We used a dataset of 521 lower extremity injuries of 487 patients to build a model that predicts the short term viability outcomes of lower extremities with vascular trauma (LEVT). All patients in the dataset had lower limb vascular injuries with an attempted salvage. Some of the patients had injuries on both limbs or vascular injuries at multiple levels on the same limb. The dataset contained a large amount of information recorded as free text descriptions. ZP reviewed these free text descriptions and extracted categorical information that was necessary for training the BN model. A summary of the dataset is shown in Table 4.7.

Amputations that are performed after an attempt to reconstruct a lower extremity are called secondary amputations. Secondary amputations may indicate unsuccessful salvage outcomes in the dataset. The LEVT dataset contained 90 lower extremities that had secondary amputations. It is, however, crucial to identify the clinical reasons of the secondary amputations since some secondary amputations may have reasons other than short-term viability. For example, several patients in the dataset, who had successful limb repairs in terms of viability, decided to undergo amputation several years after the injury. The main reason for their amputation was the pain and limited

function of their lower extremity. Although these patients had secondary amputations, they are considered as positive outcomes in terms of short-term viability and negative outcomes in terms of long-term function. Since long-term function is out of the scope of our model, these patients were labelled as positive outcomes in our training data. ZP reviewed all of the secondary amputations in the dataset and identified the clinical reasons of these decisions based on the free-text descriptions of injuries and operations. Amputations due to the causes other than short-term viability were labelled as positive outcomes in the training data. A summary of the learning dataset is shown in Table 4.8.

Table 4.7 Description of the LEVT Dataset

Data Section	Available Information
Patient Background	Id number, age, gender of the patient, and the laterality of the traumatic lower extremity
Vascular Injury	Location and type of the injured vessel,
Vascular Repair	Description and results of the vascular reconstruction operations
Ischaemia	Degree and duration of ischemia.
Soft Tissue Injury	Location and severity of soft tissue damage
Associated Injuries	Presence of associated bone, nerve and vein injuries.
Physiology	Degree of Shock
Amputations	The reason for amputation, the level of amputation (below the knee, through the knee, above the knee).

Table 4.8 Summary of the LEVT BN Training Dataset

LLVI BN Cross Validation Dataset	
Patients	487
Lower Limb Vascular Injuries	521
Secondary Amputations	90
Secondary Amputations – Short Term Viability	54
Above Knee Vascular Injuries	231
Below Knee Vascular Injuries	290
Patients with Bilateral Vascular Injuries	18
Vascular Injuries at Multiple Levels at the Same Lower Limb	16

Although the LEVT dataset had data for 521 injuries, there were not enough data for some severe but uncommon types of injuries. For example, the level of vascular injury, the type of repair operation and the presence of vascular injuries at multiple levels are important factors that are known to affect the outcome of reconstruction operations. Some combinations of these factors are uncommon but have significant consequences for the patient. The dataset contains a very small amount of

information for these combinations but their prediction is clinically important. In Chapter 6, we propose a methodology that combines results from previous research and expert knowledge with data to overcome these challenges.

4.3.3 Limb Function Dataset

The LEVT dataset also contained data about the function outcomes of 478 patients. This data was composed of the SMFA survey scores of 208 patients and SF-36 survey scores of 278 patients. We built a prototype BN that aims to predict the functional components of SMFA and SF-36 scores using the injury and treatment variables available in the LEVT dataset. However, the accuracy of the prototype was not satisfactory. The domain experts indicated that factors related to life style, such as social and economic factors, are known to be affecting the function outcome reported by patients (Harris et al., 2008; MacKenzie et al., 2005; Wegener et al., 2011). The lack of accuracy was possibly due to not including these factors in the prototype. Since we had neither data nor experts about the life style factors, we decided to leave decision support for the non-functional limb outcome out of the scope of this thesis. It would not be possible to identify this issue if we used a purely data-driven approach. Expert knowledge showed us the limitations of our resources and the requirements for developing a useful model for predicting limb function.

4.4 Challenges of Developing Useful Decision Support Models

Decision support models must be consistent with clinical knowledge to combine relevant evidence and provide evidence-based decision support. This could be achieved by models that reason with causal relations supported by clinical evidence. However, information about some of the clinically important causal relations is often not available in clinical datasets. The dataset described in Section 4.3.1 does not contain some of the most clinically important variables about physiological derangements. In this case, a decision support model must contain ‘latent’ variables in order to be consistent with clinical knowledge. The latent variables are not available in the dataset therefore their behaviour must be learned using expert

knowledge and other sources of information. In other cases, data may exist but in small amounts for some clinically important variables. The dataset in Section 4.3.2 also lacks some clinically important variables and it has only a limited amount of information for some severe and uncommon injury types. Therefore, the data needs to be supplemented by other sources of information to model the behaviour of these uncommon injuries.

The previous trauma models reviewed in Section 4.2 were primarily based on the correlations learned from their training datasets. For example, the previous models for predicting limb tissue viability imitated the previous decisions in their datasets, and ignored factors that lack data. As a result, their predictions were overfitted to their training dataset, and failed to be useful in external datasets.

BNs are well suited for modelling causal relations by combining evidence from experts, data and published research. Causal relations are often modelled based on expert knowledge as data-driven approaches have limited ability to identify causal relations (see Section 2.6.2). However, the elicitation of causal relations from experts is a challenging task requiring iterative steps of knowledge engineering to define the important variables and identify their relations. Moreover, evidence supporting the causal relations is often not presented in detail and is ambiguous to anyone else except the model's developers. Consequently, knowledge and evidence supporting the BN becomes unclear even when the BN is based on reliable evidence.

These challenges demonstrate the need of novel methodologies to develop and maintain evidence-based BN models for decision support. In the remainder of this thesis, we propose methodologies to address these issues. These novel methodologies are illustrated by two BNs developed for the trauma case study.

In Chapter 5, we propose a methodology to build BNs that reason consistent with clinical knowledge without being limited by the observed variables in data. We use this methodology to develop a BN that predicts a potentially fatal physiological derangement, ATC, and death outcome using the training dataset described in Section 4.3.1. Some of the most important variables about this outcome are latent variables that cannot be directly observed and thus not present in the dataset. A purely data-driven model either ignores the existence of these variables or tries to

estimate them from the other variables available in the dataset. Both of these approaches ignore clinical knowledge about the latent variables. Our methodology systematically combines clinical knowledge and data to learn the behaviour of latent variables, and to refine the BN model. The performance of the BN is evaluated in multiple datasets using a 10-fold cross-validation, a temporal and an external validation.

In Chapter 6, we propose a methodology to build decision support models by combining evidence provided by meta-analysis of systematic reviews with expert knowledge and data. We use this methodology to develop a BN for predicting limb tissue viability outcome using the dataset described in Section 4.3.2. Although this dataset contains a total of 521 injuries, there is small amount of data for some uncommon injury types with potentially severe consequences. Therefore, a purely data-driven approach cannot uncover the relation between these uncommon injuries and outcomes due to the limitations of the data. We conducted a systematic literature review and meta-analysis of these factors, and combine the results of the meta-analysis with expert knowledge and data to derive the structure and parameters of the BN decision support model. The accuracy of the BN is compared with purely data-based learning methods and with existing models in a 10-fold cross validation.

The BN models in Chapter 5 and 6 were developed with two trauma surgeons (Mr Zane Perkins and Mr Nigel Tai). Deriving BN structure from expert knowledge is a challenging task requiring iterative stages of knowledge engineering. The initial BN structures elicited from experts are often complicated with too many arcs and variables. The knowledge engineer and domain experts iteratively simplify the structure by removing some of the less relevant variables and relations. However, these modifications can lead to undesirable effects to knowledge encoded in the initial structure. When the derivation stages are not presented, knowledge behind the BN structure cannot be completely understood by anyone else except the developers of the BN. Knowledge engineering methodologies to systematically observe and present the effects of BN simplifications have not been thoroughly studied. Chapter 7 proposes a method of abstracting a knowledge-based BN structure elicited from domain experts. The abstraction method shows the link between the initial and final BN structures by showing the effects of each abstraction step to underlying

probability distribution. A part of the ATC BN is used as a case study to illustrate the application of the abstraction method.

In order to use a BN for EBM, evidence behind the BN should be explicitly presented to clinicians. The graphical structure of a BN shows the variables and relations included in the model but it does not present evidence supporting or conflicting with these elements. As a result, evidence behind the model becomes unclear to its users even when the model is based on substantial evidence. Chapter 8 proposes a framework for representing the evidence behind clinical BNs. The evidence framework is composed of two parts: 1) an ontology that organises evidence regarding variables, relations and fragments in a BN 2) a web page generator that presents evidence in a web page without showing the technical details of the ontology. The ATC BN is used as a case study to illustrate the evidence framework.

Chapter 5

Modelling Latent Variables with Knowledge and Data

Many medical conditions are only indirectly observed through symptoms and tests. Developing predictive models for such conditions is challenging since they can be thought of as 'latent' variables. They are not present in the data and often get confused with measurements. As a result, building a model that fits data well is not the same as making a prediction that is useful for decision makers. Chapter 4 illustrates these challenges based on the existing trauma models and available datasets. In this chapter, we present a methodology for developing BN models that predict and reason with latent variables, using a combination of expert knowledge and available data. The method is illustrated by the BN that aims to assist early stages of mangled extremity decision making by predicting acute traumatic coagulopathy (ATC), a disorder of blood clotting that significantly increases the risk of death following traumatic injuries. There are several measurements for ATC and previous models have predicted one of these measurements instead of the state of ATC itself. Our case study illustrates the advantages of models that distinguish between an underlying latent condition and its measurements, and of a continuing dialogue between the modeller and the domain experts as the model is developed using knowledge as well as data.

5.1 Introduction

Purely data-driven approaches are currently accepted as the primary, if not the only, way of developing predictive models. Because of the impressive results achieved

with such approaches by organizations like Amazon and Google, it is often assumed that this success is repeatable in other domains as long as a large enough amount of data is available. However, a purely data-driven approach can only predict the type of values recorded in a dataset, such as measurements made, decisions taken or outcomes recorded. Even when large volumes of data exist, purely data driven ML methods may not provide either accurate predictions or the insights required for improved decision-making. In this chapter, we consider the common real-world situation in which successful decision making depends on inferring underlying or latent information that is not – and can never be – part of the data. In such a situation a predictive model for decision support will contain latent variables representing this underlying state but the values of these variables will not be present in the data. We therefore need to depend on domain expertise to identify the important latent variables and to model relations between them and the observed variables.

Domain experts do not just substitute guesswork for data. They may have access to information that is not machine-readable and they should back up any judgements by reference to published research whenever possible. Yet, such expert knowledge is usually avoided in data-driven approaches using arguments such as 'avoiding subjectivity' and 'using facts based on the data' (Gelman, 2008; Tonelli, 1999). The use of latent variables is also limited: some data-driven approaches, such as regression modelling, do not include latent variables at all. Other approaches contain latent variables but these are estimated only from data values, so that the use of latent variables in these methods does not escape the limits of the data. The objectivity of data-driven approaches holds only so far as the prediction of observed values really serves the needs of users. When this is not the case, erroneous results may follow. In this chapter, we show some examples of these errors and how they are avoided by appropriate and rigorous use of domain knowledge.

We propose a pragmatic methodology to develop BN models with latent variables. Our method integrates domain expertise with the available data to develop and refine the model systematically through a series of expert reviews. We illustrate the application and results of this method with a clinical case study of a problem for which purely data-driven approaches have been tried but have not been considered to

be successful by clinicians. Our case study shows some possible reasons for these past failures. The details of the case study are provided in Chapter 4.

The remainder of this paper is as follows: Section 5.2 presents the overview of our methodology. The case study is introduced in Section 5.3 and developed further in Sections 5.4 (learning and review), 5.5 (model refinement) and 5.6 (temporal and external validations). We present our conclusion in Section 5.7.

5.2 Method Overview

The limitations of data for making predictions useful to a decision-maker can be summarised in three points:

1. *Measurement errors:* a dataset contains measurements of variables, but measurement errors mean that the true state of each variable differs from the measured data. In some domains, including clinical diagnosis, this introduces significant uncertainty about the true value, so that a data-driven model cannot accurately predict the underlying state even if it can accurately predict the associated measurement values.

2. *Sub-optimal decisions:* the objective of a decision-support model is to enable a decision-maker to determine the optimal decisions given the observed situation. A dataset may contain a 'decision' variable, that is, one that reflects the decision made (e.g. a treatment given by a clinician). A model that predicts the value of a decision variable can be useful if all the past decision-makers had similar utilities and they were completely rational in evaluating utilities with their underlying uncertainties. However, there is usually no information about the utilities involved in past decisions, and the data may have records of some decisions that were incorrect at the time or, although correct at the time, were made on outdated understanding. A model that predicts the value of a decision variable is therefore limited in its performance even if the prediction is highly accurate. Moreover, a model can only be used for 'what if' analysis – exploring the consequences of decision alternatives – if it is causal; choosing one of the decision alternatives erases the factors that influenced past decisions (Pearl, 2000). Although these

problems are well known, models that are developed to fit past decisions are common in scientific literature (see Section 3.1).

3. *Causes of outcomes:* an 'outcome' variable records what happened. But outcomes can have many causes, only some of which may be recorded in the dataset (for example, in medical applications not all interventions and treatments are recorded). A prediction based on only some causes may be useful – the missing causes simply add uncertainty – but understanding of the scope of the causes included is important to the correct application of the model. A purely data-driven approach does not resolve this problem; only an expert can detect if the data omits known causes of the outcome. If omitted causes can be identified, this information can be used either to improve the model or to clarify its scope and to assess its performance within the scope of the causes modelled.

The main aim of our method (illustrated in the flow diagram in Figure 5.1) is to overcome these limitations. We show how to develop BNs that predict and reason with latent variables using a training dataset including measurements of these variables, but not including their true state. Domain expertise is used both at the start of the development to discover latent variables and then later to refine the model in a series of expert reviews; it is during these reviews that discrepancies between knowledge and data are revealed. Expert knowledge can be used in various degrees when deriving the structure of a BN (Korb and Nicholson, 2004a). In our method, the structure of the BN is developed with domain experts by using small BN fragments for commonly occurring reasoning types as building-blocks to form the complete BN structure (Neil et al., 2000). The advantage of experts deriving the model's structure, rather than learning it from data, is to ensure causal coherence: latent variables influence measurements and decision variables influence outcomes. Hybrid approaches that combine expert knowledge and data can also be used at this stage for deriving the BN structure (Cano et al., 2011; Flores et al., 2011). Moreover, structure learning methods can be used as a complementary approach to evaluate and refine a BN structure developed by experts (Velikova et al., 2013). Of course, all causal assumptions need to be supported by the best available evidence, such as experimental results or expert consensus. Lack of knowledge of true causal

relationship is a problem and affects both expert and data-led modelling (aside from the limited capabilities of algorithms such as inductive causation (IC) (Pearl and Verma, 1991)) alike. Equally, not all causal relationships are uncertain: it is clear that an object's temperature causes the thermometer reading rather than the other way around.

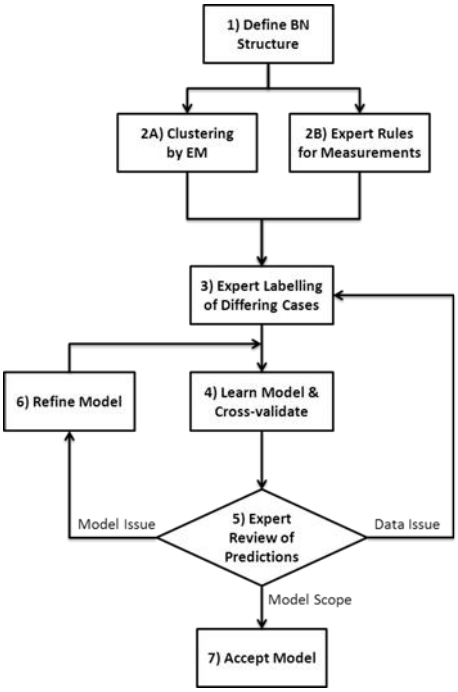


Figure 5.1 Method for Learning BN with Latent Variables

The next step is to label the latent variables in the training dataset, overcoming the problem that their values are unknown. The first label is derived from measurement data using deterministic (but not necessarily complete) rules defined by domain experts; the second uses data clustering. The experts' rules can be of any form, but are typically derived from current practice. For example, if the related measurements are continuous, these rules are threshold values for the measurements. For clustering, we use the standard Expectation-Maximisation (EM) for BNs with known structure (Lauritzen, 1995). EM is an iterative algorithm that is used for learning the parameters of a BN from a dataset with missing values. Each iteration of EM has two steps: the E-step completes the data by calculating the expected values of unobservable variables based on the current set of parameters; the M-step learns a new set of parameters from the maximum likelihood estimate of this completed data. When EM is used for parameter learning, the M-step of the final iteration calculates the BN parameters. When it is used for clustering, the unobserved variables are

labelled according to the values in the E-step of the final iteration. In our method, all of the values of the unobserved variable are missing from the dataset and we are using EM for clustering the unobserved values. Although EM can also be used for structure learning (Friedman, 1998, 1997) this is not required in our method as the BN structure is developed with domain experts. Extensions of EM that builds upon the information bottleneck (Elidan and Friedman, 2006), variational Bayesian (Attias, 1999) and hierarchical (Zhang, 2004) frameworks have been proposed for learning latent variables. Van Der Gaag et al. (2010) presents a similar approach to labelling with expert rules where they represent combinations of multiple observations with latent variables.

We now have two labels for each latent variable: one from clinical measurements and the experts' rules, the other from EM clustering. A final label is achieved by combining the two labels in cases where the labels are the same and by expert review of cases where there is a difference between the two labelling methods. We prepare a list of cases where the labels differ. Domain experts then decide the final label for each data record in this list. The experts can review other data including information that is not machine-readable and cite relevant research to support this decision. We also include a random subset of cases that were labelled consistently in the review to assess the experts' consistency with the labelling by measurements and clustering approaches. This combination of expert review and data has a number of advantages. It allows for the possibility of errors in measurement, and it uses the experts efficiently. Expert review is a costly resource and using it for every single case in the data is usually not feasible, especially if the dataset is large. Therefore, our method aims to use it only for ambiguous cases, where the labels from measurements conflict with the results of the clustering on our data.

After the expert review, we use the original dataset to which the latent variable labels have been added, to learn the BN's parameters and to evaluate its performance. We again use the EM algorithm but this time to learn the parameters, since the dataset may still contain missing values of other variables for some patients. The second use of the EM algorithm in this step should not be confused with the previous use of the same EM algorithm to label latent variables in the step 2A (see Figure 5.1). Expert constraints (Helsper et al., 2005) in the form of parameter orders can also be used if

the available data is too small for learning a part of the parameters. We use k-fold cross-validation to evaluate the performance of the BN. In this approach, the data is divided into k equal sized groups. One of the k groups is used as test data while the remaining k-1 groups are used for training the BN. The learning and testing continues iteratively until the model is validated with all of the k groups.

The inaccurate predictions of a predictive model offer useful lessons for improving the model and are the focus of the next stage of review. The BN modelling approach is well-suited for this kind of review since it concretely represents separate medical pathways leading to its predictions (Fenton and Neil, 2010; Lucas et al., 2004). When the BN model's prediction in the cross-validation differs from the value recorded in the data, the domain experts investigate the reasons for this difference to look for potential improvements to the model or clarify its scope. The domain experts look at cases where the recorded values are what is expected in their experience even though it is different to what was predicted by the model. In some cases, the domain experts may agree with the prediction of the model, and they may consider the value recorded in the data as an error or simply as an unexpected outcome. For example, in a medical decision-support scenario, survival of a patient with a severe injury burden and high blood loss might be considered to be an unexpected outcome. The expert review can also clarify scope of the model: if the recorded outcome is explained by factors that have been excluded from the model then this should be made clear to the model's users. For example, the experts might note that patient who unexpectedly survived did so as a result of a particular pre-hospital treatment, and the model could not identify this as pre-hospital interventions were out of the scope of the model. Alternatively, additional latent variables and relations that are important for the predictions can be discovered and added to the model. Since the BN's structure represents domain knowledge, any modifications must be supported by evidence.

Differences between the available data and the target subpopulation of the BN must be examined as the knowledge from these two sources is combined in our method. Correlations, caused by these differences, must be analysed and modelled in the BN structure in order to avoid developing erroneous models (Druzdzel and Díez, 2003).

5.3 Case Study: Trauma Care

We illustrate our methodology with the ATC BN which aims to provide decision support for the first outcome of the mangled extremity case study (see Section 4.1.2 for a description of three main outcomes in the case study). The details of the case study and the development and validation datasets of the ATC BN are described in Chapter 4. In this section, we recap limitations of the data driven models. Next, we discuss the significance of ATC in trauma care and challenges of predicting ATC. The structure of the ATC BN is also presented in this section.

5.3.1 Data-driven Models in Trauma Care

Several data-driven prediction models have been developed for decision support in trauma care (see Section 4.2) but with little impact in clinical practice due to the limitations discussed in Section 5.2:

1. *Measurement errors:* in the previous models to predict ATC, the presence of ATC is identified with a threshold value on a blood test called international normalised ratio (INR) (Mitra et al., 2011) despite the fact that this test has known limitations at identifying this condition. This approach has been criticised as it fails to produce useful clinical results (Brohi, 2011).

2. *Sub-optimal decisions:* models that predict the decision values in data have been developed in other areas of trauma care. One example is decision support for injured extremities which encompasses knowledge of the presence of ATC. Several data-driven models have been developed to predict amputation decisions in this domain (see Section 4.2). Although some of these models have been used as research or evaluation tools, none of them have been recommended as a decision support tool in clinical practice (Bosse et al., 2001). The output of these models shows the percentage of clinicians that made amputation decision in similar circumstances. However, recommending an amputation without relating it with patient outcomes makes it difficult to assess the model or to understand its reasoning. Moreover, recent advances in trauma care may have made some of the

decisions in the training data inappropriate for current use. A more useful prediction for the decision-maker would be to compare the function expected from a salvaged versus an amputated extremity, given the characteristics of the injury factors.

3. *Causes of outcomes:* Variables about sensation in the foot has been included in previous trauma models as a predictor of amputation even though this variable is known to indicate temporary nerve problems, therefore not recommended as a decision factor (see Section 4.2.2). Yet, some clinically important factors, such as nerve recovery and causes of nerve dysfunction, were ignored in the data-driven models as the data were not available. Considering the irreversible outcomes of amputation decisions, all relevant factors should be examined.

5.3.2 Acute Traumatic Coagulopathy

Acute traumatic coagulopathy (ATC) is one of the most critical risks regarding patient physiology in early stages of trauma care. Up to a quarter of trauma patients develop ATC soon after their injury. These patients have a considerably higher risk of bleeding and death since the body's protective mechanisms to limit bleeding are deranged. Several effective treatment options are available if ATC can be identified early. Immediate treatment is most effective however; standard laboratory tests to identify ATC take over an hour to produce results. The primary aim of the BN model is therefore to predict ATC with the information normally available within the first 10 minutes of care. It should be noted that the variables included in the BN are not limited to the ones that are available in the first 10 minutes; the predictions of the BN are generated by instantiating only those variables that can be observed in 10 minutes of care. The methodology we have described is relevant to this problem: the values of both ATC and of its causal factors are measured but none of the measurements are perfectly accurate.

5.3.3 ATC Bayesian Network

The initial structure of the BN, shown in Figure 5.2, was developed with domain experts using the AgenaRisk software (Agena Ltd, 2013). The BN structure contains two latent variables: ATC and Hypoperfusion. In addition, several other variables are

available in the training dataset but are usually unobserved in the first 10 minutes of treatment when the model is designed to be used. Each of these unobserved and latent variables is modelled with their measurements as naïve BN fragments or 'measurement idioms' (Neil et al., 2000). These naïve BN fragments were used as building blocks to form the BN structure, connected using causal relations elicited from experts. Table 5.1 shows the variables modelled with measurement idioms.

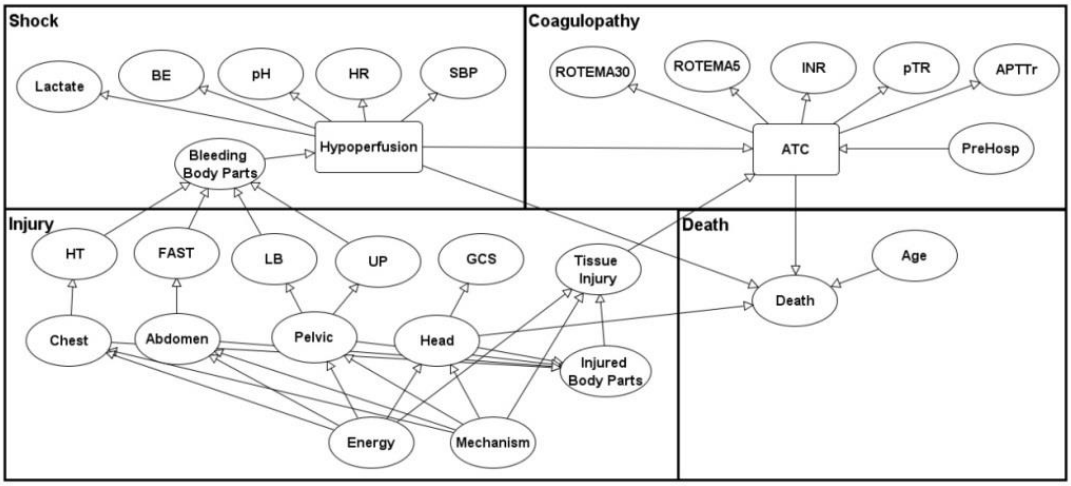


Figure 5.2 ATC BN

Table 5.1 Measurement Idioms in the ATC BN

Latent Variables	**Measurements / Markers**
ATC	ROTEMA5*, ROTEMA30*, INR*, PTR*, APTTR*
Hypoperfusion	Lactate, BE*, pH, SBP*, HR*
Variables Unobserved at First 10 Minutes	**Measurements / Markers**
Chest Injury	Haemothorax (HT)
Abdomen Injury	FAST* Scan
Pelvic Injury	Long Bone Injury (LB), Unstable Pelvis (UP)
Head Injury	Glasgow Coma Score (GCS)

**APTTR: Active partial thromboplastin time ratio, BE: Base excess, FAST: Focused assessment with sonography for trauma HR: Heart rate, INR: International normalised ratio, PTR: Prothrombin ratio, ROTEMA5 and A30: Amplitude of rotational thromboelastometry extem test at 5th and 30th minute, SBP: Systolic blood pressure*

The model is divided into four components, corresponding to the four boxes shown in Figure 5.2. The remainder of this section explains the variables and relations in each of these components briefly:

- *Coagulopathy:* the ATC variable has two states: 'Present' and 'Absent', and it can be estimated from 5 measurements. None of these measurements are available within the first 10 minutes but the variables are useful for model development. The main drivers of ATC are the degree of tissue injury and hypoperfusion. This may be aggravated by the infusion of large volumes of fluid (PreHosp).

- *Shock:* The hypoperfusion variable represents inadequate oxygen delivery to tissues as a result of blood loss, and it has three states: 'None', Compensated' and 'Uncompensated'. It can be estimated by 5 measurements: base excess (BE), lactate, pH, systolic blood pressure (SBP) and heart rate (HR). BE, lactate and pH are all relevant to the acidity of blood and they can be measured by a single, point-of-care, blood gas test. Blood gas test results are available within a few minutes. SBP and HR are continuously measured after admission to the hospital.

- *Injury:* The degree of overall tissue injury may not be known at the early stages of care. Overall tissue injury is estimated from the mechanism and energy of injury, and the number of severely injured body regions in the BN. Injury in each body part is estimated by mechanism and energy of injury, and also by clinical or radiological markers that would be expected to be available within 10 minutes of care: haemothorax (HT), FAST scan, long bone injuries (LB), unstable pelvic fracture (UP) and Glasgow coma score (GCS).

- *Death:* The model predicts death caused by physiological derangements, i.e. ATC and hypoperfusion. Age is an established independent predictor of death and has important effects on the physiological response to injury. Head injury is also a major cause of trauma deaths and thus the BN is refined to predict this (see Section 5.2.2).

Table 5.2 shows a brief description and the state-space of each variable in the ATC BN. The training and validation datasets of the ATC BN are described in Section 4.3.1.

Table 5.2 Variable Definitions and States in ATC BN

Variable	Description	States
ATC*	Acute Traumatic Coagulopathy	{Present, Absent}
ROTEM A30	Amplitude of ROTEM EXTEM at 30^{th} minute	Continuous
ROTEM A5	Amplitude of ROTEM EXTEM at 5^{th} minute	Continuous
INR	International normalised ratio	Continuous
PTR	Prothrombin ratio	Continuous
APTTr	Activated Partial Thromboplastin time	Continuous
PreHosp	Amount of liquids infused before admission to hospital.	{≥ 500ml, <500ml}
Hypoperfusion*	Decrease in the volume of blood perfusion to tissues.	{Uncompensated, Compensated, None}
BE	Base excess in blood	Continuous
pH	pH of blood	Continuous
Lactate	Amount of lactate in blood	Continuous
HR	Heart Rate	Continuous
SBP	Systolic Blood Pressure	Continuous
Bleeding Parts	Number of bleeding main body parts	{0, 1, 2, 3, 4}
Death	Risk of death in 48 hours	{Yes, No}
Age	Patient's age	{≥ 65, <65}
Tissue Injury	Severity of the overall tissue injury defined by injury severity score (ISS)	{Profound (ISS ≥ 45), Severe (45 > ISS ≥ 30), Moderate (30 > ISS ≥ 15), Mild (ISS < 15)}
Injured Parts	Number of severely injured main body parts	{0, 1, 2, 3, 4}
Chest	Severe chest injury	{Present (AIS ≥ 3), Absent (AIS < 3)}
Abdomen	Severe abdomen injury	{Present (AIS ≥ 3), Absent (AIS < 3)}
Pelvis & Extremity	Severe pelvis and extremity injury	{Present (AIS ≥ 3), Absent (AIS < 3)}
Head	Severe head injury	{Present (AIS ≥ 3), Absent (AIS < 3)}
Energy	Energy of Injury.	{Low, High}
Mechanism	Mechanism of Injury	{Penetrating, Blunt}
HT	Haemothorax	{Present, Absent}
UP	Unstable pelvis	{Present, Absent}
LB	Long bone injury	{Present, Absent}
GCS	Glasgow coma scale	Integer between [3,15]
FAST	FAST scan result	{Positive, Negative}

*ATC and Hypoperfusion variables were not available in the datasets.

5.3.4 Issues with ATC Measurements

The true state of ATC, which is the main outcome of our model and a crucial factor in trauma care, cannot be directly observed in practice, even after all the laboratory measurements have been completed. The ATC state is estimated using laboratory measurements such as the clotting time of a blood sample. However, none of these measurements can estimate the underlying ATC state with complete certainty. One measurement is the INR which is the normalised ratio of the clotting time of a patient's blood plasma to the clotting time of a healthy person. INR, and its clinically interchangeable measure prothrombin ratio (PTR), are the clinically accepted standard for diagnosing ATC (Frith et al., 2010). A normal INR value is 1, meaning that a patient has the same clotting time as a healthy person, and higher INR values indicate coagulation problems. However, there is not a clear borderline to distinguish normal coagulation from coagulopathy. Given that the actual mechanism of coagulation is complex and incompletely understood, INR and similar measurements have limitations that lead to uncertainty in the diagnosis of coagulopathy:

1. INR only tests blood plasma, disregarding other components essential to clotting such as the contribution made by platelets and the blood vessel wall.

2. INR does not measure the strength of a formed clot, the primary abnormality in ATC. It only measures the time it takes to form a clot.

3. INR is designed to monitor the effects of the drug Warfarin; it is not specifically designed for trauma.

Developing and validating a model that predicts INR values is convenient, but predicting INR is quite different from predicting the underlying coagulopathy state. For example, Mitra et al. (2011) used an INR of 1.5 as a threshold value for classifying ATC. However, a patient with an INR of 1.3 may have serious coagulation problems.

Consequently, the true underlying coagulopathic state of some patients cannot be known with certainty until a completely accurate way of measuring coagulopathy is discovered. Until then, clinicians will continue to estimate coagulopathy using their clinical judgement together with available measurements and observations. These

clinical judgements are not recorded in the hospital database. Only the data about INR and similar measurements are recorded in the dataset. The situation is similar for 'Hypoperfusion' which is the other latent variable in our model.

5.4 Learning

5.4.1 Initial Labelling with Expert Thresholds and Clustering

The latent variables were labelled twice using two different methods: first using measurement thresholds that reflect current clinical understanding (Brohi et al., 2003; Davenport et al., 2011; Frith et al., 2010), and then by clustering using the EM algorithm (2A and 2B of Figure 5.1). The thresholds used for labelling the ATC and Hypoperfusion variables are shown in Table 5.3. As a result of missing data, a number of patients could not be labelled. The labelling criteria for Hypoperfusion (see Table 5.3) are not complete so this state could not be labelled for several patients. Clustering was performed using the EM algorithm on the BN structure shown in Figure 5.2. EM uses all of the observed values and the BN structure to classify the data into coherent groups based on the maximum likelihood estimate of the latent variables. We used EM to classify the data into two coherent ATC states and three coherent hypoperfusion states.

Table 5.3 Criteria for Labelling ATC and Hypoperfusion from Measurements in Data

ATC		
No		*Yes*
INR≤ 1.2		INR> 1.2
Hypoperfusion		
None	*Compensated*	*Uncompensated*
BE≥-2 & Lactate≤2 & SI<0.9	SBP<90 & BT>4	BE<-4 & Lactate>4 & BT>0
Alive & BT in 12 Hours=0	-4≤ BE<-2 & 2<Lactate≤4 & BT>0	Pre-hospital cardiac arrest
Died in > 48 hours & BT=0		Death from haemorrhage

BT= Blood Transfusion in 12 Hours, SI = Shock Index, BE= Base Excess, SBP= Systolic Blood Pressure

5.4.2 Expert Review of the Labelling Differences

We compared the labels given by the measurement threshold and clustering approaches and prepared a list of the patients with differing labels, no label and a random subset of other cases. Three domain experts independently reviewed these cases and provided an expert label. All clinical information was available to the experts to assist labelling. The experts were blind to the labels assigned by the measurement threshold and EM clustering methods. The consensus between the experts' labels was assigned as the final label. Table 5.4 shows the number of cases reviewed for the two latent variables.

Table 5.4 Number of Cases Reviewed by Domain Expert

Hypoperfusion			**ATC**		
Label Differs	*No Label*	*Label Same*	*Label Differs*	*No Label*	*Label Same*
114	57	17	27	10	17
		Total: 188			*Total:* 54

This method required the domain experts to review 188 (31%) and 54 (9%) of the 600 cases respectively to label the hypoperfusion and ATC categories. Table 5.5 and Table 5.6 show the number of measurement threshold labels changed after the review: for example Table 5.5 shows that 6 patients classified as coagulopathic on the basis of the INR threshold were re-classified to non-coagulopathic by the expert review. At the end of this step, each latent variable had a single set of labels that were obtained from the combination of measurement threshold and clustering approaches, and the expert review of the differing labels.

Table 5.5 Measurement Threshold ATC Labels Changed by Expert

ATC Label Review - Measurements			
	After Review		
Measurements	*Yes*	*No*	*Unlabelled*
Yes	57	6	-
No	3	524	-
Unlabelled	1	5	4
			Total: 600

Table 5.6 Measurement Threshold Hypoperfusion Labels Changed by Expert

Hypoperfusion Label Review – Measurements				
	After Review			
Measurements	*Uncomp.*	*Comp.*	*None*	*Unlabelled*
Uncomp.	62	9	4	-
Comp.	1	52	5	-
None	1	6	403	-
Unlabelled	-	17	35	5
				Total: 600

5.4.3 Learning and Cross-Validation

The result of the expert review (step 3 of Figure 5.1) is a dataset now including values for the latent variables for almost all patients. The ATC value of 4 patients and Hypoperfusion value of 5 patients remained unlabelled after the expert review because the expert was not confident about the correct value. We used the standard EM algorithm to learn the parameters of the model. The performance of the model trained on the RLH data was tested by 10-fold cross validation. Only the variables that can be observed in the first 10 minutes of treatment are instantiated for generating the predictions in 10-fold cross validation.

Performance of a model can be measured in terms of its discrimination, calibration and accuracy. Discrimination measures whether the model can distinguish the patients with the event. A model that has well discriminatory performance gives higher probabilities to the patients with the event, and lower probabilities to the patients without the event. Calibration measures whether the predicted probability represents the correct probability on average. For example, when a model predicts 10% chance of survival for a group of patients, 10% of these patients are expected to survive if the model is well calibrated. Accuracy measures whether the predicted outcomes are close to the actual outcomes by combining features of discrimination and calibration. Medlock et al. (Medlock et al., 2011) recommends using multiple performance measures to quantify different aspects of the model performance.

We used multiple performance measures to assess the discrimination, accuracy and calibration of the ATC BN as recommended by the Medlock framework (Medlock et al., 2011). The discrimination of the ATC BN was evaluated with receiver operating characteristic (ROC) curves, sensitivity and specificity values. The area under the

ROC curve (AUROC) is 0.90 and 0.81 for the prediction of ATC and death respectively. Brohi (2011) argues that a useful prediction model for coagulopathy must operate with at least 90% sensitivity: the BN achieves specificities of 71% for ATC and 44% for death when operating with 90% sensitivity. The initial performance of the model on the cross-validation dataset can be seen in Table 5.7.

Table 5.7 Initial Cross Validation Results

	ATC	Death
AUROC	0.90	0.81
Specificity*	71%	44%
Specificity**	83%	67%
Brier Score	0.06	0.09
Brier Skill Score	0.32	0.15

*At 0.90 sensitivity **At 0.80 Sensitivity

The accuracy of the BN was evaluated with the Brier score (BS) and Brier skill score (BSS) (Brier, 1950; Weigel et al., 2007). BS is the mean squared difference between the predicted probability and actual outcome. The score can take values between 0 and 1; 0 indicates a perfect model and 1 is the worst score achievable. BSS measures the improvement of the model's prediction relative to a reference prediction which is often the average probability of the event in the data. BSS can take values between negative infinity and 1; a negative value indicates a worse prediction than the average probability and 1 indicates a perfect model. The BN has BS of 0.06 and BSS of 0.32 for ATC predictions, BS of 0.09 and BSS of 0.15 for death predictions.

The calibration of the BN was assessed with the Hosmer-Lemeshow test (Hosmer and Lemeshow, 1980). This test divides the data into 10 subgroups, and calculates a chi-square statistic comparing the observed outcomes to the outcomes expected by the model in each subgroup. Low p-values indicate a lack of calibration. Hosmer-Lemeshow test is strongly influenced by the sample size. In large datasets, small differences between the expected and observed outcomes can lead to low p-values but the visual representation of this test provides a concise summary of the model calibration.

The BN was well calibrated for both ATC and death predictions with Hosmer-Lemeshow statistics of 9.7 (p=0.29) and 6.7 (p=0.57) respectively. Figure 5.3 is a visual representation of the model's calibration for ATC predictions. The similarity

between the expected and true outcomes in each subgroup shows that the model was well calibrated.

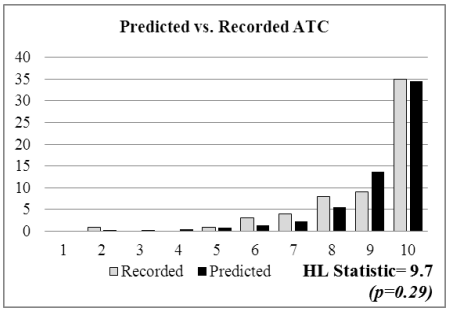


Figure 5.3 Model Calibrations for ATC Predictions

5.4.4 Inaccurate Predictions and Unexpected Clinical Outcomes

After the learning and cross-validation steps, we reviewed the inaccurate predictions of the model with the domain experts (step 5 of Figure 5.1). We divided the predictions, given by cross validation of the model, into ten bins according to the predicted probability, and prepared a contingency table that compares the predictions of the model to the outcome values in data for each bin as shown in Table 5.8.

Table 5.8 Predictions and Recorded Outcomes

ATC Prediction	**Outcome in Data**	
P	*ATC=Yes*	*ATC=No*
1.0 > *P* ≥ 0.9	0	0
0.9 > *P* ≥ 0.8	1	**1**
0.8 > *P* ≥ 0.7	5	**1**
0.7 > *P* ≥ 0.6	15	**5**
0.6 > *P* ≥ 0.5	7	***8***
0.5 > *P* ≥ 0.4	7	***8***
0.4 > *P* ≥ 0.3	1	***8***
0.3 > *P* ≥ 0.2	7	***25***
0.2 > *P* ≥ 0.1	6	***40***
0.1 > *P* ≥ 0.0	***12***	440
Total	61	535

The negative outcomes with ATC prediction over 0.1, and the positive outcomes with ATC prediction less than 0.1 (shown in bold in Table 5.8) were considered as the possibly inaccurate predictions since 10% of the patients were initially labelled with ATC and thus 0.1 was our prior probability. A clinician reviewed the data and

patient notes for each of these 108 cases and analysed the possible causes of each unexpected prediction:

a. **Expert agrees with the prediction:** The actual outcome is unexpected, possibly requiring further clinical investigation. Another possible explanation is incorrectly recorded data.

b. **Expert expects the recorded outcome:** The model was considered to be making inaccurate predictions for these cases. The clinician decided that the outcome value in the data is clinically expected and analysed the causes of the inaccurate predictions. These inaccuracies could be caused by an error in the model structure.

Table 5.9 gives a summary of this review: the domain experts agreed with about a third of the apparently inaccurate predictions. During the review, domain experts explained why they agreed with the individual predictions or recorded outcomes which led to a number of refinements to the model and to the clinicians' understanding of the data. Death predictions were also reviewed by the same approach. We describe these issues and the way the model was refined in the following section.

Table 5.9 Inaccurate Predictions and Expert Review

ATC Prediction P	Prediction differs from the recorded outcome	Expert agrees with the prediction
$0.9 > P \geq 0.8$	1	0 (0%)
$0.8 > P \geq 0.7$	1	1 (100%)
$0.7 > P \geq 0.6$	5	5 (100%)
$0.6 > P \geq 0.5$	8	4 (50%)
$0.5 > P \geq 0.4$	8	5 (63%)
$0.4 > P \geq 0.3$	8	6 (75%)
$0.3 > P \geq 0.2$	25	8 (32%)
$0.2 > P \geq 0.1$	40	6 (15%)
$0.1 > P \geq 0.0$	12	2 (17%)

5.5 Model Refinement

Three issues were found from the review of inaccurate predictions:

1. ATC may develop in some patients soon after the blood test used for INR and other measurements was taken.

2. Some of the deaths recorded in the dataset were most likely due to conditions other than ATC.

3. There are mechanisms of coagulopathy other than ATC that may be occurring in patients in the dataset.

These issues all challenge the supposed objectivity of data and reinforce the need to combine data with expert review. The following sections describe these issues in more detail.

5.5.1 Incipient Coagulopathy

A group of patients who had normal values for their initial ATC measurements (see Table 5.1) showed significant signs of ATC in a second set of measurements that were conducted soon after. Moreover, these patients had severe injury burden and poor perfusion; they were therefore at high clinical risk of developing ATC. The ATC model predicts high risk of coagulopathy for these patients but the value in the data is negative since only the initial measurements were considered while labelling the ATC state of patients with measurement thresholds and clustering approaches.

Coagulopathy is a dynamic phenomenon that develops in time, so the results of measurements are dependent on the time they are carried out. Variations in the interval between the injury and the arrival at the hospital add further uncertainty. Therefore, the domain experts considered the prediction of those patients with ‘incipient coagulopathy’ as a clinically useful feature of the BN.

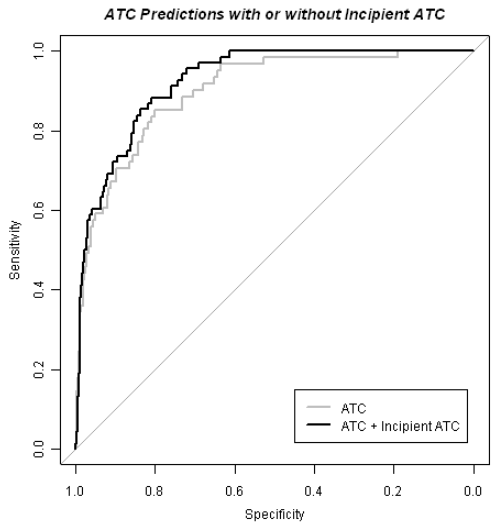


Figure 5.4 Predictions with Incipient Coagulopathy

We relearnt the ATC BN and recalculated its performance in a cross validation when patients with incipient ATC were also considered as positive outcomes. The structure of the ATC BN was not changed in this analysis. Figure 5.4 compares the ROC curves for ATC prediction based on only the initial measurements with the one for patients with incipient ATC. The AUROC is 0.92, and the model achieves specificity of 79% with sensitivity of 90%, BS of 0.06 and BSS of 0.39.

Prediction of incipient coagulopathy shows the difference between the clinically useful models and the models that predict measurements in data well. The patients with incipient coagulopathy would count as incorrect predictions for a purely data-based approach, and such an approach would try to change the parameters to 'correct' these predictions. In contrast, the expert was able to explain the apparent anomaly and show that predicting incipient coagulopathy was useful; this was not obvious at the beginning of the model development.

5.5.2 Other Causes of Death

The review revealed that a large proportion of deaths that could not be predicted by the BN were the result of head injuries and thus these deaths were expected by the domain experts. The ATC model is designed to predict risk of death relevant to bleeding and coagulopathy, so the initial model does not predict deaths related to head injuries. However, the model structure is easily modified to predict these deaths since we already have a head injury variable in the model which is used to estimate

the overall tissue injury burden of patients. By adding an arc between head injury and death variable, we increase the accuracy of the model for death prediction. Although death might be considered to be the least ambiguous outcome in a clinical dataset, our experience shows that this is not the case when there is a mismatch between the modelled and actual cause of death.

This simple modification increased the accuracy of death predictions significantly. The AUROC increased from 0.81 to 0.88 as shown in Figure 5.5. The specificity of the BN is increased from 44% to 72% when it is operated at 90% sensitivity level. BS and BSS also indicated an increased accuracy in the death predictions: BS of the BN decreased from 0.09 to 0.08; and BSS increased from 0.15 to 0.23. This change had no impact on ATC prediction.

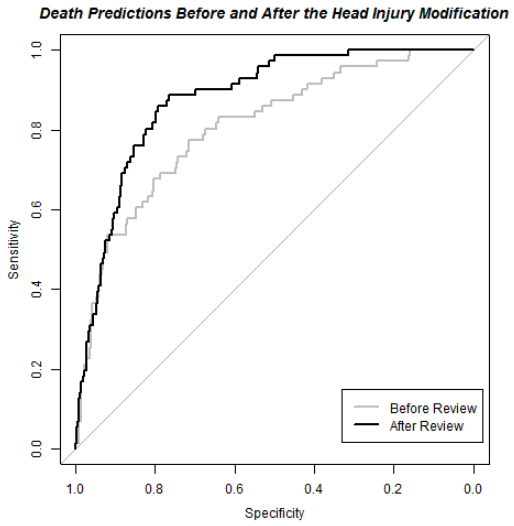


Figure 5.5 Predictions with Head Injury Modification

5.5.3 Unmodelled Mechanisms of Coagulopathy

The aim of this BN is to predict ATC, which is driven by a combination of the degree of tissue injury and hypoperfusion following traumatic injuries. The scope of the model has to be clearly defined since other forms of coagulopathy exist. For example, the anticoagulation medicine Warfarin makes a person coagulopathic without any traumatic injury, and predicting drug induced coagulopathy is out of scope for this model.

Another important cause of traumatic coagulopathy is a catastrophic brain injury. These injuries seem to effect coagulation via a different mechanism to ATC. The review of unexpected predictions showed that 9 of the 12 coagulopathic patients that the BN model could not accurately predict had severe head injuries, and in 7 of these patients brain injury was fatal. It is likely that these patients were suffering from a coagulopathy caused by brain injury (BIC) rather than ATC.

BIC is now documented as being outside the scope of our BN. This issue was not clear at the beginning of the model development even though the clinicians were aware of the phenomenon; it was identified as a result of the review of inaccurate predictions with the domain expert. If prediction of the BIC is required by the users, the model structure can be adapted accordingly by adding two variables 'BIC' and 'Coagulopathy' as shown in Figure 5.6. In this model fragment, 'Coagulopathy' variable represents the overall coagulopathy risk that sums the risk of 'BIC' and 'ATC' variables.

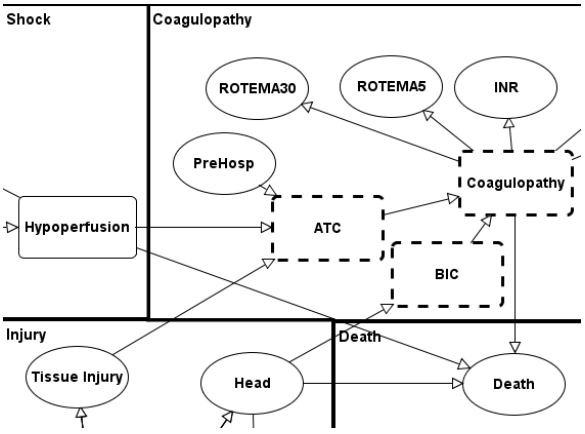


Figure 5.6 BN Structure Refined for Brain Injury Induced Coagulopathy

5.6 Temporal and External Validation

Further validation of the model, with the head injury and incipient coagulopathy modifications, was done on the test and external datasets (see Section 4.3.1 for a description of these datasets). These datasets are composed of exactly the same variables as the training dataset as all datasets were collected as a part of an international collaboration. The ATC states in the test and external datasets were labelled using the methodology described in Section 5.2.

A temporal validation of the ATC BN was done using the test dataset. In the temporal validation, the AUROC is 0.94 for the ATC predictions, and 0.92 for the death predictions. At 90% sensitivity level, the BN achieves 92% specificity for predicting ATC, and 79% specificity for predicting death (see Table 5.10). The ATC BN was well calibrated for both ATC and Death predictions at the temporal validation (see Figure 5.7a and Figure 5.7b)

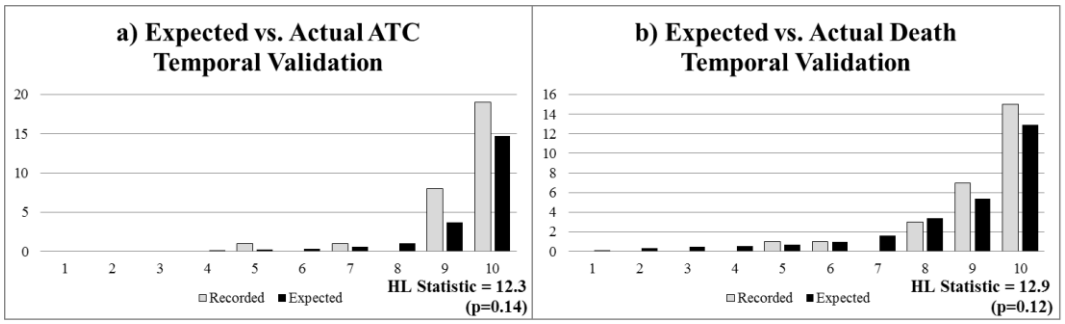


Figure 5.7 Calibration of a) ATC b) Death predictions at Temporal Validation

An external validation of the ATC BN was done using the external dataset. In the external validation, the AUROC is 0.90 for ATC predictions, and 0.91 for death predictions. The BN achieves 88% specificity and 90% sensitivity for ATC predictions; 84% specificity and 90% sensitivity for death predictions (see Table 5.10). The calibration for ATC and Death predictions on the external data is shown in Figure 5.8a and Figure 5.8b respectively. The model was well calibrated for ATC predictions. For death predictions, the discrimination of the model was accurate but there were more deaths in the external data than it was expected by the model. A part of these deaths were caused by factors such as neck and chest injuries which are outside the scope of the ATC BN.

Table 5.10 Temporal and External Validation Results

	Temporal Validation		**External Validation**	
	ATC	**Death**	**ATC**	**Death**
AUROC	0.94	0.92	0.91	0.90
Specificity*	92%	79%	88%	81%
Specificity**	94%	87%	90%	84%
Brier Score	0.05	0.06	0.07	0.10
Brier Skill Score	0.48	0.30	0.27	0.30

*At 0.90 sensitivity **At 0.80 Sensitivity

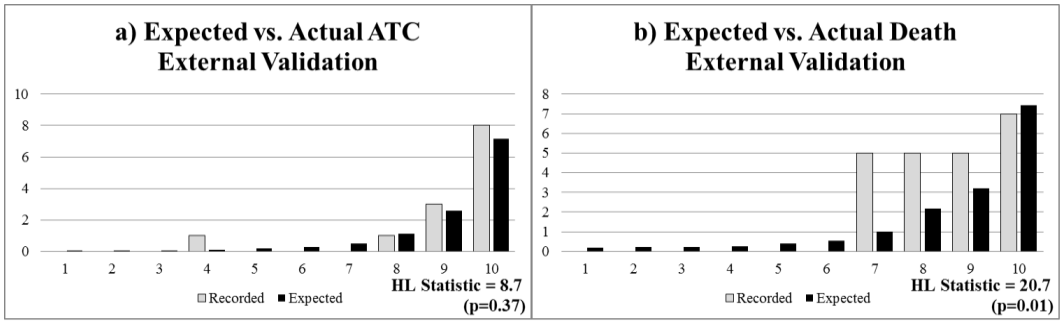


Figure 5.8 Calibration of a) ATC b) Death predictions at External Validation

5.7 Conclusion

This chapter proposed a method for developing and refining BNs with latent clinical conditions, using a combination of expert knowledge and data. The method is successfully applied to a clinical case study about the prediction of ATC in trauma care. Our method addresses the problems related to measurement errors and causes of outcomes by:

1. Making a clear distinction between a latent variable that we wish to predict for decision support and any measurements of this variable that may be recorded in a dataset; both latent and observed variables are represented explicitly in the BN model.

2. Using iterative expert review of the model to refine the model and to understand the relationship between the data and the real decision problem.

Our methodology systematically integrated domain expertise into model development at two stages. Firstly, the 'true' but unobserved state was added to a dataset by combining labelling by observed measurements with data clustering in an expert-elicited BN structure. Focussing the detailed expert review on the cases labelled differently in these two steps saves time compared to a review of all cases. Secondly, the experts examined differences between the model's predictions and the data.

In our case study, this examination revealed several issues initially neglected by our experts and emphasised the difference between useful predictions for the decision-maker and an accurate prediction of measurements in data. Other latent and observed

causes of predicted outcomes, which were not clear at the beginning, were modelled during the review. These issues were resolved either by refining the model or by acknowledging the scope of its applicability, which were not obvious at the initial stages of model development.

The case study demonstrates significant improvements in predictions from the iterative expert reviews and refinements. Identifying and including the other causes of death increased the specificity for death predictions from 44% to 72% when the model is operated at 90% sensitivity. Similarly, identifying the clinically important patient subgroup with incipient coagulopathy increased the specificity for ATC predictions from 72% to 79% at 90% sensitivity. The ATC BN performed well in the temporal and external validations.

Chapter 6

Building Bayesian Networks using the Evidence from Meta-analyses

Meta-analysis is an important statistical tool for EBM as it combines evidence from multiple studies to infer the overall effect and variation. By combining the results from individual datasets, meta-analysis can provide evidence based on a larger number of patients compared to the size of the individual datasets. This is beneficial especially for uncommon medical conditions, such as mangled extremities, where data is typically available in small amounts. Chapter 4 illustrated the need for combining evidence from different sources to develop useful decision support models in this domain.

However, many medical studies report 'univariate' relations between a single factor and a single outcome. RCTs, for example, analyse the effect of a single treatment by using randomisation to decrease the confounding effect of other variables. Similarly, many observational studies report the relation between individual risk factors and outcomes even when the dataset contains information about multiple factors. Moreover, medical studies rarely present their raw data (Vickers, 2006).

Meta-analysis can effectively combine evidence about univariate relations but decision support from such evidence is limited in most circumstances. Clinical decisions are often complex (Buchan et al., 2009). They require decision maker to evaluate multiple factors that may interact with each other. For example, separate meta-analyses can be conducted to combine the evidence about the individual effect

of a treatment and a comorbidity factor. However, if the treatment and comorbidity factor interacts with each other, their joint effect may be completely different from their individual effects. In this case, decision support provided by the meta-analysis of individual effects may be invalid for a patient who is exposed both to the treatment and the comorbidity factor (see Marshall (2006) and Rawlins (2008) for a discussion of generalisation of clinical evidence). More useful decision support can be provided by combining the evidence about all plausible causes and interaction effects. Statistical techniques, such as multivariate meta-regression, are available for combining evidence about multivariate relations, but clinical studies rarely publishes information detailed enough to use for these techniques (Vickers, 2006).

In this chapter, we present a methodology for building decision support models that reason consistent with the best available evidence and accounts for the complexity of clinical decisions. Our methodology aims to build BNs based on the evidence from meta-analysis, expert knowledge and data. As discussed in Chapters 2, 3 and 5, BNs offer a powerful framework for providing decision support by combining different sources of evidence. However, a systematic methodology to build a BN by using meta-analysis results has not been proposed. Our methodology combines the evidence from meta-analysis with expert knowledge and data to define the structure and parameters of a BN. Our methodology uses auxiliary BNs to learn the parameters of the BN used for decision support. We apply this methodology to develop a BN that predicts the short-term viability outcomes of lower extremities with vascular trauma.

In the remainder of this chapter, Section 6.1 summarises the meta-analysis techniques. Section 6.2 describes our methodology for defining BN structure and parameters based on the results of a meta-analysis. Section 6.3 and 6.4 present the application and results of this method for the case-study, and Section 6.5 presents the conclusions.

6.1 Meta-analysis

Meta-analysis is a statistical method for combining evidence from different RCTs or observational studies. It is often used as a part of systematic literature reviews to combine the statistics from reviewed studies.

A RCT or an observational study may compare the outcome of patients who were exposed to a treatment or a risk factor against those who were not exposed to this factor. For example, a researcher, who aims to investigate the effects of bone fractures to the outcomes of lower extremity surgery, examines the fractured lower extremities N_P in the data. He records the number of fractured extremities that had a successful operation S_P, and the number of extremities that had a failed operation F_P. Afterwards, he examines the lower extremities that were not fractured N_A, and records the number of successful S_A and failed outcomes F_A among these extremities (see Table 6.1).

Table 6.1 Numbers Presented in the Example about Mangled Extremity

	Success	**Fail**	**Total**
Fracture - Present	S_P	F_P	$N_P = S_P + F_P$
Fracture - Absent	S_A	F_A	$N_A = S_A + F_A$

The results of this study can be presented in several ways including counts, conditional probabilities and odds ratios. Perhaps the crudest way is to present the counts. The counts must be transformed into conditional probabilities or other forms of ratios in order to compare and combine results from multiple studies. The results can be presented by two conditional probabilities: the probability of a successful outcome given a fractured lower extremity $P(S|P) = S_P / N_P$ and the probability of a successful outcome given a non-fractured lower extremity $P(S|A) = S_A/N_A$. Rather than presenting two conditional probabilities, the effect of a bone fracture can also be summarised as a single odds ratio. The odds ratio can be calculated by dividing the odds of having a successful outcome with a fractured extremity by the odds of having the same outcome with a non-fractured extremity:

$$Odds\ Ratio = \frac{S_P/F_P}{S_A/F_A}$$

Meta-analysis can be used to combine various statistics including conditional probabilities and odds ratios. Since BN parameters are composed of conditional probabilities, we focus on the meta-analysis of conditional probabilities in the remainder of this chapter.

6.1.1 Fixed and Random Effects Meta-analysis

Evidence can be combined based on either a fixed effect or a random effects model in a meta-analysis. The fixed effect model assumes that all of the individual studies in the analysis share the same true effect and that there is no variation between the studies (see Figure 6.1a). Therefore, the individual studies are expected to be centred on the true effect value, and the only source of variation is dependent on the sample size of the studies. However, it is often implausible to assume, especially for observational studies, that a single effect is common to all studies. The random effects model assumes that several known or unknown factors may cause variations (heterogeneity) in the true effect size between the studies. The random-effects model accounts for variation between the studies as well as variation within the studies (see Figure 6.1b).

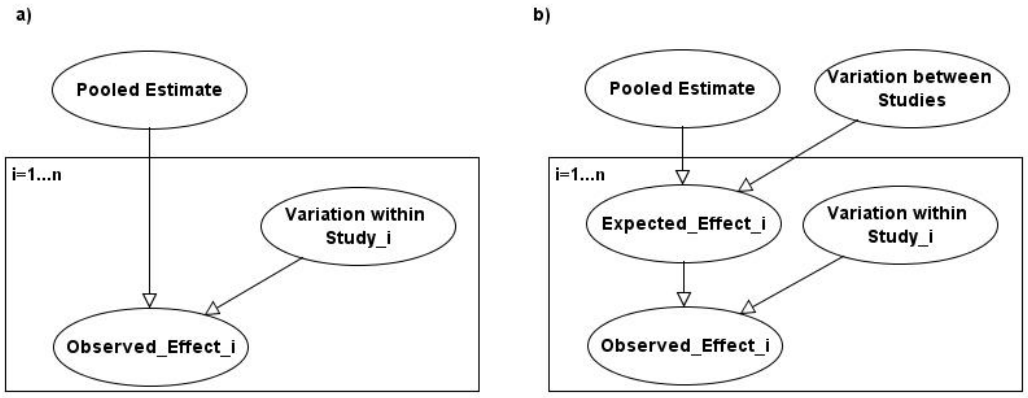


Figure 6.1 Illustration of a) Fixed Effects b) Random Effects Models in Meta-analysis

The assumptions behind the fixed effects model are not realistic in most real-world cases. Observational studies, for example, are likely to be heterogeneous. Besides, large observational studies are not necessarily preferable to smaller ones as their data may contain less detail and more errors (Egger et al., 2001). In the fixed effect model, the contribution of each study to the pooled estimate is weighted by their sample sizes: the studies with larger sample size accounts for most of the combined

evidence whereas small studies may practically have no effect. The random-effects meta-analysis is more conservative in allocating weight to sample sizes as it also takes heterogeneity into account.

6.1.2 Bayesian meta-analysis

Meta-analysis can be conducted using either a frequentist or a Bayesian approach. Computation of the frequentist approach is simpler and readily implemented in popular statistical software such as SPSS and R. The Bayesian approach has several advantages over the frequentist approach including the following (Sutton and Abrams, 2001):

1. It offers a unified modelling framework to model the variation between and within the studies.
2. The results of the meta-analysis can be presented in a predictive distribution that takes both heterogeneity and the uncertainty of the pooled estimate into account.
3. Individual study effects do not necessarily follow the normal distribution in Bayesian meta-analysis models.
4. Prior information can be included into the analysis. Priors must be chosen with care: it is often useful to conduct a sensitivity analysis for different prior alternatives.

6.1.3 A Bayesian meta-analysis model for combining probabilities

Conditional probabilities from multiple studies can be combined using the Bayesian meta-analysis model shown in Figure 6.. This model takes the variation between the studies into account, and it does not assume normality for the distribution of individual studies.

The binomial distribution is the probability distribution of the number of positive outcomes in n independent experiments where the probability of a positive outcome is p for every experiment. In this model, the result of each individual study i is modelled with the binomial distribution:

$$r_i \sim Binomial(p_i, n_i)$$

Where r_i is the number of positive outcomes observed in the study i, p_i is the true study probability of the study i, and n_i is the sample size of the study i.

When combining estimates from different studies, a transformation can be used to model the pooled estimates with the normal distribution. For probability values, the logit transformation can be used for this purpose. The normal distribution is convenient for modelling the pooled estimate and variation between studies. In our model, the logit transformation of the true study probability p_i follows the normal distribution. The mean μ of this distribution represents the transformed pooled estimate, and the variance τ^2 represents the variation between studies.

$$logit(p_i) = \theta_i$$

$$\theta_i \sim Normal(\mu, \tau^2)$$

The predictive distribution of the conditional probability for a future study can also be calculated by a logit transformation.

$$\theta_{new} \sim Normal(\mu, \tau^2)$$

$$logit(p_{new}) = \theta_{new}$$

Finally, priors need to be chosen for the mean and variance of the normal distribution. The ignorant priors shown below can be used if informative priors are not available.

$$\mu \sim Normal(0{,}1000)$$

$$\tau \sim Uniform(0{,}2)$$

In order to calculate the posteriors of μ, τ^2 and p_{new}, we enter the observed number of positive outcomes r_i and sample sizes n_i from each reviewed study. The posteriors can be calculated by using the dynamic discretisation algorithm (Neil et al., 2007) in the AgenaRisk software (Agena Ltd, 2013) or the Markov Chain Monte Carlo (MCMC) sampling technique in the OpenBUGS software (Lunn et al., 2009).

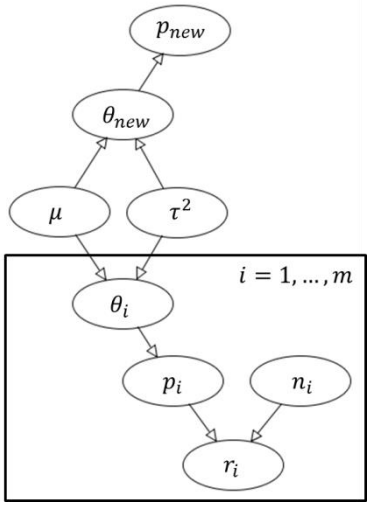


Figure 6.2 Bayesian Meta-analysis model for pooling proportions

6.2 Building BNs based on Meta-analysis

Previous section described a Bayesian meta-analysis technique for pooling proportions. In this section, we present a methodology that uses the evidence from a meta-analysis of proportions to define the structure and parameters of a BN decision support model. Our methodology integrates the evidence from meta-analysis with data and clinical knowledge to build the BN. Our methodology assumes that expert knowledge, a meta-analysis of univariate relations from a relevant systematic review and some data about multivariate relations is available. However, the data is not large enough to learn the behaviour of some of the relations in the BN.

6.2.1 Structure

Development of a BN structure can be defined in two stages: selecting variables, and identifying the relations between those variables. Our aim is to use clinical evidence in both of these stages.

In our method, domain experts use the meta-analysis as a guide for selecting the important variables for the BN. They review every variable that is found to have a clinically significant effect in the meta-analysis. During the review, they define the mechanistic relations between each of these variables and the outcome, and describe how these variables clinically affect the outcomes. Expert knowledge about the

mechanistic relations allows us to 1) build a causal BN structure based on clinical knowledge 2) examine whether each variable is within the scope of the model.

The mechanistic relations between the observed factors and outcomes often depend on several clinical factors that cannot be observed. In this case, latent variables may be required to represent the clinical knowledge in the BN. For example, a meta-analysis may show that a complicated surgery has worse outcomes than its alternatives. A decision support model, with only the surgery and outcome variables, may therefore never recommend this surgery. However, the reason for the worse outcomes can be that the surgery is only applied to patients with more severe conditions. Even though its outcomes are worse compared to the average, it may perform better in patients with severe conditions. Such knowledge can only be provided by domain experts as the meta-analysis contains only the univariate effects. Moreover, a latent variable representing the severity of injury is required to model this knowledge in the BN structure.

Knowledge about mechanistic relations also allows knowledge engineers to understand whether each variable is within the scope of the BN. Some variables may be irrelevant considering the aims and scope of the model even though they have significant effects in the meta-analysis. For example, although the complicated surgery performs well for severe conditions, it may not be included in the BN if such patients are outside its scope.

6.2.2 Parameters

Since most studies, especially in the medical domain, publish the results about univariate relations, the meta-analysis of such studies provides a probability conditioned on a single variable, such as $P(Y|X_1)$. Such probability distributions can be used for the parameters of BN variables with a single parent but BNs often contain variables with multiple parents. The parameters of these variables require probabilities conditioned on multiple variables such as $P(Y|X_1, \ldots, X_n)$. The probability distributions from a meta-analysis of univariate relations cannot be used for such BNs.

In this section, we present a parameter learning method for combining the results of a meta-analysis with data to learn the parameters of a BN variable with multiple parents. The Bayesian framework of this method assumes that the data generating process of the reviewed studies is similar, but not necessarily the same as the available data. The differences between the subpopulations of the data and the previous research must be evaluated before applying this method in order to avoid developing erroneous BNs (Druzdzel and Díez, 2003). In the remainder of this section, we illustrate the proposed method by a simple example in Section 6.2.2.1, and we present the generalisation of the method in Section 6.2.2.2.

6.2.2.1 Illustration of the Parameter Learning Method

In this section, we illustrate our parameter learning method with the simple BN shown in Figure 6.3.

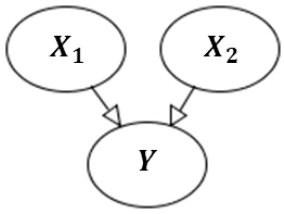


Figure 6.3 Simple BN for Illustrating the Parameter Learning Method

This BN has 3 variables and each of the variables has 2 states:

$$X_1 = \{x_1^1, x_1^2\}$$

$$X_2 = \{x_2^1, x_2^2\}$$

$$Y = \{y^1, y^2\}$$

Table 6.2 NPT of the Variable Y

	$X_1 = x_1^1$ $X_2 = x_2^1$	$X_1 = x_1^1$ $X_2 = x_2^2$	$X_1 = x_1^2$ $X_2 = x_2^1$	$X_1 = x_1^2$ $X_2 = x_2^2$
$\boldsymbol{Y = y^1}$	$P(y^1 \mid x_1^1, x_2^1)$	$P(y^1 \mid x_1^1, x_2^2)$	$P(y^1 \mid x_1^2, x_2^1)$	$P(y^1 \mid x_1^2, x_2^2)$
$\boldsymbol{Y = y^2}$	$1 - P(y^1 \mid x_1^1, x_2^1)$	$1 - P(y^1 \mid x_1^1, x_2^2)$	$1 - P(y^1 \mid x_1^2, x_2^1)$	$1 - P(y^1 \mid x_1^2, x_2^1)$

Table 6.2 shows the NPT of the variable Y. We require 4 parameters for this NPT: $P(y^1|x_1^1, x_2^1)$, $P(y^1|x_1^1, x_2^2)$, $P(y^1|x_1^2, x_2^1)$ and $P(y^1|x_1^2, x_2^2)$.

The parameters of the variable Y can be learnt from data using the maximum likelihood estimate (MLE) approach. For example, $P(y^1|x_1^1, x_2^1)$ can be estimated by dividing $M[y, x_1^1, x_2^1]$ to $M[x_1^1, x_2^1]$, where $M[y, x_1^1, x_2^1]$ represents the count of data instances where $Y = y^1$, $X_1 = x_1^1$ and $X_2 = x_2^1$, and $M[x_1^1, x_2^1]$ represents the count of data instances where $X_1 = x_1^1$ and $X_2 = x_2^1$.

$$P(y^1|x_1^1, x_2^1) = \frac{M[y^1, x_1^1, x_2^1]}{M[x_1^1, x_2^1]}$$

Suppose we have a dataset with a sample size of $M = 250$ to learn the parameters of the BN in Figure 6.3. Table 6.3 shows a part of the relevant counts from this imaginary dataset. There is only 3 data instances where $Y = y^1$, $X_1 = x_1^1$ and $X_2 = x_2^2$ as shown by the first row of this table.

Table 6.3 Some Relevant Counts from the Data

Counts in the Data	
$\boldsymbol{M[y^1, x_1^1, x_2^1]}$	3
$\boldsymbol{M[y^1, x_1^1, x_2^2]}$	25
$\boldsymbol{M[x_1^1, x_2^1]}$	10
$\boldsymbol{M[x_1^1, x_2^2]}$	160
$\boldsymbol{M[x_2^1]}$	230
$\boldsymbol{M[x_2^2]}$	20
$\boldsymbol{M}$	250

Our aim is to estimate the parameters of the BN. Although the overall sample size of the data is not small, there is not an adequate amount of data for learning some of the parameters. For example, there are only a few data instances to learn the probability of $P(y^1|x_1^1, x_2^1)$ since $M[y^1, x_1^1, x_2^1] = 3$ and $\mathrm{M}[x_1^1, x_2^1] = 10$.

As well as the data, suppose we have the results of a meta-analysis that analyses the relation between Y and X_1. This meta-analysis pools the conditional probabilities $P(y^1|x_1^1)$ reported in multiple studies. The result of the meta-analysis is reported by the mean, $\mu_{pnew}(y^1|x_1^1)$, and variance, $\sigma^2_{pnew}(y^1|x_1^1)$, of the predictive distribution of the pooled conditional probability (see Table 6.4). The way that these statistics are calculated is described in Section 6.1.3.

Table 6.4 Predictive Distribution Parameters from the Meta-analysis

Meta-analysis of $P(y^1\|x_1^1)$ Predictive Distribution Parameters	
$\mu_{pnew}(y^1\|x_1^1)$	0.2
$\sigma^2_{pnew}(y^1\|x_1^1)$	0.005

The results of the meta-analysis cannot be directly used for the BN parameters since the variable Y is conditioned on both X_1 and X_2 in the BN model whereas it is conditioned only on X_1 in the meta-analysis. In other words, there is no parameter to use $P(y^1|x_1^1)$ directly in the NPT of the variable Y (see Table 6.2).

In the remainder of this section, we present a novel technique that combines the data shown in Table 6.3 and the meta-analysis results shown in Table 6.4 to learn the parameters $P(y^1|x_1^1, x_2^1)$ and $P(y^1|x_1^1, x_2^2)$ for the NPT of the variable Y. The generalisation of this method for a larger number of parents and states is described in Section 6.2.2.2.

Figure 6.4 shows a BN representation of the implemented technique. The BN representation is divided into five components that are described in the remainder of this section:

1. **Data:** This part uses the binomial distribution to model the relation between the CPDs that we aim to estimate and the observed counts in the data. For example, the number of data instances where $X_1 = x_1^1$, $X_2 = x_2^2$ and $Y = y^2$, shown by $M[y^1, x_1^1, x_2^1]$, has a binomial distribution where the probability parameter is $P(y^1|x_1^1, x_2^1)$ and the number of trials parameter is $M[x_1^1, x_2^1]$. The binomial distributions used in this part are shown below:

$$M[y^1, x_1^1, x_2^1] \sim Binomial\big(M[x_1^1, x_2^1], P(y^1|x_1^1, x_2^1)\big)$$

$$M[y^1, x_1^1, x_2^2] \sim Binomial\big(M[x_1^1, x_2^2], P(y^1|x_1^1, x_2^2)\big)$$

$$M[x_2^1] \sim Binomial(M, P(x_2^1))$$

$$M[x_2^2] \sim Binomial(M, P(x_2^2))$$

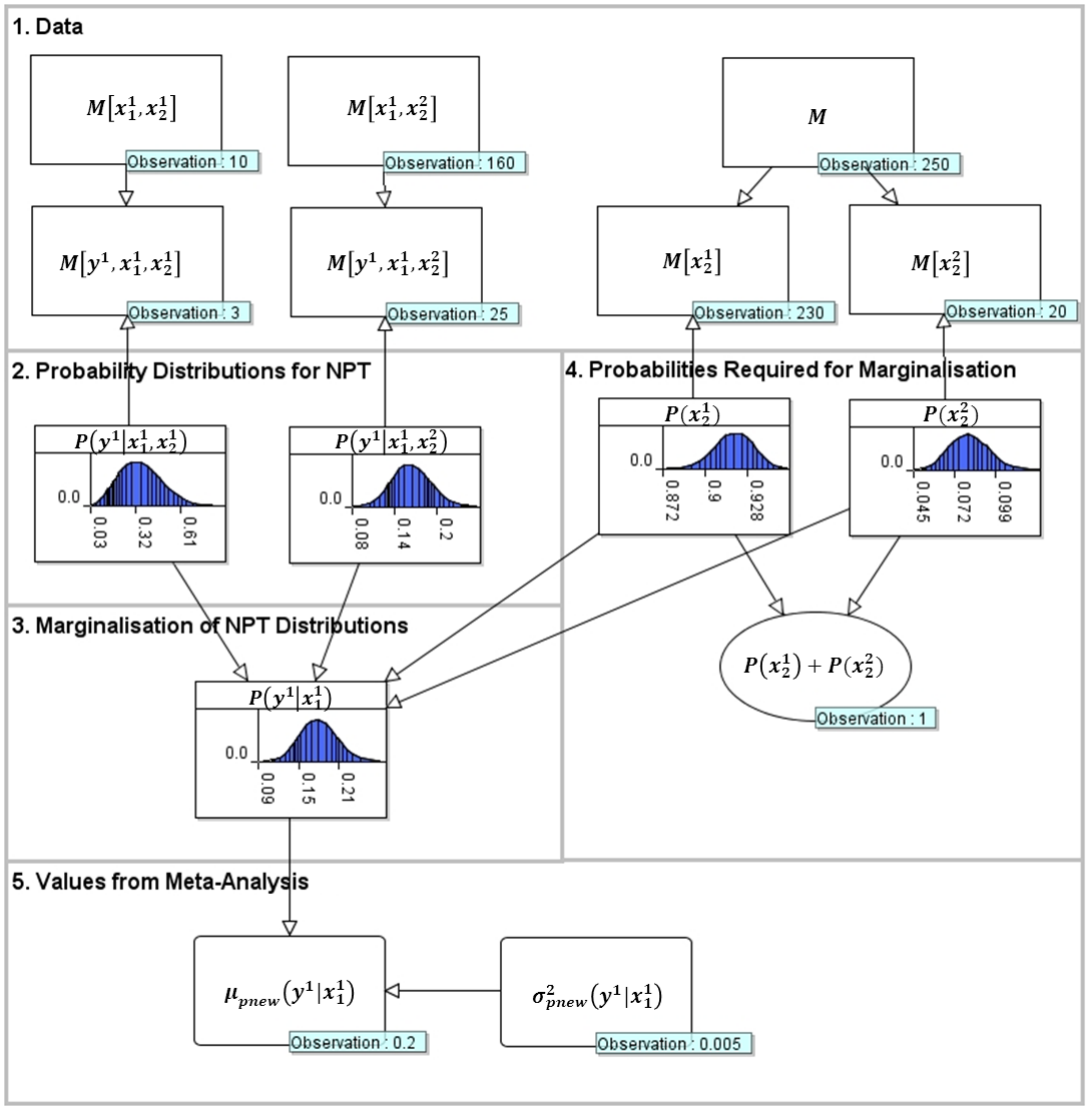


Figure 6.4 BN Representation of the Auxiliary Parameter Learning Model

2. **Probability Distributions for NPT:** This part contains the CPDs that we aim to estimate for the NPT of Y. We assign a uniform prior for these distributions:

$$P(y^1|x_1^1, x_2^1) \sim Uniform(0,1)$$

$$P(y^1|x_1^1, x_2^2) \sim Uniform(0,1)$$

3. **Marginalisation of NPT Distributions:** Since the variable Y is conditioned only on 1 variable in the meta-analysis and 2 variables in the BN, we model the probability distribution from the meta-analysis, $P(y^1|x_1^1)$, as the marginalisation of the probability distribution from the BN parameters $P(y^1|x_1^1, x_2^1)$ and $P(y^1|x_1^1, x_2^2)$:

$$P(y^1|x_1^1) = \sum_{X_2}(P(y^1|x_1^1, X_2) * P(X_2))$$

$$= P(y^1|x_1^1, x_2^1)P(x_2^1) + P(y^1|x_1^1, x_2^2)P(x_2^2)$$

4. **Probabilities Required for Marginalisation:** In order to calculate the marginalisation in part 3, we need the probability distributions of $P(x_2^1)$ and $P(x_2^2)$. We assign a uniform prior for these variables. We also assign a constraint to ensure that sum of $P(x_2^1)$ and $P(x_2^2)$ equals to 1.

$$P(x_2^1) \sim Uniform(0,1)$$

$$P(x_2^2) \sim Uniform(0,1)$$

$$\sum_{X_2} P(X_2) = P(x_2^1) + P(x_2^2) = 1$$

5. **Values from Meta-analysis:** The pooled estimate $\mu_{pnew}(y^1|x_1^1)$ from the meta-analysis is modelled with the normal distribution centred on the marginalisation shown in part 3. We use $\sigma^2_{pnew}(y^1|x_1^1)$ from the predictive distribution as the variance of this normal distribution. The normal distribution is truncated to a unit interval as it represents a probability value, denoted by $TNormal_{[0,1]}(\mu, \sigma^2)$. The values from the meta-analysis are modelled as:

$$\mu_{pnew}(y^1|x_1^1) \sim TNormal_{[0,1]}\left(P(y^1|x_1^1), \sigma^2_{pnew}(y^1|x_1^1)\right)$$

After the observations from the data and meta-analysis is entered to the BN (see Figure 6.4), the posteriors for $P(y^1|x_1^1, x_2^1)$ and $P(y^1|x_1^1, x_2^2)$ can be calculated. Note that, the NPT of Y requires point estimates for $P(y^1|x_1^1, x_2^1)$ and $P(y^1|x_1^1, x_2^2)$ whereas our model calculates the entire probability distribution of these parameters. Therefore, we take the mean of these distributions for the point estimates required for the NPT.

In the following section, we describe the generalisation of this technique for estimating parameters of variables with more parents or states.

6.2.2.2 Generalisation of the Parameter Learning Method

Let Y be a BN variable that has n parents, and $\bar{X} = \{X_1, X_2, \dots, X_n\}$ be the set of parents of Y (see Figure 6.5). Both Y and its parents have multiple states:

$$Y = \{y^1, \dots, y^k\}$$

$$X_i = \{x_i^1, \dots, x_i^k\}$$

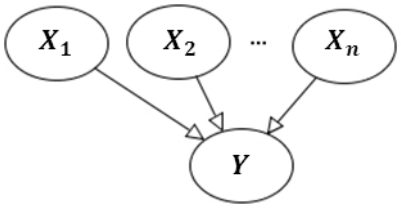


Figure 6.5 BN Model

Our dataset contains a total of M data instances about $\bar{X}$ and Y (see Table 6.5). We also have pooled conditional probability and variance estimates of the predictive distribution of $P(Y|X_i)$ from a meta-analysis (see Table 6.6). The way that these statistics are calculated is described in Section 6.1.3. The predictive distribution is a recommended way of presenting the results of a meta-analysis as it represents the uncertainty from both pooled estimate and heterogeneity (Higgins et al., 2009). However, the meta-analysis only provides us with the univariate conditional probability estimates; conditional probabilities such as $P(Y|X_1, X_2)$ are not available.

Table 6.5. Sample Learning Dataset

	Y	**X_1**	**…**	**X_n**
1	y^4	x_1^3	…	x_2^2
2	y^2	x_1^2	…	x_2^1
:	:	:		:
:	:	:		:
M	y^1	x_1^1	…	x_2^4

Figure 6.6 shows an abstract graphical illustration of the generalised auxiliary parameter learning model. This model is a generalisation of the model shown in Figure 6.4. This illustration is not a BN; it is a schema for building an auxiliary parameter learning BN for any number of states and parent variables. The size of the

auxiliary parameter learning BN grows rapidly with increasing number of parents and states.

Table 6.6 Sample Meta-analysis Results

	$\boldsymbol{\mu_{Pnew}}$	$\boldsymbol{\sigma^2_{Pnew}}$
$\boldsymbol{P(Y\|X_1)}$.	0.13	0.007
$\boldsymbol{P(Y\|X_2)}$.	0.21	0.025
:	:	:
:	:	:
$\boldsymbol{P(Y\|X_n)}$.	0.19	0.001

In Figure 6.6, the variables shown by circles are unknown variables that will be estimated by the model. The variables shown by rounded rectangles are observed with the values from the meta-analysis, and the variables shown by rectangles are observed from the dataset. The constraints that sum probabilities to 1 are not included in this figure to simplify the illustration. By running this auxiliary model, we estimate probability distributions for the parameters $P(Y|\bar{X})$ required by the NPT of Y. Since the BN requires only a point estimate of the parameter, not the whole distribution; we use the mean of this distribution as the BN parameter.

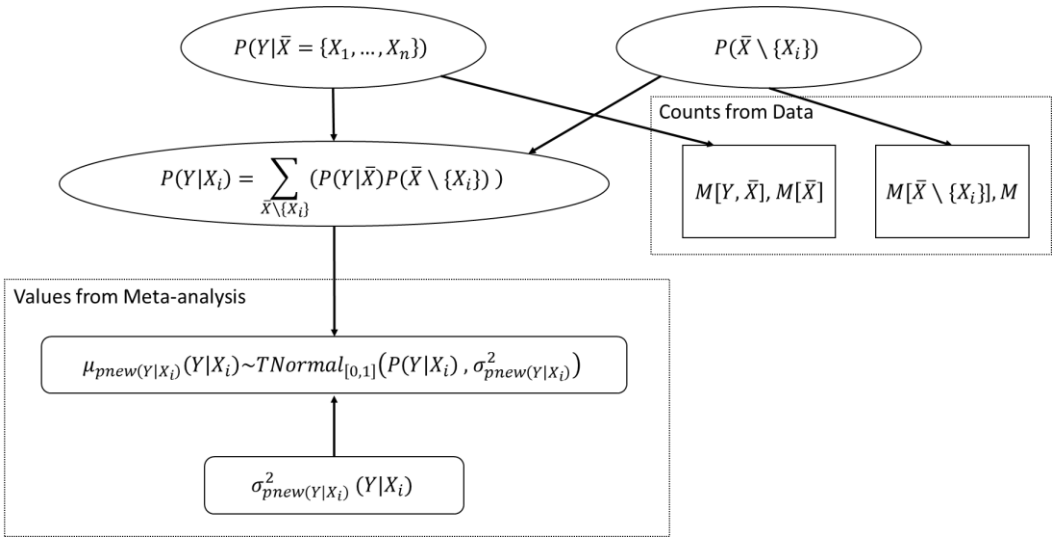


Figure 6.6 Graphical Illustration of the Generalised Auxiliary Parameter Learning Model

According to our model, the data related to Y, i.e. $M[\bar{X}, Y]$, is generated by the binomial distribution with the probability of success $P(Y|\bar{X})$ and the number of trials $M[\bar{X}]$.

$$M[Y, \bar{X}] \sim Binomial(M[\bar{X}], P(Y|\bar{X}))$$

$M[\bar{X}, Y]$ represents the count of data instances for specific values of $X_1, \dots, X_n$ and Y. For example, $M[x_1^1, x_2^3, \dots, x_n^4, y^2]$ represents the number of data instances where $X_1 = x_1^1, X_2 = x_2^3, \dots, X_n = x_n^4$ and $Y = y^2$. Similarly $M[\bar{X}]$ represent the number of data instances where $X_1, \dots, X_n$ have certain values.

Our aim is to estimate the CPD of $P(Y|\bar{X})$. We assign a uniform prior for this distribution; informative expert priors can also be used when available.

$$P(Y|\bar{X}) \sim Uniform(0,1)$$

The meta-analysis results are conditioned on a fewer variables than the CPD in the BN. Therefore, the expected values of the meta-analysis results are modelled as a marginalisation of the CPD. The meta-analysis provided the pooled conditional probability estimates about $P(Y|X_i)$ that are marginalisations of $P(Y|\bar{X})$

$$P(Y|X_i) = \sum_{\bar{X} \setminus \{X_i\}} P(Y|\bar{X}) P(\bar{X} \setminus \{X_i\})$$

$P(\bar{X} \setminus \{X_i\})$ is also estimated by the following binomial distribution.

$$M[\bar{X} \setminus \{X_i\}] \sim Binomial(M, P(\bar{X} \setminus \{X_i\}))$$

Where M denotes the total number of data instances, and $M[\bar{X} \setminus \{X_i\}]$ denotes the counts of data instances with $\bar{X} \setminus \{X_i\}$. $P(\bar{X} \setminus \{X_i\})$ has a uniform prior

$$P(\bar{X} \setminus \{X_i\}) \sim Uniform(0,1)$$

The pooled estimates from the meta-analysis $\mu_{Pnew(Y|X_i)}$ are modelled with a normal distribution that is centred on the marginalisation of the CPD. The normal distribution is truncated to a unit interval, i.e. [0 – 1], as it represents a probability. The variance of the truncated normal distribution $\sigma^2_{P(Y|X_i)}$ represents the degree of uncertainty we assign to the meta-analysis results. We enter the mean and variance of the predictive distribution in meta-analysis as observations for $\mu_{Pnew(Y|X_i)}$ and $\sigma^2_{Pnew(Y|X_i)}$. We use the truncated normal distribution as it is convenient to define the expected value and variance parameters for it but $\mu_{Pnew(Y|X_i)}$ may not be normally distributed as it represents a probability value between 0 and 1.

$$\mu_{Pnew(Y|X_i)} \sim TNorm_{[0,1]}(P(Y|X_i), \sigma^2_{Pnew(Y|X_i)})$$

Finally, we introduce constraints to ensure that probability distributions sum up to 1.

$$\sum_{Y} P(Y|\bar{X}) = 1$$

$$\sum_{\bar{X}\setminus\{X_i\}} P(\bar{X} \setminus \{X_i\}) = 1$$

$$\sum_{Y} P(Y|X_i) = 1$$

6.3 Case-study

The method described in Section 6.2 was used for developing a BN that aims to predict the short term outcomes of a traumatic lower extremity with vascular injury. The LEVT BN is designed for the limb-saving stage of the treatment: it estimates the risk of a failure of a salvage attempt, which makes amputation inevitable due to inadequate blood supply and nonviable soft tissue in the lower extremity.

The BN is built in collaboration with the Trauma Sciences Unit at the RLH and the ISR. Two trauma surgeons (ZP and NT) were involved in development of the LEVT BN. The LEVT dataset (see Section 4.3.2) and a meta-analysis of a systematic review were used for developing the LEVT BN. The remainder of this section describes the meta-analysis and development methodology of the LEVT BN.

6.3.1 Meta-analysis for Lower Extremity Vascular Trauma

Our first step was to conduct a systematic review and meta-analysis of the factors affecting outcomes of lower extremity vascular injuries. The systematic review was conducted by ZP. The studies published between 2000 and 2012 were searched using Medline, EMBASE and CINAHL databases. Another trauma registrar followed the same search procedure to evaluate the consistency of the inclusion criteria. Forty-four articles, containing information about 3054 lower extremity repairs, were

included in the systematic review. The protocol for systematic review is published in the PROSPERO register of systematic reviews (Perkins et al., 2012).

Meta-analysis of the systematic review was conducted by the author. The complete meta-analysis included pooling of conditional probabilities, odds ratios and risk differences regarding 23 variables in 44 studies reviewed in the systematic review. In this section, we use a part of the pooled conditional probabilities that are relevant to the LEVT BN as pooled conditional probabilities are naturally suited for learning BN parameters. We used the model described in Section 6.1.3 to pool the conditional probabilities and calculate the predictive distributions. The calculations were done in AgenaRisk (Agena Ltd, 2013). Table 6.7 shows the means and variances of these predictive distributions. In the following sections, we describe how these results were used for defining the BN structure and parameters.

Table 6.7 Mean and Variances of the Predictive Distributions from the Meta-analysis

Clinical Factor	**Predictive Distribution**	
	μ_{Pnew}	σ^2_{Pnew}
Arterial Repair		
Graft	0.11	0.009
Primary Repair	0.05	0.002
Anatomical Site		
Femoral	0.04	0.004
Popliteal	0.14	0.005
Tibial	0.10	0.018
Associated Injuries		
MAI* - present	0.22	0.045
MAI* - absent	0.10	0.006
Soft tissue - present	0.28	0.066
Soft tissue - absent	0.09	0.009
Fracture - present	0.14	0.013
Fracture - absent	0.02	0.001
Nerve - present	0.12	0.022
Nerve - absent	0.05	0.016
Complications		
Shock - present	0.12	0.047
Shock – absent	0.06	0.030
Ischaemia time > 6 hrs.	0.24	0.050
Ischaemia time ≤ 6 hrs.	0.05	0.009
CS+ - present	0.31	0.008
CS+ - absent	0.06	0.002

*MAI= Arterial Injuries at Multiple Levels, +CS: Compartment Syndrome

6.3.2 Deriving the BN structure

The structure of the BN was defined using the methodology described in Section 6.2.1. A domain expert (ZP) identified the variables that are found to have clinically significant effect in the meta-analysis. ZP described the mechanistic relation between each of the variables and the predicted outcome, which were modelled in a causal BN structure. Knowledge about the mechanistic relations helped us to identify the variables that are outside the scope of the BN. For example, nerve injuries were not included in the model even though it is found to increase the probability of amputation in the meta-analysis. The domain expert indicated that the outcomes related to limb function are outside the scope of the LEVT BN, and the amputations related to nerve injuries are often caused by pain and poor function outcomes.

Table 6.8 Observed and Latent Variables in LEVT BN

Observed Variables	Latent Variables
Arterial Repair	Blood Supply
Anatomical Site	Ischemic Damage
Multiple Levels (MAI)	Microcirculation
Soft Tissue Injury	Soft Tissue Cover
Associated Fracture	
Shock	
Ischemic Time	
Ischemic Degree	
Compartment Syndrome	
Repair Failure	
Number of Injured Tibials	
Nonviable Extremity	

Several latent variables were introduced while the domain expert identified the mechanistic relations between the observed clinical factors and outcomes. These variables were clinically important but neither the dataset nor the reviewed studies contained them as they cannot be directly observed. For example, both graft repairs and soft tissue injuries have higher probabilities of amputation in the meta-analysis. However, each of these factors is related to amputation through a different pathway. Graft repairs can lead to amputation when the repaired artery bleeds or gets blocked, and thus cannot deliver enough blood to the lower extremity. A variable about the degree of blood supply is required to model this relation. Although the degree of blood supply can be estimated by several measurements, the precise state of this variable is difficult to observe and therefore is not in the dataset. Soft tissue injuries

can lead to amputation if there is not enough viable soft tissue to cover the injuries and to repair the wounds. Similarly, a latent variable about the degree of soft tissue cover is required to model this relation into the BN model. Table 6.8 shows a list of the observed and latent variables in the LEVT BN structure. These variables are described in the remainder of this section.

The information in our dataset was more detailed, for some variables, compared to the information reported in the meta-analysis. For example, soft-tissue injuries were modelled with more detailed states in the BN as the dataset had more information about this variable. Similarly, information about the degree of ischemia were present in the dataset but not in the meta-analysis. Therefore, the BN contains more detail about some variables compared to the information obtained from the meta-analysis.

Model Structure

The LEVT BN is divided into 5 components, corresponding to the 5 boxes shown in Figure 6.7. The remainder of this section describes the LEVT BN by summarising the meanings of the variables and relations in each of these components:

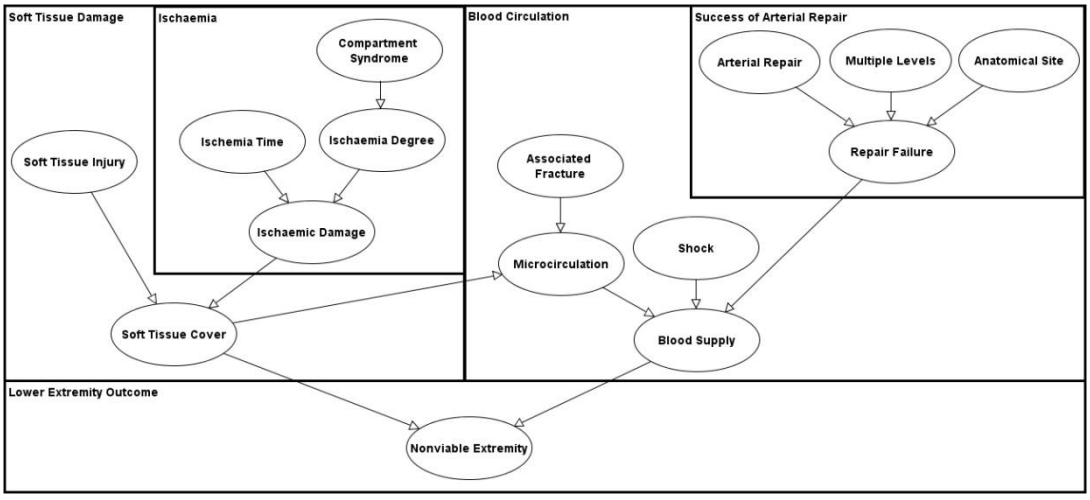


Figure 6.7 LEVT BN Structure

- **Lower Extremity Outcome:** The aim of the LEVT BN is to predict the risk of failure of an attempted salvage for a lower extremity with vascular injury. The 'Nonviable Extremity' variable represents extremities that are amputated as a result of nonviable tissue. A lower extremity can sustain life if there is adequate blood flow from the vessels and enough viable soft tissue to cover

the vessels. ‘Nonviable Extremity’ is the main outcome variable that the LEVT BN aims to predict.

- **Ischaemia:** Ischaemia is the deficiency of blood supply as a result of an arterial injury or obstruction. Ischaemia causes permanent damage to tissues if it continues for a prolonged time. Since our model is built for lower extremities with vascular injuries, most of the extremities within the scope of our model will be partly or completely ischemic until the vascular injury is repaired. The severity of ischaemic damage depends on the time elapsed since the beginning of ischemia (Ischaemic Time) and the degree of obstruction (Ischaemic Degree). A major cause of ischemia is a compartment syndrome, which causes complete obstruction of blood flow due to increased pressure in the muscle compartments of a lower extremity.
- **Soft Tissue Damage:** This part of the model predicts the projected amount of viable soft tissue cover in the lower extremity. Adequate amount of soft tissue cover is necessary to repair the tissues and protect them from infection. Therefore, soft tissue cover is one of the main factors affecting the viability outcome. Our model estimates the amount of soft tissue cover based on the amount of non-viable tissue due to the direct damage from the injury (Soft Tissue Injury) and ischemia (Ischaemic Damage).
- **Success of Arterial Repair:** This part of the model predicts the success of a vascular repair operation. ‘Arterial Repair’ variable represents the type of the repair operation, and have two states: ‘Graft’ and ‘Primary Repair’. ‘Graft’ represents bypassing of the injured artery by a vein harvested from the patient. ‘Primary repair’ represents a simpler repair operation such as stitching of a small laceration in the artery. ‘Graft’ repairs have higher rate of failure compared to ‘Primary Repair’ as this operation is more complicated and applied to more severe cases. Injury characteristics often define the type of the arterial repair. For example, an arterial injury cannot be treated by primary repair if a significant part of the artery has been torn apart and thus a graft is necessary.

  The ‘Multiple Levels’ variable represent whether vascular injuries are present at multiple levels of the same extremity. Repair of such injuries have a higher

probability of failure as clots are more likely to form when the artery is injured at multiple levels.

'Anatomical Site' variable represents the location of the main arterial injury. The injury can be at above the knee (femoral artery), at the knee (popliteal artery) or below the knee (tibial artery). Reconstruction of a femoral artery often has better outcomes compared to a popliteal or a tibial artery.

- **Blood Circulation:** 'Blood Supply' variable represents the degree of blood supply to the lower extremity. This variable essentially depends on the 'Repair Failure' variable. If the vascular repair fails, the extremity will not have adequate blood supply; so there is a deterministic relation between the negative repair failure and inadequate blood supply. However, a successful arterial repair may not guarantee adequate blood supply throughout the lower extremity; side factors including 'Shock' and 'Microcirculation' can also affect the outcomes. The 'Shock' variable represents an overall deficiency of blood supply throughout the body. 'Microcirculation' represents the severity of injury in the smaller vessels of the lower extremity.

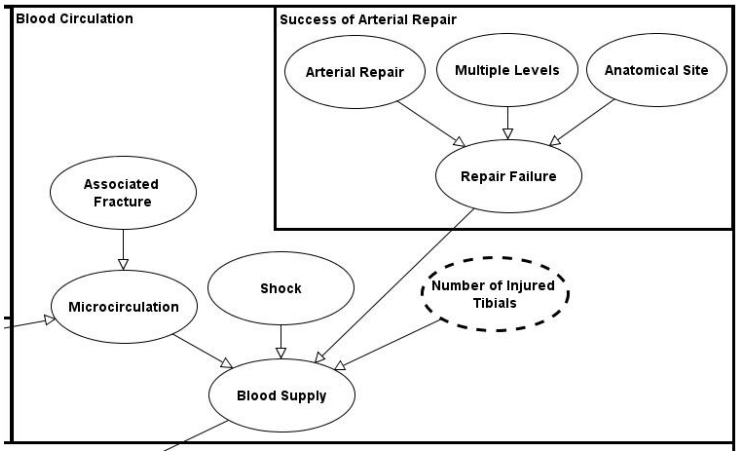


Figure 6.8 LEVT BN Modification for Below the Knee

The main branch of artery that carries blood to the lower extremity divides into three branches below the knee. In other words, a single main branch of artery supplies blood for the tissues above the knee whereas three branches, called tibial arteries, supply blood for the tissues below the knee. In order to model this difference, we modified the BN structure for injuries below the knee by adding a variable about the number of injured tibial arteries. This

modification is shown by the variable with dashed lines in Figure 6.8. Modelling of tibial arteries is important since the extremity is more likely to survive if all of the tibial arteries are not injured, even when the arterial repair fails. Apart from this difference, the BN models for above the knee and below the knee injuries are exactly the same. Table 6.9 shows the description and states of each variable in the LEVT BN.

Table 6.9 Description and States of Variables in LLVI BN

Variable	Description	States
Anatomical Site	Level of arterial injury	{Femoral, Popliteal, Tibial}
Arterial Repair	Surgical method for treating arterial injury	{Primary, Graft, Ligation}
Associated Fracture	Associated fracture at the level of arterial injury	{True, False}
Blood Supply	Predicted degree of blood supply after repair	{Low, Medium, High}
Ischaemic Damage	Degree of soft tissue damage caused by ischaemia	{Low, Medium, High}
Ischaemic Degree	Degree of obstruction in blood supply	{None, Partial, Complete}
Ischaemic Time	Duration of obstructed blood supply	{<1hr, <3hr, <6hr, ≥6hr}
Microcirculation	Degree of microcirculation problems at the level of arterial injury	{Normal, Severe, Deranged}
Multiple Levels	Presence of arterial injuries at multiple levels	{True, False}
Nonviable Extremity	Presence of a non-survivable lower extremity	{True, False}
Number of Injured Tibials	Number of tibial arteries injured	{0, 1, 2, 3}
Shock	Presence of uncompensated shock	{True, False}
Soft Tissue Cover	Degree of soft tissue damage due to injury and ischaemia	{Low, Medium, High}
Soft Tissue Injury	Degree of soft tissue damage cause by injury	{None, Mild, Moderate, Severe, Profound}
Repair Failure	Failure of arterial repair due to occlusion or bleeding	{True, False}

6.3.3 Learning Parameters

We examined the amount of data available for learning each parameter in the BN. Data were too small, or not available, to learn the parameters for some relations. In this section, we describe the techniques used for learning the parameters when they have 1) small amount of data, 2) no data at all (latent variables) and 3) adequate amount of data.

Variables with Small Data

We learned the parameters that had less than 20 data instances from a combination of meta-analysis and data using the technique described in Section 6.2.2. Table 6.10 shows the data available for learning the NPT of the 'repair failure' variable. There was small amount of data to learn some of these parameters (shown by bold in Table 6.10) therefore we learned these parameters by combining the meta-analysis results with the data. We used the mean and variance of the predictive distributions from the meta-analysis (see Table 6.7) as observations to μ_{Pnew} and σ^2_{Pnew} variables in the Bayesian learning model (see Figure 6.6).

The meta-analysis provides us the pooled probabilities of an unsuccessful outcome conditioned on each individual clinical factor (see Table 6.7). The variable equivalent to an unsuccessful outcome is 'nonviable extremity' in the LEVT BN. However, the meta-analysis results could also be used for the NPT of the 'repair failure' variable as 1) our model assumes a deterministic relation between an unsuccessful outcome and repair failure 2) the parents of 'repair failure' can influence 'nonviable extremity' through only one pathway. For example, the 'arterial repair' variable can affect 'nonviable extremity' through the following pathway in our model:

$$\underset{(AR)}{Arterial\ Repair} \rightarrow \underset{(RF)}{Repair\ Failure} \rightarrow \underset{(BS)}{Blood\ Supply} \rightarrow \underset{(NE)}{Nonviable\ Extremity}$$

The probability provided by the meta-analysis, $P(NE = True|AR = Graft)$, is equivalent to marginalisation of 'repair failure' and 'blood supply' from this pathway:

$$P(NE = True|AR = Graft) = \sum_{BS,RF} P(NE = True|BS)P(BS|RF)P(RF|AR = Graft)$$

In our model, a repair failure always leads to inadequate blood supply, and inadequate blood supply always leads to a nonviable extremity:

$$P(BS = Low|RF = True) = 1, P(NE = True|BS = Low) = 1$$
$$P(BS = Low|RF = \neg True) = 0, P(NE = True|BS = \neg Low) = 0$$

By using these values in the marginalisation equation above, we get: $P(NE = True|AR = Graft) = P(RF = True|AR = Graft)$. Consequently, we can use the probabilities from the meta-analysis for learning the relation between 'repair failure' and its parents.

The meta-analysis does not provide any information about the multiplicity of tibial arteries (see Section 6.3.2 how the multiplicity of tibial arteries is modelled in the BN). Therefore, we model the number of injured tibial arteries as a risk modifier. Our model assumes that a repair failure leads to a non-viable extremity if all 3 tibial arteries are injured. However, there is a chance of a successful outcome, which is learned from data, if only 1 or 2 tibial arteries are injured.

Table 6.10 Amount of Data Available for Learning Parameters of Repair Failure Variable

AR* MAI* AS*	Graft True Femoral	Graft True Popliteal	Graft True Tibial	Graft False Femoral	Graft False Popliteal	Graft False Tibial	Primary True Femoral	Primary True Popliteal	…
RF* Data	**14**	**6**	**2**	71	115	38	**1**	**3**	…

**AR: Arterial Repair, MAI: Multiple Levels, AS: Anatomical Site, RF: Repair Failure*

We used the OpenBUGS software (Lunn et al., 2009) to calculate the posteriors of the auxiliary learning model for this case study. Since OpenBUGS uses MCMC sampling, it is necessary to assess the convergence of the Markov chain to ensure that sampled values cover the entire distribution. We used Gelman and Rubin (1992) diagnostic technique, sample plots and autocorrelation plots to assess the convergence. None of the diagnostic techniques can prove convergence but they can assist detecting the lack of convergence. We discarded the first 10,000 samples in

MCMC as the burn-in samples, and calculated the posterior distributions based on the next 70,000 samples.

Table 6.11 shows extracts of the NPTs of the 'Repair Failure' and the amount of data available for learning its parameters. The values in bold and italic fonts are the parameters learned by combining the results of the meta-analysis with the data, and the values in normal fonts are the parameters learned purely from the data. The parameters learned from two approaches differ substantially for smaller amounts of data. The effects of this difference to the model performance are discussed in Section 6.4.1.

Table 6.11 Learning from Data, and from Combining Data and Meta-Analysis Results

AR*:	Graft		Graft		Graft		…	Primary		Primary		…
MAI*:	True		True		True			True		True		
AS*:	Femoral		Popliteal		Tibial			Femoral		Popliteal		
RF*:							….					
True	0.17	***0.17***	0.39	***0.41***	0.99	***0.49***		0.01	***0.14***	0.01	***0.29***	...
False	0.83	***0.83***	0.61	***0.59***	0.01	***0.51***		0.99	***0.86***	0.99	***0.71***	
Data:	14		6		2			1		3		

**AR: Arterial Repair, MAI: Multiple Levels, AS: Anatomical Site, RF: Repair Failure*

Latent Variables

The BN contained several latent variables as described in Section 6.3.2 (see Table 6.8 for a list of these variables). Ranked nodes were used to model the NPT of these variables (Fenton et al., 2007). A ranked node is an approximation of the truncated normal distribution to the multinomial distribution with ordinal scale (see Section 2.6.1 for a more detailed description of ranked nodes). We used the framework proposed by Fenton et al. (2007) to elicit the parameters of ranked nodes.

For each of the latent variables we first asked the domain experts to describe the relation between the variable and its parents. Afterwards, we selected a suitable ranked node function and elicited initial weights that imitate the described relation. We presented the behaviour of the ranked node under various combinations of observations to the domain experts, and refined the weights based on their comments.

Adequate Amount of Data

After the parameters with small or no data were defined, the remainder of the parameters were learned purely from the data. The EM algorithm was used to learn the parameters as the dataset contained missing values. The parameters that were already defined from the experts or meta-analysis were kept fixed while EM was applied.

6.4 Results

The performance of the LEVT BN for predicting the 'Nonviable Extremity' variable was tested by a 10-fold cross-validation in the LEVT dataset. The AUROC was 0.90. The BN had 80% specificity when operated at 80% sensitivity, and 70% specificity when operated at 90% sensitivity. The LEVT BN had a BS of 0.06 and a BSS of 0.33. Hosmer-lemeshow test was used to assess the calibration of the BN. The BN was well calibrated with a Hosmer-lemeshow statistic of 12.7 (p-value: 0.13, see Figure 6.9).

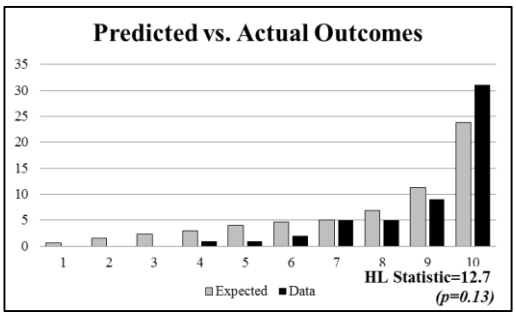


Figure 6.9 Calibration of the LEVT BN

6.4.1 Parameters from Pure Data vs. Hybrid Approach

The LEVT BN was learned by a hybrid approach that reinforces the data with the meta-analysis results when there is small amount of data to learn a parameter (see Section 6.3). We compared the results of this approach to a purely data-based parameter learning algorithm. We learned the parameters of the same BN structure purely from data and compared it with the LEVT BN learned by the hybrid approach. Note that, the parameters of the ranked nodes, which were elicited from

the experts, were kept the same in both models. The purely data-based parameter learning had poor results in all measures. The AUROC was 0.48, the specificity was 13% at 80% sensitivity, and the Hosmer-lemeshow test indicated poor calibration (p-value: 0.01). Both BS and BSS indicated poor performance for the purely data-based parameter learning as well (see Table 6.12).

The LEVT BN has a quite complicated structure compared to the data available for learning some of its parameters. As a result, the purely data-based approach overfits the data, which leads to poor results in the cross-validation.

Table 6.12 Results of Parameter Learning from Data and Hybrid Approach

	Hybrid Approach	**Data**
AUROC	0.90	0.48
Specificity (at 90% Sensitivity)	70%	7%
Specificity (at 80% Sensitivity)	80%	13%
Hosmer-Lemeshow Test	12.7 (p=0.13)	20 (p=0.01)
Brier Score	0.06	0.10
Brier Skill Score	0.33	0.02

6.4.2 Mangled Extremity Severity Score

MESS (Johansen et al., 1990) is a well-known scoring system developed for providing decision support in treatment of mangled extremities. MESS calculates a score based on the injury mechanism, the degree of shock, the ischemic status and the patient's age. If the score is above a certain threshold value MESS recommends an amputation (see Section 4.2.2 for the results of MESS in different validation studies). In the LEVT dataset, MESS had an AUROC of 0.75 for predicting the 'Nonviable Extremity' variable (Figure 6.10). Its specificity was 40% when operated at 90% sensitivity and 60% when operated at 80% sensitivity. Both the accuracy and calibration of MESS were worse than the predictions of the LEVT BN as shown in Table 6.13. BS and BSS could not be calculated for MESS since the outputs of MESS are not probabilities.

Table 6.13 Results of LEVT BN and MESS

	LEVT BN	MESS
AUROC	0.90	0.75
Specificity (at 90% Sensitivity)	70%	40%
Specificity (at 80% Sensitivity)	80%	60%
Hosmer-Lemeshow Test	12.7 (p=0.13)	20 (p=0.01)

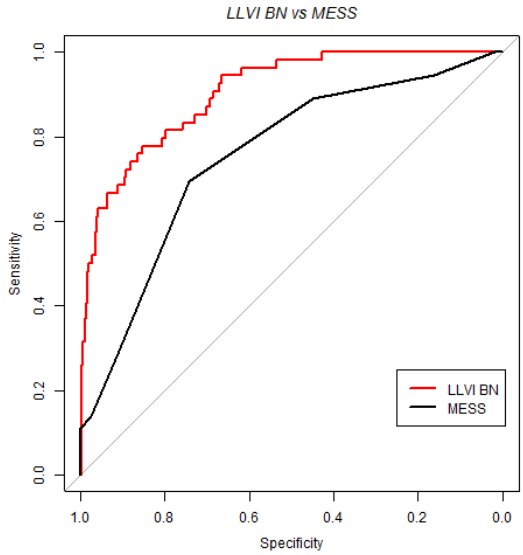


Figure 6.10 ROC Curves for LEVT BN and MESS

6.4.3 Learning BN Purely From Data vs. LEVT BN

Clinical knowledge was an essential factor in development of the LEVT BN. In order to assess the effects of using knowledge in modelling, we compared the performance of the LEVT BN to 3 other BNs that were developed purely from the data using 3 different structure learning algorithms. We used a score based learning algorithm (hill climbing (HC) algorithm with BIC score (Korb and Nicholson, 2004b; Margaritis, 2003; Schwarz, 1978)), a constraint based learning algorithm (grow shrink (GS) algorithm (Margaritis, 2003)), and a hybrid algorithm that combines the score and constraint based approaches (max-min hill climbing (MMHC) algorithm (Tsamardinos et al., 2006)) to learn each of these BNs. Expert knowledge was not used in development of these BNs, and the BNs do not contain any latent variables.

Table 6.14 Results of LEVT BN and Structure Learning Methods

	LEVT BN	HC	MMHC	GS
AUROC	0.90	0.83	0.83	0.84
Specificity (at 90% Sensitivity)	70%	37%	41%	54%
Specificity (at 80% Sensitivity)	80%	71%	80%	69%
Hosmer-Lemeshow Test	12.7 (p=0.13)	11.5 (p=0.17)	15.3 (p=0.05)	8.5 (p=0.39)
Brier Score	0.06	0.07	0.07	0.07
Brier Skill Score	0.33	0.22	0.21	0.23

We first imputed the missing values in the dataset using the Amelia package (Honaker et al., 2013) in the R statistical software (R Core Development Team, 2013) as all of the structure learning algorithms require complete datasets. Afterwards, we learned a separate BN structure using each of the learning algorithms. These algorithms are readily implemented in the BNLearn package (Scutari, 2010) of the R statistical software.

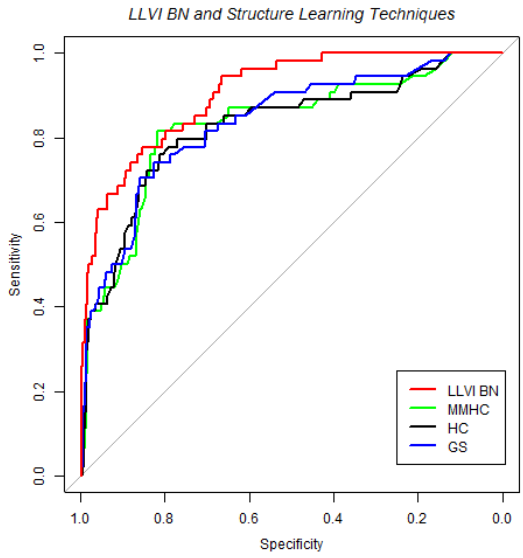


Figure 6.11 ROC Curves for LEVT BN and Structure Learning Methods

We assessed the performance of each BN for predicting the 'Nonviable Extremity' variable in a 10-fold cross validation. The LEVT BN has a larger AUROC than the BNs learned purely from the data (see Figure 6.11). The LEVT BN had substantially better performance at operating points with higher sensitivity levels. At the 90% sensitivity level, the predictions of the LEVT BN have over 70% specificity which is, on average, 27% more than the BNs learned from the data. Both BS and BSS also

indicated better performance for the LEVT BN (see Table 6.14). The Hosmer-lemeshow test indicated that the LEVT BN, HC and GS were calibrated at 95% confidence level.

Figure 6.12 shows the BN structures learned by MC (Figure 6.12a), MMHC (Figure 6.12b) and GS (Figure 6.12c) methods. In Section 6.4.1, we observed that a purely data-driven approach overfits the data for the LEVT BN as the BN structure is too complicated for the available data. In order to avoid overfitting, the structure learning methods learned simpler BN structures. Consequently, their predictions were better than the purely data-based parameter learning algorithm in Section 6.4.1.

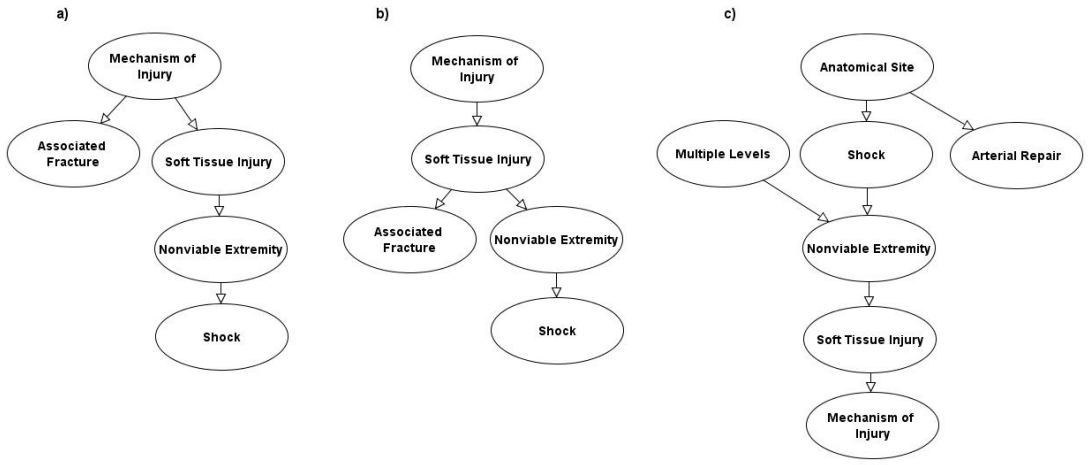


Figure 6.12 BN Structures Learned by a) HC b) MMHC c) GS methods

One of the disadvantages of a purely data-driven structure learning method is that the learnt structure is often not meaningful to domain experts (see Section 3.3). In this case study, the structure learning methods built different, and sometimes contradicting, BN structures. For example, while the 'Shock' variable is a cause of the 'Nonviable Extremity' variable in the GS model (see Figure 6.12c); the same variable is the consequence of the 'Nonviable extremity' variable in the HC and MMHC models (see Figure 6.12a and Figure 6.12b).

The potential use of the model for evidence based medicine may be limited if the BN structure is not meaningful to experts. First of all, the BN cannot be supported by the evidence from the clinical literature if its structure does not make sense to domain experts (see Chapter 8). The relations defined in the BN must be consistent with clinical knowledge in order to identify the relevant evidence for them.

The predictions of a BN can also be explained if its structure is consistent with clinical knowledge. If causal pathways in the BN are aligned with clinical knowledge, its predictions can be explained in a way that makes sense to domain experts. However, it is not possible to explain the predictions of some machine learned BN apart from saying that the input variables are correlated with the outcome.

6.5 Conclusion

This chapter presented a methodology to build BNs for decision supports based on clinical evidence from meta-analyses, expert knowledge and data. Meta-analysis results were used to identify the BN variables based on the evidence from previous research. We proposed a Bayesian learning technique that can combine the meta-analysis results with data to learn the BN parameters. Our method was successfully applied to the trauma case-study, in which we developed an accurate model for predicting short-term outcomes of lower extremities with vascular injuries. The techniques presented in this paper can be applied to a wider scope of problems than trauma care. A meta-analysis is an important source of evidence but it is often used to analyse simple relations, conditioned on a few variables, therefore it does not account for the complexity of disease mechanisms. By combining evidence from the meta-analysis with data and knowledge, our method enables evidence to be used in more complicated decisions by taking causal pathways and interaction effects into account.

In our case study, we developed an accurate model for a clinical problem where the previous modes have not been successful. The case study demonstrated the benefits of integrating meta-analysis results and expert knowledge into BN development. The BN built by our approach performed better than the structure learning techniques and the scoring system compared: the AUROC was 0.90 for our BN, 0.84 for the best performing structure learning method and 0.75 for the MESS scoring system.

The LEVT BN contained latent variables that were modelled by ranked nodes. The parameters of these variables were elicited from domain experts. It would also be possible to estimate the parameters of these variables using the methodology

presented in Chapter 5. In this case, each step of the EM algorithm would estimate the parameters of a ranked node rather than a normal NPT. The EM algorithm has not been implemented to learn ranked nodes. Such implementation would be useful for domains with limited data as ranked nodes require fewer parameters than complete NPTs.

As further research, the parameter learning method described in this Chapter could be expanded with qualitative expert constraints (see Section 2.6.3 for a review of the Bayesian parameter learning methods with qualitative expert constraints). This expansion would allow expert knowledge to be integrated with meta-analysis results and data when learning parameters. In Bayesian parameter learning methods, such as the one we described in Section 6.2.2, we estimate the entire distribution of the parameters. This gives us both expected value of the parameter and a variance showing the degree of uncertainty for this estimate. Usually, the expected value of this distribution is used as the BN parameter, and the variance is ignored. However, if the BN structure is meaningful and aligned with knowledge, the variance could be used to show our degree of understanding about the parameters. Ways of integrating this variance to the parameter estimation techniques could be investigated.

Chapter 7

Abstractions in Bayesian Networks

The graphical structure of a BN makes it a technology well-suited for developing evidence-based decision support models from a combination of domain knowledge and data. However, the available data seldom match the variables in the structure that is elicited from experts, whose models may be quite detailed; consequently, the structure needs to be abstracted to simplify parameter estimation. Up to now, this abstraction has been informal, loosening the link between the final model and the experts' knowledge. In this chapter, we propose a method for abstracting the BN structure by using four 'abstraction' operations: node removal, node merging, state-space collapsing and edge removal. Approximations introduced by the abstraction operations can be identified from changes in the conditional independence (CI) assertions of a BN.

7.1 Introduction

A knowledge-based BN aims to model the data-generating process of a problem domain by encoding knowledge about influences and independences between the important variables of the domain. They are often developed through multiple stages as the knowledge engineers and the domain experts refine the model iteratively (Laskey and Mahoney, 2000). The initial knowledge model is often large and detailed, and some elements of the model may need to be simplified or *abstracted* as data is lacking or the parameters are too difficult to elicit. However, even simple abstraction operations, such as removing a node, can result in numerous and

complicated alternative BNs which are difficult for the knowledge engineers to evaluate without a structured method. The effects of these abstractions must be carefully examined by the domain experts to prevent any unwanted changes in the modelled knowledge of the data generating process. Moreover, the way that the final BN has been derived needs to be presented thoroughly so that the knowledge base of the model and its derivation is understandable.

Our aim is to present a method of abstracting a BN structure. The method is developed for knowledge engineers building a BN structure with domain experts. The method provides a set of abstraction operations which together:

1. Allow a BN to be simplified by removing and merging nodes, removing edges and reducing the number of states

2. Distinguish abstractions that add to the knowledge base from those compatible with the knowledge elicited so far, so that the added knowledge can be confirmed by domain experts

3. Provide a way to show the link between the initial and abstracted models, in the form of a derivation that captures the complete sequence of abstraction operations

The method can be used to help knowledge engineers select the most suitable model refinements by evaluating alternative abstractions, in consultation with domain experts. The selection may also be guided by considering the availability of data or compatibility with causal relationships.

Our knowledge engineering method is based on well-known techniques mainly used for learning and inference problems (Choi and Darwiche, 2006; Shachter, 1986; Van Engelen, 1997; Wellman and Liu, 1994). Our main contribution is to explore the knowledge engineering aspect of these operations.

The remainder of this chapter is organised as follows: Section 7.2 gives an overview of the relation between knowledge and conditional independencies (CI) in BNs. Section 7.3 introduces abstraction as a knowledge engineering method and Section 7.4 describes the abstraction operations of this method and examines their

compatibility properties. These operations are illustrated by a medical case-study in Section 7.5. Section 7.6 shows the graphical representation of the abstraction operations. Finally, Section 7.7 discusses the motivation from ABEL (Patil et al., 1981), and Section 7.8 presents the conclusions.

7.2 Knowledge and Conditional Independencies

The aim of a knowledge-based BN is to model the data-generating process of a domain by encoding knowledge about influences and independences in the BN structure. A satisfactory modelling of this knowledge is when a BN G_X is able to represent the joint probability distribution P_X of the data-generating process (see Figure 7.1).

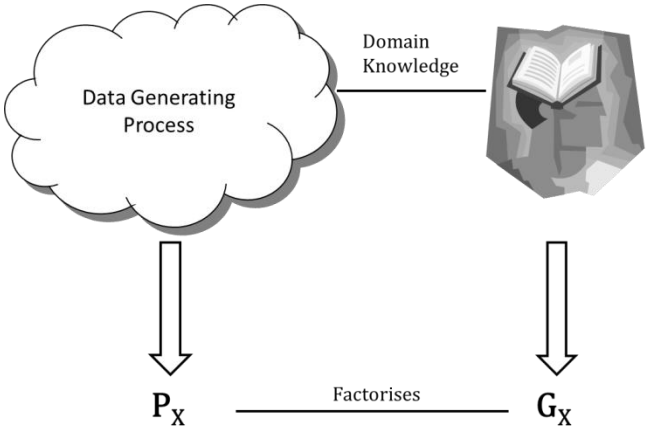


Figure 7.1 Knowledge-based Bayesian Network

A BN structure encodes a set of CIs; therefore its ability to represent the data-generating process depends on the CIs the BN asserts. Consequently, compatibility of the BN abstractions can be evaluated by their effect on the CI assertions of the BN. The preliminaries regarding CIs and BNs are provided in Section 2.4.

7.3 Abstraction as a Knowledge Engineering Method

In this section, we present an overview of our abstraction method for knowledge-engineered BNs. Our approach is to construct the BN structure based on expert knowledge before using data, statistics from relevant studies, and numbers elicited

from domain experts to parameterise the BN. The first step of our method is to elicit the BN structure about the domain. The initial structure should be considered as a knowledge-model of the domain so it should include all relevant variables and relations without being limited by issues such as availability of data or complexity of the model. Knowledge engineering techniques for eliciting the BN structure are discussed in Section 2.6.1.

When the initial structure of the BN is complete, we compare it with the resources available for parameterising the BN. The data and the statistics from relevant studies may not be enough to learn the parameters for all of the variables in the BN. For example, information about some of the variables may not be observed in practice, or they may not be recorded in the data. Moreover, the NPTs of some variables may be too large to elicit from the domain experts or to learn from the data. Consequently, the initial BN structure has to be abstracted in order to have a parameterised and working BN.

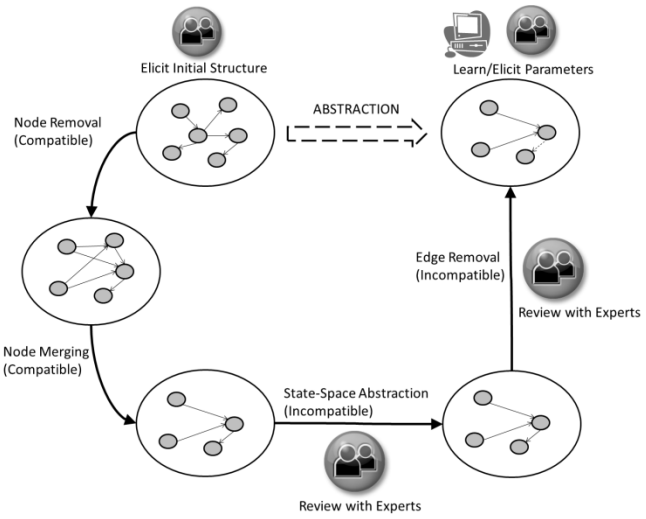


Figure 7.2 Overview of abstraction as a method of model development

We propose 4 abstraction operations that are classified as compatible or incompatible as shown in Table 7.1. Node removal and node merging are 'compatible' abstractions meaning that these operations can always be applied without adding new CI conditions to the variables shared between the initial and the abstracted BN. State-space abstraction and edge removal are 'incompatible' abstractions that bring additional CI conditions to the BN structure. The probability distribution of the data-generating process that factorises on the initial BN structure

may not factorise after incompatible abstractions are applied. Therefore, the domain experts should review the effects of the incompatible abstractions on the knowledge-base of the BN. The compatibility properties of the abstraction operations are defined in Section 7.4.1. Figure 7.2 illustrates the overview of our method. In summary:

- We use compatible abstraction first to reduce the number of variables in the model
- We then make approximations with incompatible abstractions to improve learning and elicitation, often to complement the compatible abstractions

However, the order of the abstraction operations is not restricted to this pattern; any operation can be applied to any network.

Table 7.1 Abstraction Operations

Operation	Compatibility
Node removal	Compatible
Node merging	Compatible
State-space abstraction	Incompatible
Edge removal	Incompatible

The purpose of the BN must be considered when removing variables: some variables may have primary importance, and they should not be removed even when no data is available; whereas other variables may be removed without affecting the reasoning mechanism of the model significantly. The node removal operation is described in Section 7.4.2.

Node merging can be used for a set of variables that together represent an abstract concept. Each of the merged variables is described as a part of the definition of the abstract concept. The node merging operation is demonstrated in Section 7.4.3.

The states of a variable can be collapsed in order to decrease the size of its NPT thereby making it easier to learn or elicit the parameters with available resources. Domain experts should review the CI conditions that are added as a result of state-space abstraction. The state-space abstraction operation is described in Section 7.4.4.

Edge removal can also be used to simplify the parameter-space of a BN. Edge removal is an incompatible abstraction; therefore, its effects on the CI conditions should be reviewed by domain experts. Data, if available, can be used to assess the effects of edge removals by statistical independence tests or model selection scores. The edge removal operation and approaches for reviewing its effects are described in Section 7.4.5.

7.4 Abstraction Operations

7.4.1 Compatibility of Abstractions

A BN structure has fewer variables, states or relations than the initial – knowledge – model after the abstraction operations are applied. A crucial factor to consider is whether the abstracted BN G_A is a compatible abstraction that is able to represent the probability distribution of the remaining variables. In this section, we present the definition of compatible abstractions in the case where some variables are removed from the model[1]. In the following sections we will expand this definition for the node merging operation, and discuss the compatibility of the state-space abstraction and edge removal operations.

A compatible node removal operation is able to represent the probability distribution of the remaining variables in the BN. Let G_X be a BN structure where the probability distribution P_X of the set of variables X factorises. When a set of nodes are removed from this BN, the abstracted BN structure G_A must be able to represent the probability distribution of the set of remaining variables A. Since $A \subseteq X$, the probability distribution P_A of the set of variables A is simply a marginalisation of P_X, and the compatible abstraction G_A is able to factorise $P_A = \sum_{X-A} P_X$. This is possible if the abstraction operation does not introduce additional independence assertions to the BN structure so that G_A asserts the subset of the d-separations in G_X.

[1] This section uses the concepts of d-separation, active trails and CI to define the compatibility of abstractions, see Section 2.4 for the definition of these concepts.

Compatible Abstraction: G_X is an compatible abstraction of G_A if $dsep_{G_A}(D; E|F) \rightarrow dsep_{G_X}(D; E|F)$ and $\neg dsep_{G_X}(D; E|F) \rightarrow \neg dsep_{G_A}(D; E|F)$ where D, E, F three sets of variables in A.

When an abstraction operation adds a CI assertion, therefore is incompatible, the knowledge engineers must carefully evaluate the differences brought by the additional CI comparing it to the initial knowledge-based model (see Figure 7.3).

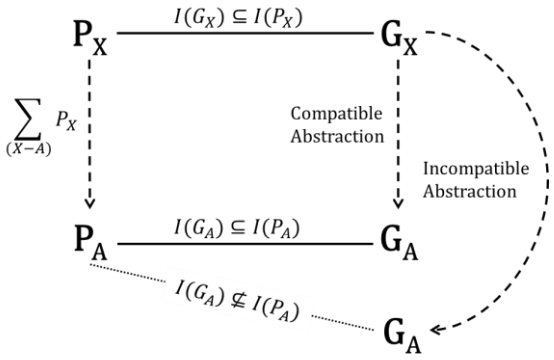


Figure 7.3 Compatible and Incompatible Abstraction

In the remainder of this section, we present each of the abstraction operations and discuss their compatibility properties.

7.4.2 Node Removal

Any node in a BN can be removed without adding independence assertions using Shachter's topological operations (Shachter, 1988, 1986). The node removal operation is based on the concepts of barren nodes, covered edges and edge reversals:

1. **Barren Nodes:** Nodes that do not have any descendants in the BN are called barren nodes. Removing barren nodes does not add independence assumptions for the rest of the BN: $dsep_{G_A}(D; E|F) = dsep_{G_X}(D; E|F)$ where $(X - A)$ are barren nodes and D, E, F are three sets of variables in A.

2. **Covered Edge and Edge Reversal:** An edge $Y \rightarrow Z$ in a graph G is *covered* if $PA_Y^G = PA_Z^G - \{Y\}$ that is if the set of parents of Y and set of parents of Z excluding Y are equivalent. Covered edges can be reversed without adding

any independence assertions to the BN structure. This is useful since the outgoing edges from a node can be reversed to make the node barren, and to remove the node without adding CIs, when these edges are covered.

3. **Adding Edges:** Adding edges to a BN structure do not add CI assertions to G_X, but they remove CI assertions, since all the previously active trails are still present and the added edges may add further active trails: $I(G_E) \subseteq I(G_X)$ where G_E is derived from adding some edges to G_X. As a result, any edge in a BN can be covered, without introducing CIs, by adding edges between the variables that share the edge and their parents.

In summary, we can cover any edge in a BN by adding more edges to the BN structure, and this will not add more independence assertions. We can reverse any edge and make any node barren as we can cover any edge in the BN structure. Consequently, we can remove any node as a compatible abstraction without adding independence assertions to the BN (Shachter, 1988).

Removing a node can increase the parameter-space and computational complexity of a BN especially if the removed node has many children. As the number of children of a variable increases, more edges are required to cover the outgoing edges from the variable. For example, we need to add 4 edges in order to remove the node R from the BN structure in Figure 7.4a with a compatible abstraction. As a result, the number of parents that C, D and E have is tripled (Figure 7.4c).

Suppose that we need to reverse an edge $A \rightarrow B$ so that A becomes a barren node. If another directed path exists from $A \rightarrow \cdots \rightarrow B$ then reversing the edge $A \rightarrow B$ would introduce a cycle. However, in this case the node B has another parent 'X' on this path; since X is a descendant of A and the BN is acyclic we can be sure that there is no directed path from X to B via A. In general, since the BN is acyclic, the set PA_B^G must contain at least one node that does not have an additional directed path leading to B. This node is not greater than the other parents of B in the partial order on edges defined by the BN's graph. It is therefore always possible to select a node from PA_B^G so that its edge to B can be reversed without introducing a cycle and is therefore possible to make any node in a BN barren without introducing a cycle.

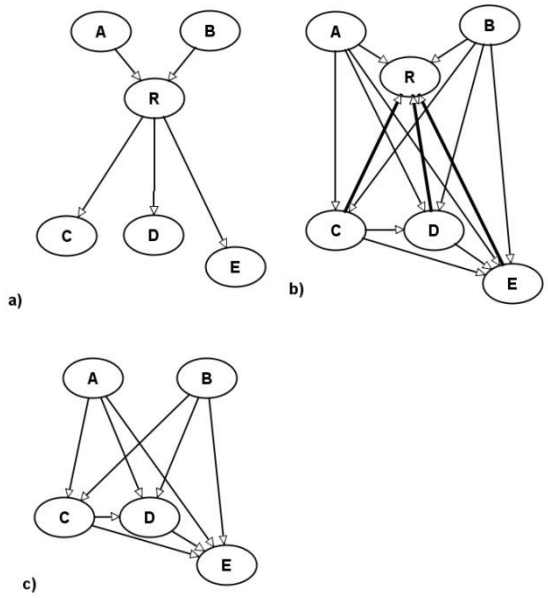


Figure 7.4 (a) R with multiple children (b) Making R a barren node (c) R removed

Since the BN's graph defines only a partial, not total order, on nodes, we may have to choose the order in which the edges leading from a node R are reversed as we transform the BN to make R a barren node. Changing the order of edge reversal can change the structure of the final BN. For example, we have to reverse $R \rightarrow D$ and $R \rightarrow C$ in order to remove R from the BN shown in Figure 7.5a. If we start by reversing $R \rightarrow D$ the resulting structure has a total of 6 edges as shown in Figure 7.5b. However, if we start by reversing $R \rightarrow C$, the resulting structure has 7 edges as shown in Figure 7.5c.

The order of edge reversals can be selected in a way to minimise either the number of edges added or the size of the final state space. Having the least number of additional edges does not guarantee that the state-space is minimised. A BN with fewer edges can have a larger state-space due to the state-space of its individual variables.

Making an exhaustive search for possible equivalent abstractions gets increasingly difficult as the number of children of the removed node increases. The outgoing edges from a node that has n children can be reversed in $n!$ possible orders. However, an exhaustive search can be unnecessary from a knowledge engineering perspective. A relation in a BN can represent causality or association. Edges added between the children of a removed node are assumed to represent association due to

the missing parent, therefore they can be modelled in any direction. However, some of the directions can make more sense to the domain experts even though these variables are assumed to be independent when the state of missing parent is known.

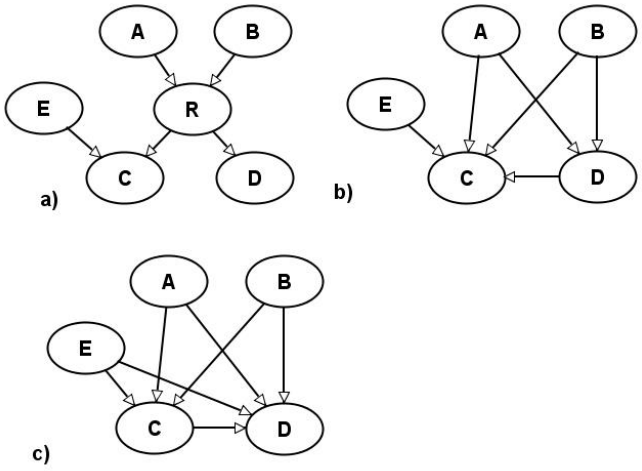


Figure 7.5 (a) Initial BN (b) Equivalent Abstraction (c) Equivalent Abstraction

If the domain experts have no preference about the directions of edges, following heuristics can be useful to choose the order of removing nodes without making an exhaustive search:

- In order to add the minimal number of edges, start reversing from the edge $R \rightarrow X$ where X has the least number of parents. When multiple edges must be reversed to remove R, an edge is added from X and $PA_X^G - \{R\}$ in order to cover edges directed to other children of R that are reversed later. Therefore, if we start from reversing the node that has the smallest number of parents, fewer edges will be added to cover the edges that are reversed later.

- Similarly, in order to have a minimal increase in the size of the state-space, start reversing from the edge $R \rightarrow X$ where X has the smallest number of parameters.

Node removal is a well-known technique that has been primarily used for inference problems (Shachter, 1990, 1988, 1986). It is a crucial operation for knowledge engineering of BNs as it explicitly shows the number and possible direction of edges that must be added for abstraction. A node removal operation is straightforward when the removed node has a single or no child. In this case, no edges are added and

the edge directions are maintained. However, removing a node with multiple children can result in multiple equivalent BN structures with many additional edges as shown in Figure 7.5. The additional edges are necessary for compatibility but, if further simplifications are required, incompatible abstractions may follow.

Using Edge Removal with Node Removal

The edge removal operation (see Section 7.4.5) can complement node removal when a large number of edges are added due to node removal. Rather than making ad-hoc approximations, the domain experts should evaluate the CIs modelled by each of the new edges, and remove the ones that are considered to be trivial. An example of a node removal followed by edge removals is shown by the case-study in Section 7.5.

7.4.3 Node Merging

The second of the compatible abstraction operations is merging multiple nodes $T_1, \dots, T_n$ into a single node M. Two nodes T_i and T_j can be merged into a single node M when an edge can be added between T_i and T_j and this edge is reversible by adding extra edges to cover it. The network with the merged variable combines the networks with this edge in both directions. The merging operation is a compatible abstraction since covering and reversing the edge between T_i and T_j does not add CI conditions to the BN structure (see Section 7.4.2). When we merge the nodes T_i and T_j into M, the state-space of M becomes the Cartesian product of the state-spaces of T_i and T_j.

Node merging does not change the probability distribution that factorises on the BN. The main difference is that T_{i} and T_{j} cannot be observed separately, they must be observed or unobserved together. Therefore, while comparing the CI conditions between G_X and G_A observing M is equivalent to observing T_{i} and T_{j} together.

An example of the merging operation is shown in Figure 7.6 where the nodes X and Y are merged into XY. The first step is to add an edge between X and Y as shown in Figure 7.6b. Next, this edge is covered and reversed as shown in Figure 7.6c. The final BN with the merged variable XY (see Figure 7.6e), can be seen as equivalent to the BNs with edges in both directions (see Figure 7.6d).

Multiple nodes can be merged into a single node by repeating the merging operation pairwise. For instance, the nodes X, Y and Z in Figure 7.6a can be merged pairwise by first merging X and Y to XY (see Figure 7.6e), then merging XY and Z to XYZ as shown in (see Figure 7.7).

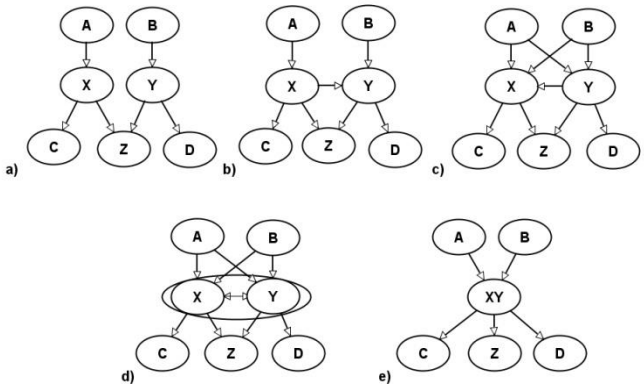


Figure 7.6 (a) Initial BN (b) $X \rightarrow Y$ added (c) $X \rightarrow Y$ reversed (d) BNs with $X \rightarrow Y$ and $X \leftarrow Y$ combined (e) XY merged

A BN with merged nodes has the same or fewer CI assertions than the initial BN. The edges added after edge reversal and covering operations may remove some CIs. Moreover, some of the CI encoded in the initial BN disappears after merging since the merged variables cannot be observed separately. For example, A and B is independent given that X and Y is observed in Figure 7.8a, but this independence disappears after merging X and Y as shown in Figure 7.8b.

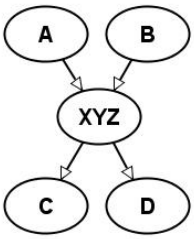


Figure 7.7 XY and Z merged

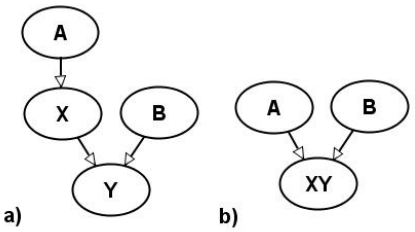


Figure 7.8 (a) Initial BN (b) X and Y Merged

The merging operation introduces a cycle if there is a directed path between the nodes to be merged, T_i and T_j, that includes some other node. These cycles have to

be eliminated since BNs are acyclic graphs. Our solution is to reverse some of the edges in these directed paths until none of the directed paths remain between T_i and T_j. In other words, if there is a directed path between T_i and T_j that includes other nodes, we break this directed path by reversing some of the edges in it. There are two issues to consider while reversing these edges:

- The reversed edges must not introduce a cycle as well.
- The edges must be covered before reversing (see Section 7.4.2).

For example, we cannot immediately merge X and Y in Figure 7.9 since there is a directed path $X \rightarrow B \rightarrow D \rightarrow Y$ between X and Y that includes the nodes B and D. We have to reverse one of the edges between $X \rightarrow B$, $B \rightarrow D$ or $D \rightarrow Y$ to prevent the merging operation from introducing a cycle.

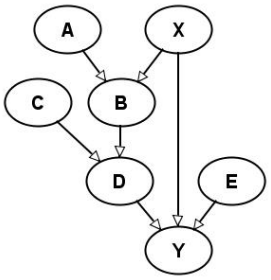


Figure 7.9 BN before X and Y are merged

We consider whether any of $X \rightarrow B$, $B \rightarrow D$ or $D \rightarrow Y$ are covered or not. If one of them is covered, it can be reversed. In this case, none of these edges are covered therefore we need to add more edges to cover and reverse them:

- Parents of $B = \{A, X\}$, parents of $X = \{\,\}$ therefore the edge $X \rightarrow B$ is not covered. The edge $A \rightarrow X$ has to be added in order to make this edge covered (see Figure 7.10a).
- Parents of $D = \{C, B\}$, parents of $Y = \{D, X, E\}$; therefore the edge $D \rightarrow Y$ is not covered. The edges $C \rightarrow Y$, $B \rightarrow Y$, $X \rightarrow D$ and $E \rightarrow D$ must be added (Figure 7.10b).
- Parents of $B = \{A, X\}$, parents of $D = \{C, B\}$; therefore the edge $B \rightarrow D$ is not covered. The edges $A \rightarrow D$, $X \rightarrow D$ and $C \rightarrow B$ must be added (Figure 7.10c).

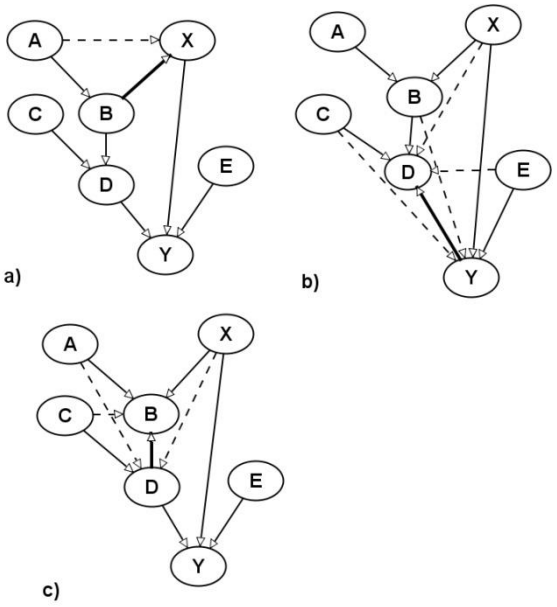


Figure 7.10 (a) $X \rightarrow B$ reversed (b) $D \rightarrow Y$ reversed (c) $B \rightarrow D$ reversed

Figure 7.10a, Figure 7.10b and Figure 7.10c shows the model structure when the edges $X \rightarrow B, D \rightarrow Y$ and $B \rightarrow D$ are covered and reversed respectively. The edges that were added are shown by dashed lines, and the reversed edges are shown by bold lines in these figures. The BN in Figure 7.10a does not have a directed path between X and Y therefore the nodes X and Y can be merged. The BNs in Figure 7.10b and Figure 7.10c, on the other hand, still have directed paths $X \rightarrow B \rightarrow Y$ and $X \rightarrow D \rightarrow Y$ respectively. We need to reverse more edges in order to merge X and Y in these BNs. Since we already have a structure without a directed path between X and Y (Figure 7.10a), we will continue with this structure.

The BN in Figure 7.11 is a compatible abstraction of the BN shown in Figure 7.9. The probability distribution $P(A, B, C, D, XY, E)$ can be represented in the same way in these two BNs assuming that observing XY is equivalent to observing X and Y at the same time in the initial BN.

The BN in Figure 7.11 has one more edge compared to the initial BN. It is possible to select the sequence of reversals that leads to the fewest number of additional edges or to the smallest state-space. In our example, reversing $X \rightarrow B$ (Figure 7.10a) added the least number of edges compared to reversing $D \rightarrow Y$ (Figure 7.10b) or $B \rightarrow D$ (Figure 7.10c).

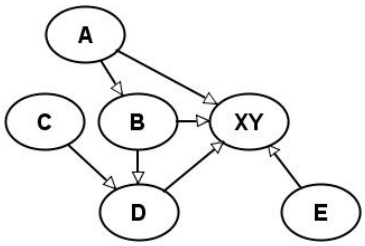


Figure 7.11 Compatible Merging of *X* and *Y*

In general, it is possible to merge any pair of nodes in a BN. In the worst case the process described above will lead to a fully connected BN, with no CI conditions. Since all fully connected BNs that are formed by the same set of nodes have equivalent CI conditions (i.e. none), any (acyclic) combination of edge directions can be reached by an appropriate sequence of reversals of covered edges. Therefore, any edge can be reversed in a fully connected network, and any pair of nodes can be merged.

Although possible, not all merging operations are sensible from a knowledge engineering perspective. In Figure 7.9, X is a cause of Y, and B and D are the intermediate variables between the cause and effect. Merging the cause and effect, and leaving the intermediate variables in the model would not be sensible in most knowledge engineered BNs. One would prefer to merge an intermediate variable with either its cause or its effect in order to simplify a causal BN.

Using Node Merging with State-Space Abstraction

Abstraction by merging variables is suitable when multiple variables in the model are parts of the definition of a more abstract concept. For example, one option to model an engine fault in a BN is to represent it with a single variable with 'Yes/No' states. Alternatively, the same concept can be modelled in detail with multiple variables each representing faults in different components of the engine such as faults in pistons, alternators and crankshaft. The merging operation makes it possible to show the link between the abstracted and detailed representations in this case. The variables about the individual engine components can be merged into the abstract variable about overall engine fault. After node merging the state-space of the abstract variable becomes the Cartesian product of the states of the merged variables. However, we would expect the abstract variable to have more abstract states as well.

State-space abstraction will follow node merging operation in such cases (see Section 7.4.4). An example of a node merging followed by state-space abstraction is shown in the case-study in Section 7.5.

7.4.4 State-space Abstraction

State-space abstraction collapses multiple states of a variable into a single state. For example, suppose a variable has 4 states named {None, Moderate, Severe, Profound}. We can collapse 'Moderate', 'Severe' and 'Profound' states into a single state called 'Present', as a result, the variable will have 2 states named {None, Present}. Consequently, the NPT of the variable requires fewer parameters and its learning, or elicitation, becomes simpler. This operation is often used in combination with node merging (See Section 7.4.3).

State-space abstraction is an incompatible abstraction although it makes no change in the BN structure (Chang and Fung, 1990; Wellman and Liu, 1994). The changes in the CI and their effects to the domain representation should be discussed with the experts before making state-space abstractions.

7.4.5 Edge Removal

Edge removal is an incompatible abstraction that always adds CI assertions to the BN structure (see Section 7.2). Therefore, an edge should not be removed if it is not recommended by experts or statistical evidence. By adding CIs, edge removal decreases the parameter space and computational complexity of the BN. Edges added after the node removal and merging operations, and weak relations represented in the BN structure are suitable candidates for edge removal. The remainder of this section presents knowledge and data driven approaches for assisting edge removal.

Using Domain Knowledge for Selecting Edges to be Removed

Before removing an edge, the effects of removing the edge should be reviewed with domain experts. The review focuses on the strength of the relation modelled by the edge, and the CI conditions introduced due to removing it. It may not be feasible to

review the CI conditions between all subsets of variables especially if the BN is large. In this case, the domain experts can identify the important variables in the domain, and limit the review to these variables.

Edge removal can be used with node removal and node merging as these operations can add many edges to the BN structure (see Section 7.4.2 and 7.4.3). For example, when the node R is removed from the BN in Figure 7.12a 3 edges are added between the variables B, C and D as shown in Figure 7.12b. Removing these edges simplify the BN but it also adds CIs that are not present in the initial BN. For example, removing $C \rightarrow D$ and $B \rightarrow C$ as in Figure 7.12c adds two new CI conditions: $(B \perp C|A)$ and $(C \perp D|A)$. The knowledge engineers should discuss with the domain experts whether these CIs are acceptable in the domain.

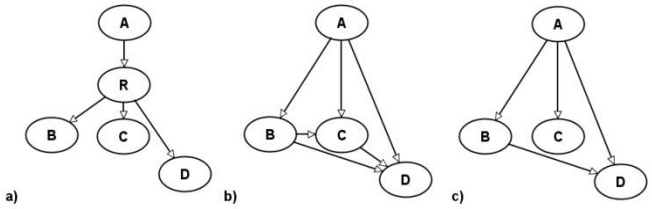


Figure 7.12 (a) Initial BN (b) Node R removed (c) Edges $B \rightarrow C$ and $C \rightarrow D$ removed

Using Data for Selecting Edges to be Removed

If data is available, statistical tests, including χ^2, Margaritis (2004) and Dash and Druzdzel (2003), can be used to assist edge removal. These tests can identify the statistically significant CI assertions between variables. Alternatively, model selection scores, such as Bayesian Information Criterion (Schwarz, 1978), can be used to compare the effects of removing edges. Another possible approach is to assess the sensitivity of certain variables to edge removals (Renooij, 2010).

7.5 Case-Study: Shock

In this section, we use the ATC BN as a case study to illustrate the abstraction methodology. We apply the abstraction methodology to derive the shock fragment of the ATC BN from a more complicated BN structure elicited from the domain experts (see Section 5.3.3 for a description of the ATC BN). In the remainder of this section,

Sections 7.5.1 and 7.5.2 describe clinical knowledge modelled in the shock fragment and the initial BN structure elicited from the experts. Sections 7.5.3 – 7.5.6 applies each of the abstraction operations to the initial BN.

7.5.1 Background

Blood is the medium through which oxygen, which is vital for metabolism, is delivered to tissues. Patients with significant blood loss are unable to adequately perfuse their tissues with blood and thus unable to adequately deliver oxygen to the tissues. The body tissues start to die if starved of oxygen for a prolonged time. The body responds in several ways to compensate for the effects of blood loss. First, the heart rate increases in order to maintain normal perfusion and blood pressure. However, as the blood loss increases the increased heart rate cannot compensate and the blood pressure decreases. Second, as oxygen delivery to the tissues decreases, less efficient (anaerobic) metabolism takes over to compensate for the lack of oxygen. The by-products of anaerobic metabolism increase the acidity of blood and tissues, which in turn causes the respiration rate to increase. A patient in whom these mechanisms are operating as a result of bleeding is said to be in a state of haemorrhagic shock.

Table 7.2 Definitions of Variables in the Shock fragment

Variable	Definition
Bleeding Body Parts	The number of bleeding body compartments.
Hypovolemia	Decrease in the volume of blood in circulatory system. Heart rate (HR) increases as a result of this to maintain normal blood pressure.
Cardiac Output	Volume of blood ejected from left side of the heart in 1 minute. As the blood loss increases cardiac output and blood pressure (SBP) will fall. Urine output and Glasgow coma score (GCS) is dependent on the perfusion of blood to kidney and brain respectively.
PVS	Vascular resistance to the flow of blood in peripheral arterial vessels. Degree of PVS can be estimated by capillary refill time (CRT).
Oxygen Delivery	Amount of oxygen delivery to the tissues. Body temperature (Temp) indicates the degree of overall oxygen delivery to the tissues.
Metabolic Acidosis	As oxygen delivery to the tissues decreases, less efficient (anaerobic) metabolism takes over to compensate for the lack of oxygen. The by-products of anaerobic metabolism increase the acidity of blood and tissues, which in turn causes the respiration rate (RR) to increase. The degree of metabolic acidosis can be estimated by pH, base excess (BE) and lactate values in blood.
Death	The risk of death in 48 hours

It is not possible to directly observe the state of the many variables explained above. For example, the oxygen perfusion to the tissues cannot be measured precisely. Instead, these 'latent' variables are inferred from related measurements and clinical observations. Reasoning of this type suggests the potential of a BN to detect shock.

7.5.2 Shock Fragment of the ATC BN

We developed a BN structure for reasoning about the physiology of bleeding patients and for predicting the risk of death from shock. We did not consider the limitations of data at this stage; our aim was to model all main variables and relations indicated by the domain experts. The structure of the initial BN is shown in Figure 7.13. The circular nodes in this BN are the main clinical variables, and rectangular nodes are the measurements and observations relevant to the main clinical variables. Table 7.2 shows the definition of the main variables in the BN. We do not have data about some of these variables since they either cannot be directly observed or are not recorded in practice. We choose to simplify the BN structure using abstraction operations before applying the methodology described in Section 5 to learn the parameters of this BN fragment.

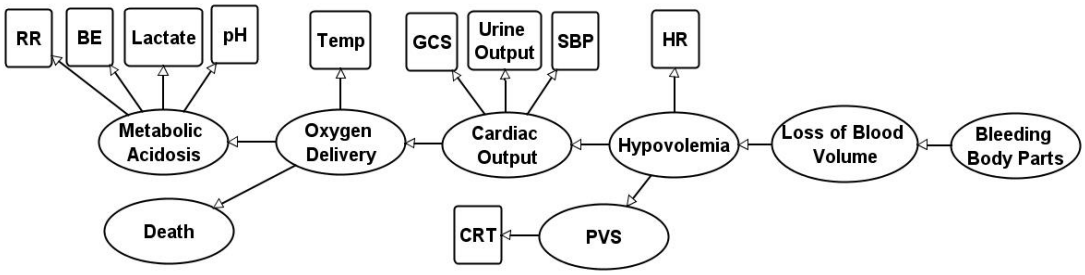


Figure 7.13 Initial Structure of the Physiology BN

7.5.3 Node Removal

We removed the Metabolic Acidosis, Loss of Blood Volume, CRT, PVS, Urine Output, RR, Temp and GCS variables from the BN structure. These variables are selected with domain experts, considering both the objectives of the BN and availability of data.

Removing the blood loss and metabolic acidosis variables requires 1 and 4 edge reversal operations respectively. The measurement variables CRT, RR, Temp and Urine Output are already barren nodes therefore they can be removed without any edge reversal operations. The PVS variable become barren after CRT is removed so no edge reversal is needed for it either. The resulting BN structure after the node removal operations can be seen in Figure 7.14.

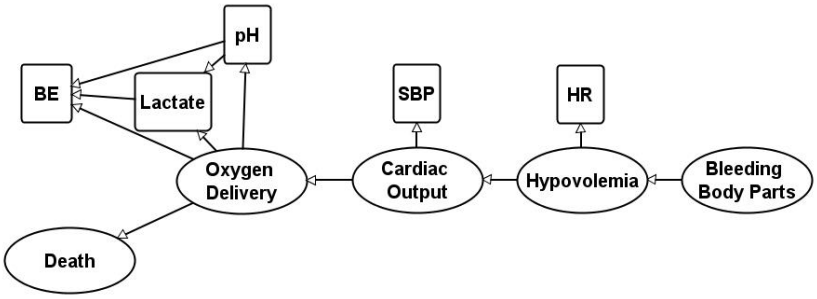


Figure 7.14 BN Structure after Node Removals

Removing metabolic acidosis with a compatible abstraction made the abstracted BN more complex: 3 edges were added between the children of this variable and one of its children (BE) has 3 parents now. These edges do not represent causal relations but removing them will introduce CIs that were not present in the initial BN. Before simplifying the graph, the effects of removing each of these edges, in terms of CI assertions, should be investigated by the domain experts (see Section 7.4.5).

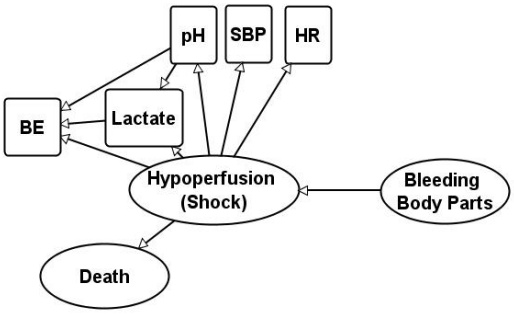


Figure 7.15 BN Structure after Node Merging

7.5.4 Node Merging

The state of the cardiac output, oxygen delivery and hypovolemia variables cannot be directly observed. The domain experts do not recommended removing these variables as they represent an important physiological mechanism about bleeding.

The parameters of these variables can be estimated by combining the observed data of their measurements and expert knowledge using the methodology described in Chapter 5. However, the domain experts indicated that the BN structure could be further simplified before starting to learn or elicit the parameters.

The domain experts stated that oxygen delivery, cardiac output and hypovolemia are elements of a more abstract physiological definition that can be called hypoperfusion or shock. Shock is clinically defined as a metabolic disturbance due to the failure of the circulatory system to maintain adequate perfusion to vital organs. Hypovolemia and cardiac output is associated with failure of the circulatory system, oxygen delivery and its relation to death represents the metabolic disturbance. Therefore these 3 variables can be merged into a single variable. We did not have to reverse any edges during merging since the merged node do not have directed path that includes a non-merged node between them. The resulting BN after merging is shown in Figure 7.15.

Table 7.3 States of Shock before and after State-Space Abstraction

Before state-space abstraction[1]		**After state-space abstraction**
A,A,A	→	Abnormal
N,N,N : : N,A,A	→	Normal

[1]A: Abnormal, N: Normal

7.5.5 State-Space Abstraction

The node merging operations have simplified the structure of the BN resulting in a hypoperfusion variable that is easier to elicit clinically. However, the state-space of the hypoperfusion variable still requires simplification since it is formed by the Cartesian product of the states of each merged variable. For simplicity we defined binary states for the merged variables: {Normal, Abnormal}. The domain experts stated that Shock is present when all of the 3 factors are abnormal. In other words, we are interested in two states of Shock variable: the state when all of the factors are

abnormal, and when all of the factors are not abnormal. An illustration of this state-space abstraction is shown in Table 7.3.

7.5.6 Edge Removal

Several edges were added between BE, Lactate and pH as a result of removing the 'Metabolic Acidosis' variable. We listed the CI conditions that removing each of these edges can bring, and discussed these CIs with the domain experts. For example, when we remove pH → BE, pH becomes independent of BE given that Shock or the variables that are merged into Shock are observed. This CI was not present in the initial BN. The resulting BN structure after edge removals can be seen in Figure 7.16.

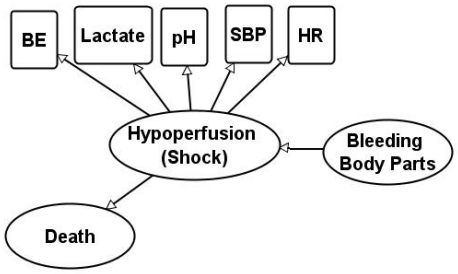
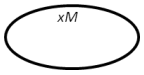
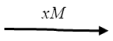
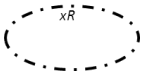
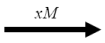
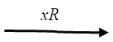
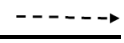


Figure 7.16 Final Abstracted BN after Edge Removal

7.6 Graphical Notation for Abstractions

In this section, we present a graphical notation that shows the order and type of abstraction operations applied to a BN. The notation shows how the abstracted structure is derived from the initial BN by showing each abstraction step. This is essential for communicating the knowledge-base of a BN or for deriving a more detailed version of the BN when, for example, more data become available. We use both the initial and the abstracted BN structures to show the steps of abstraction. These structures are annotated by the symbols shown in Table 7.4. In the remainder of this section, we illustrate the application of these symbols using the case study.

Table 7.4 Symbols for Abstraction Operations

Operation	Symbol	Operation	Symbol
Node merging		Edge added after node merging	
Node removal		Edge reversed after node merging	
State-space abstraction		Edge added after node removal	
		Edge removal	

Removed nodes are shown with dashed boundaries and annotated by '*x*R*'* where *x* is an integer that indicates the order of the abstraction operation and R indicates a removal operation. Edges that are added as a result of node removal operation are shown by an '*x*R*'* annotation with the same 'x'. In the case study, first abstraction operation is the removal of the 'Metabolic Acidosis' node, and this operation adds several edges to the model. Both the variable, and the added edges are annotated by '1R' in the initial and abstracted BNs respectively as shown in Figure 7.17.

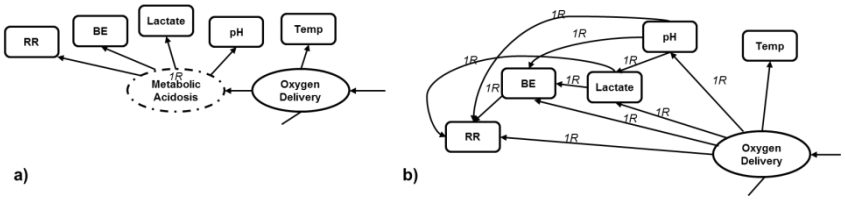


Figure 7.17 Notation for initial (a) and abstracted (b) BN fragment for removal of Metabolic Acidosis variable

Nodes involved in a merging operation are annotated by '*x*M' where *x* is the order of the abstraction operation and M stands for 'merging'. In the shock example, we merged three nodes: 'Oxygen Delivery', 'Cardiac Output' and 'Hypovolemia'. The node resulting from this merging operation is called 'Hypoperfusion (Shock)'. We annotate each of these variables with '2M' since it is the second abstraction operation and it is a merging operation. The initial and abstracted BNs after this operation can be seen in Figure 7.18.

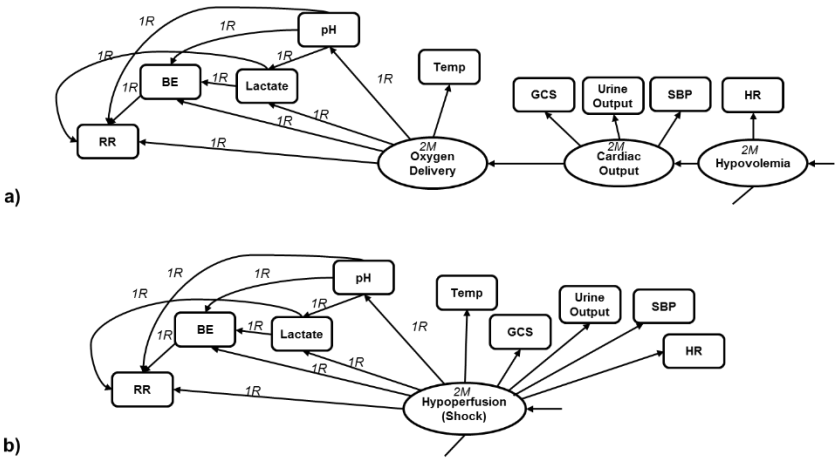


Figure 7.18 Notation showing before (a) and after (b) merging multiple variables into Shock variable

A node merging operation may introduce edges (see Section 7.4.3). In this case, both the added and the reversed edges are annotated by '*x*M', and the reversed edges are shown by bold lines.

Variables with state-space abstractions are shown by a double lined boundary, and removed edges are shown by dashed lines. The order of state-space abstraction and edge removal does not change the BN structure (other than the removed edge itself) so the order is not annotated. However, the collapsed or removed states after state-space abstraction must be documented (for example, as in Table 7.3). In the case study, the state-space of 'Shock' is abstracted and the edges between 'pH', 'RR', 'BE' and 'Lactate is removed in order to simplify the BN, as shown in Figure 7.19.

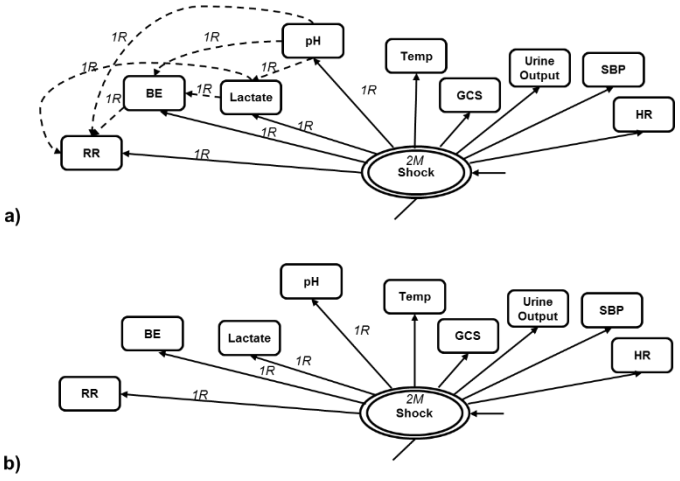


Figure 7.19 Notation showing edge removals and state-space abstraction (a) and the BN structure after edges are removed (b)

Figure 7.20 shows the initial and abstracted structure of the Shock BN with all the abstraction operations annotated. The graphical notation presented in this section clarifies the derivation of the final abstracted structure from the initial – detailed – structure. Multiple graphs each showing a particular step of abstraction should be recorded if the BN is simplified with multiple abstraction operations.

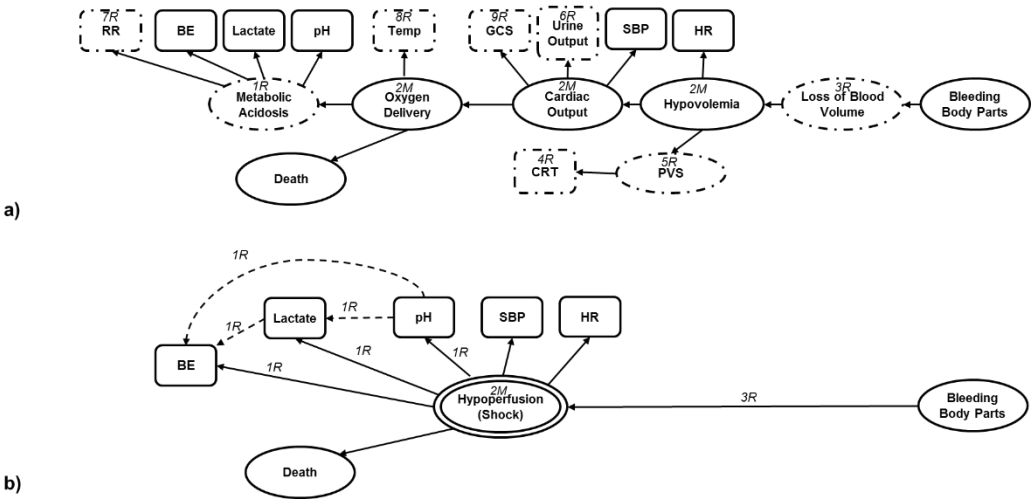


Figure 7.20 Initial (a) and abstracted (b) bleeding physiology BN with abstraction notation

7.7 Motivation from ABEL

ABEL (Patil, 1981; Patil et al., 1981) was a pioneering clinical expert system that was developed for diagnosing acid-base disturbances of patients. Given laboratory data about a particular patient, ABEL generates the relevant causal diagrams from its knowledge-base, which is known as the patient specific model (PSM). It reasons by abstracting and elaborating these causal models to make diagnostic inferences, which is considered to be similar to how clinicians express their decisions. The causal models at the higher levels are directed acyclic graphs like BNs. Although widely referenced, PSMs have not become a commonly used approach for developing clinical decision support models.

The causal diagrams of ABEL have several differences from the BN formalism. Firstly, each node in a PSM represents a single state of a variable, whereas the nodes in a BN are variables that can have multiple states. Secondly, the lower abstraction levels of PSM can have feedback loops which are always eliminated at higher levels; but BNs are acyclic graphs. Thirdly, PSM do not reason probabilistically and its

reasoning mechanism does not take the prior probabilities of diseases into account; BNs have superior probabilistic reasoning algorithms that are able to calculate complex learning and inference problems. Finally, BNs are lacking techniques for abstracting their knowledge-base for different levels of detail in a similar way to the abstraction mechanism in ABEL. Abstraction is clearly necessary for developing knowledge-based BNs for complex domains, and for explaining these models to external users. Our work on this paper was initially motivated by ABEL, notably its hierarchical structure and abstraction operations.

7.8 Conclusion

This chapter proposed abstraction as a knowledge engineering method for simplifying a BN structure. The method is illustrated by a medical case study about haemorrhagic shock. Our method provided:

1. A sufficient set of operations that simplify a BN by removing and merging nodes, removing edges and reducing the number of states

2. The compatibility properties of each abstraction in terms of CIs added to the BN structure

3. A graphical notation that captures the sequence and type of abstraction operations, and thereby showing the link between the knowledge-base and the abstracted BN

Some of the abstraction operations in our method are based on existing techniques that has been mainly used for learning and inference problems. This paper emphasises the potential of these techniques for following a systematic approach to knowledge engineering. The compatible abstraction operations do not add CIs but they can make the BN structure increasingly complex by adding edges. Incompatible abstraction can be used to simplify the structure but they approximate the BN structure by adding new CI conditions. Trade-offs between the approximations and complexity must be considered carefully. The next stage in this research is to implement these abstraction operations in BN software such as AgenaRisk (Agena Ltd, 2013). The BN software, augmented with the abstraction operations, would

guide the user through BN development by showing the impacts of compatible abstractions and presenting the approximations resulting from incompatible abstractions. Moreover, it would be beneficial to evaluate practical impacts of our method to BN development.

Chapter 8

Evidence behind Clinical Bayesian Networks

Clinicians may reject a prognostic model if they are not convinced that the model's performance, for their patients, will be similar to its published performance in validation studies (Moons et al., 2009a; Wyatt and Altman, 1995). It may be difficult to evaluate the prospective performance of a model in clinical practice if the reasoning mechanism and clinical evidence behind the model is not completely understood (Wyatt and Altman, 1995). The graphical structure of a BN is well suited for representing causal and associational reasoning between a large number of variables. However, the current representation of BNs is not descriptive enough to show the details of clinical evidence behind a BN. In many clinical BNs, the names of variables are often short and ambiguous, the relations modelled by edges are not explained, and the corresponding evidence from the clinical literature and data is not presented. Consequently, domain knowledge and evidence behind these BNs are clear only to their developers.

The clarity of a BN and relevant evidence also enables other domain experts to recommend local modifications in the BN when new evidence becomes available. For example, a part of the BN may become obsolete due to recent evidence and a modification may be required. Prognostic models are often redesigned from scratch in such circumstances by disregarding all information present in the previous model (Altman et al., 2009; Moons et al., 2009a; Royston et al., 2009). BNs allow local modifications as their parameters can be defined locally, and their graphical structure can be built consistent with clinical evidence. However, evidence relevant to all parts

of a BN, including relations, variables, and a group of variables (BN fragments), must be defined in order to identify and refine the parts with obsolete evidence.

In this chapter, we present an evidence framework for representing the evidence behind clinical BNs. Our framework is composed of two elements: an ontology for organising evidence, and a browser for presenting evidence. The main aim of the evidence framework is to organise and present evidence that supports or conflicts with a clinical BN. We use the ATC BN (see Section 5.3.3) as a case study to illustrate the evidence framework. Evidence behind ATC BN can be browsed online at the ATC BN website (ATCBN, 2013).

8.1 Challenges of the Evidence Framework

The challenges of describing the evidence behind a BN can be summarised in two points:

1. **Organising Evidence:** The evidence about a BN must be detailed enough to prevent ambiguities about its relevance, source and type. First of all, the evidence can be relevant to different parts of the BN: some can be relevant to a particular variable or relation whereas others can be relevant to an entire group of variables. The evidence framework must be able to store evidence relevant to the fragments of the BN as well as to the individual variables and relations. Secondly, there may be different items of evidence. A part of the BN may have evidence from the data, and another part may be based on clinical publications. Thirdly, conflicting evidence must also be taken into account in order to have a comprehensive description of the evidence. The evidence framework must contain all of the relevant evidence even if it belongs to something not modelled in the BN. The type of the evidence must be recorded to show whether the evidence supports or conflicts with the BN model. Finally, the meaning of the variables must also be clear. Variable names in a BN structure are often short and may be ambiguous (see Section 3.3); more detailed information must be given in the evidence framework.

2. **Presenting Evidence:** Another challenge is to present evidence in a clear and simple way without losing any important information. Although the evidence

framework may require a complicated structure to store the evidence data, users do not necessarily need to see the technical details of the structure when browsing the evidence. Evidence should be presented in a user-friendly environment that is compatible with commonly available software such as web page browsers.

The first challenge shows that BNs require a supporting evidence structure that is able to cope with the complications of organising and recording evidence with adequate amount of detail. The evidence structure must be flexible to deal with changing requirements in different applications. For example, some clinical models may require additional types of evidence to be defined, and the evidence structure must be flexible in addressing such requirements. Moreover, the evidence structure must be able to answer the queries about its completeness by showing the parts of the BN that lacks evidence.

The second challenge can be overcome by an evidence browser that is automatically generated from the evidence structure. The evidence browser aims to present the evidence in a clear and understandable way without necessarily showing the details of the structure that organises evidence data.

In order to overcome these challenges, we propose an evidence framework that is composed of two elements: a structure for organising the evidence data and a web page for browsing the evidence. We use ontology technology to build the structure and organise the data about evidence. The web page is automatically generated from the ontology. In the following sections, we describe the structure (Section 8.2) and browser (Section8.3) parts of the evidence framework.

8.2 Evidence Structure

We use the web ontology language (OWL) (W3C, 2013) framework for modelling and organising the evidence data, and the Protégé software (version 4.3.0) (Knublauch et al., 2004) to create and populate the OWL ontologies. In Section 8.2.1, we give a brief introduction to ontologies. We present the structure of the evidence ontology in Section 8.2.2, illustrate how data is entered to this structure in

Section 8.2.3, and show queries for assessing the completeness of evidence in Section 8.2.4.

8.2.1 Introduction to Ontologies

Ontology is a formal definition of the relations among terms (Berners-Lee et al., 2001), and OWL is a flexible language for building ontologies (W3C, 2013). Our primary reason for storing the evidence data in an ontology is their flexibility in building, modifying and querying a data structure. Our aim is to develop a general structure for recording and presenting evidence for BNs in medical domain but we do not assume that our evidence structure will satisfy the needs in all other BN applications. For example, some clinical applications may require additional types of evidence to be defined in the ontology structure, and ontologies offer a simple and robust framework for making such changes. Making similar modifications in relational databases is, however, difficult and time-consuming due to their highly structured schema and query system.

An OWL ontology is composed of individuals, classes, object properties and data properties. Individuals represent objects in the domain, and classes represent the sets that individuals belong. Properties define the relations between individuals: object properties define the relation between two objects, and data properties define the relation between an object and a data value. In the remainder of this chapter, individuals are represented by purple diamonds, classes are represented by yellow circles, object properties are represented blue squares, and data properties are represented by green squares as shown in Figure 8.1.

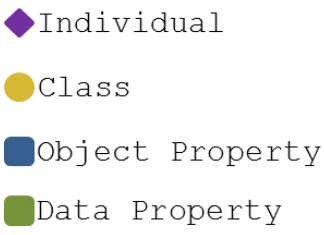


Figure 8.1 Representation of Ontology Elements

An example of a relation between two individuals is shown in Figure 8.2. In this example, John and Anna are individuals, and hasMother is an object property that defines the relation between these individuals.

Figure 8.2 Individuals and Properties

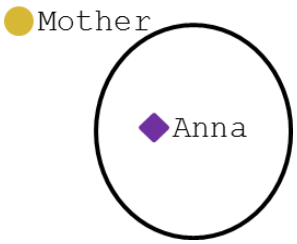


Figure 8.3 Classes and Individuals

Members of a class can be defined by the user or it can be inferred from classes and property characteristics. For example, we can either manually define that Anna is a member of the Mother class (see Figure 8.3) or we can infer this from the characteristics of the hasMother property. Let the hasMother property have 3 characteristics:

1. The inverse of the hasMother property is the hasChild property
2. The domain of the hasMother property is the Child class
3. The range of the hasMother property is the Mother class

Based on each of these characteristics, we are able to infer that (see Figure 8.4):

1. Anna hasChild John – inferred from the inverse of hasMother
2. Anna belongs to the Mother class – inferred from the domain of hasMother
3. John belongs to the Child class – inferred from the range of hasMother

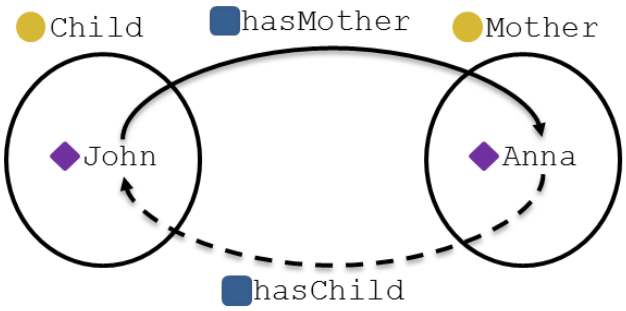


Figure 8.4 Inferred Classes and Properties

Classes can be defined in multiple levels of hierarchy. For example, we can define that the Mother class is a subclass of the Woman class, and that the Woman class is a subclass of the Human class (see Figure 8.5). In this case, all individuals that are members of the Mother class are also members of the Woman and Human classes. An individual can belong to multiple classes with different hierarchies unless the classes are explicitly defined as disjoint. For example, Anna can belong to both Mother and Child classes but she cannot belong to both Human and Item classes as these are disjoint classes.

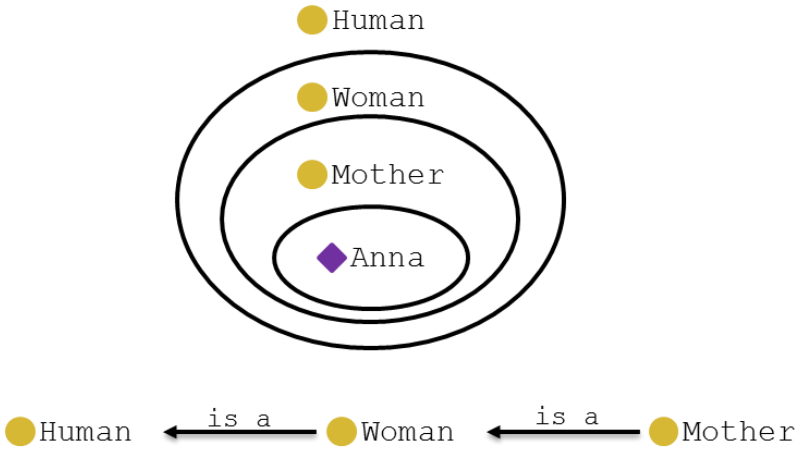


Figure 8.5 Class Hierarchy

Data properties describe the relation between an individual and a data value. Figure 8.6 shows the relation between an individual and a string type of data value using a data property.

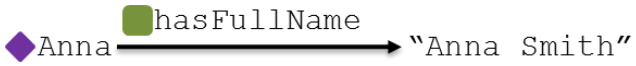


Figure 8.6 Data Property Example

OWL contains many other features for defining properties and objects. A thorough description of OWL is, however, beyond the scope of this chapter, and the readers are referred to Allemang and Hendler (2011), and Segaran et al. (2009) for an introduction to ontology modelling and OWL. In the following section, we describe the classes, and properties of the evidence ontology.

8.2.2 Evidence Ontology

Our evidence structure is based on three main classes: BN element, evidence and source (see Figure 8.7). The BN element class contains the nodes, edges and BN fragments that form a BN. Each node in a BN represents a variable, and each edge represents a relation. We use the terms node and variable, and edge and relation interchangeably throughout this chapter. The evidence class defines the type and description of evidence. The source class describes the source of evidence; which can be a dataset, a domain expert or a scientific publication. Individuals of the evidence class may have multiple sources, for example, a relation in the BN structure may have evidence from several publications.

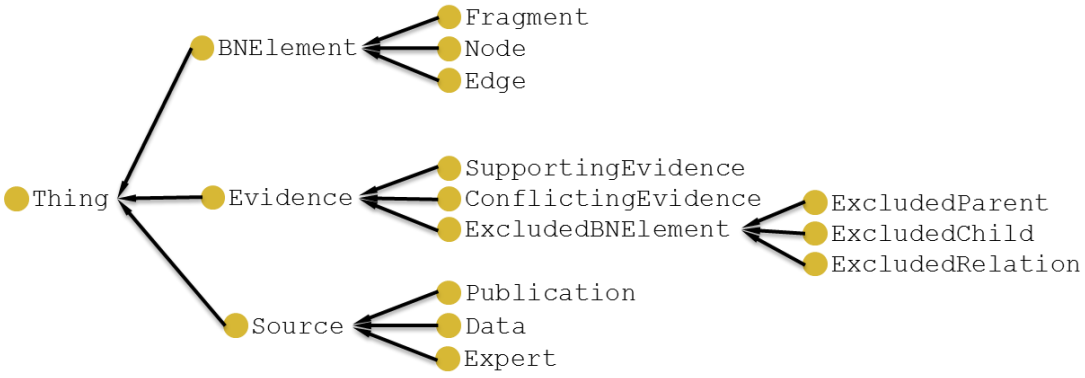


Figure 8.7 Class Hierarchy of the Evidence Ontology

In the remainder of this section, we describe the subclasses and properties related to the BN element, evidence and source classes.

8.2.2.1 BN element

The BN element class has 3 subclasses: fragment, node and edge. This section describes the object and data properties related to each of these subclasses.

Fragment

A BN is composed of variables and edges, and it is necessary to clarify the meaning of these in order to describe knowledge behind the BN model. Several studies used BN substructures to assist BN development and to give a more concise summary of a BN model (see Section 2.6.1). A BN substructure represents a part of the BN that describes an important concept in its domain. Our ontology defines BN substructures in a general way as 'BN fragments'. Any group of variables can be defined as a BN fragment in the evidence ontology. The complete structure of the BN is always defined as a BN fragment containing all of the other fragments and nodes in the BN.

The information about a BN fragment is stored within the fragment class. Figure 8.8 shows the object and data properties related to this class. A BN fragment may contain nodes and other – smaller – BN fragments. Members of a fragment can overlap with other fragments. Description of clinical knowledge modelled in the BN fragment, and evidence relevant to the entire BN fragment can be stored in the evidence ontology.

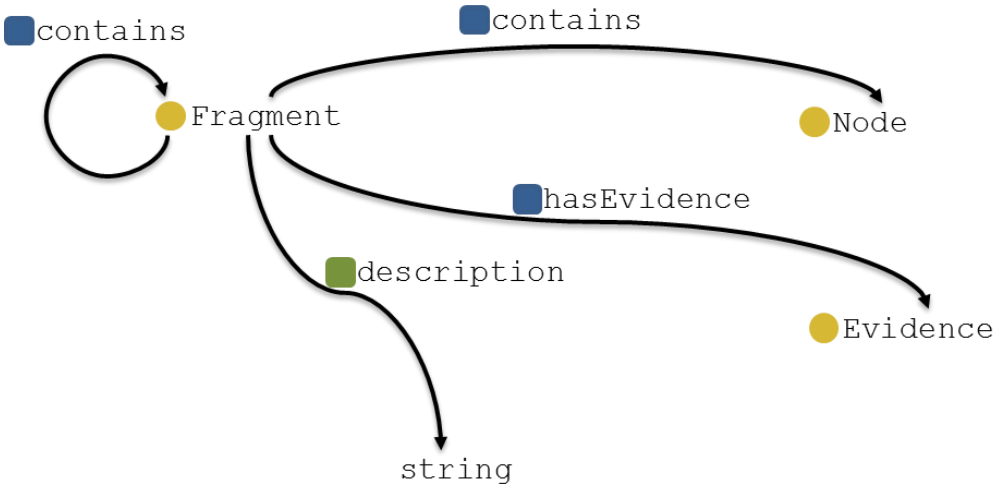


Figure 8.8 Object and Data Properties related to Fragment Class

Subclasses of the fragment class can be defined if certain type of BN fragments, such as idioms (Neil et al., 2000) and object oriented BNs (Koller and Pfeffer, 1997), are used repetitively in the BN. In the ATC BN, we use measurement idiom structure multiple times to define parts of the BN structure (see Section 5.3.3). In order to distinguish the measurement idioms from other BN fragments in the evidence framework, we could add a class called 'MeasurementIdiom' as a subclass of the fragment class (see Figure 8.9).

Figure 8.9 Subclasses of Fragment Class

Node

Individuals of the node class represent the variables in a BN. Knowledge modelled in a BN cannot be understood without having a clear understanding of the meaning of its variables. Variable names are often short, consisting of a few letters, therefore they may not clearly describe the concept that is represented by the variable (see Section 3.3). For example, a BN variable named 'Heart Rate' may be sufficient to show that this variable represents a measurement of the patient's heart rate; but this name may not be descriptive enough if the time and location of this measurement is important for the use of the BN. In order to avoid such ambiguities, the evidence ontology stores a description of each BN variable (see Figure 8.10).

The variables in a BN can be discrete or continuous. For discrete variables, we define the names of the states in the evidence ontology. Continuous variables have infinitely many states therefore we define the upper and lower bounds of the distribution.

Since the edges of a BN are directed, we have an object property for both the edges that are directed to the variable (hasIncommingEdge) and the edges that are directed away from the variable (hasOutgoingEdge) (see Figure 8.10). The hasParent and hasChild properties show the parent – child relations between the BN variables. These 4 properties about edge directions and parent – child relations are inferred from the object properties related to the edge class described in the remainder of this section. The within property shows the BN fragments that contain the variable.

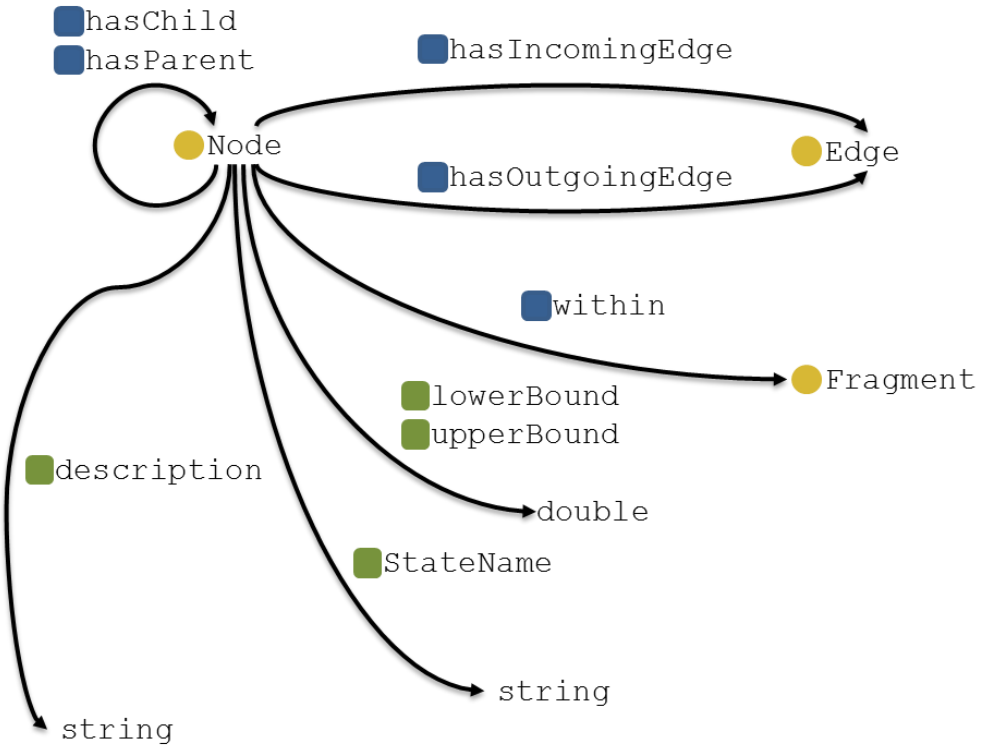


Figure 8.10 Object and Data Properties related to Node Class

Different sources can be used to define the parameters of different BN variables as BN parameters can be learnt locally. For example, we can use data to define the parameters of some variables, and use expert knowledge to elicit the parameters of others. For each variable, we record the source used for defining its parameters (see Figure 8.11).

Some variables may not be modelled in the BN even when evidence exists that they are relevant to the problem domain. For example, the knowledge engineers and domain experts may choose to exclude some variables to keep the BN simple. The evidence about the excluded variables and relations is recorded to have a comprehensive evidence base for the BN (see Figure 8.11).

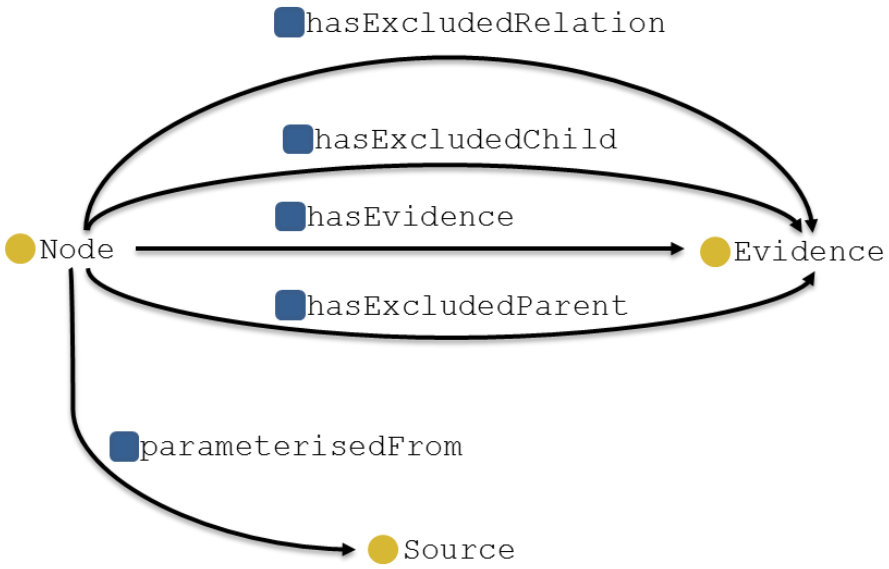


Figure 8.11 Object Properties related to Node, Evidence and Source Classes

Edge

Edges represent the relations in a BN. BNs are directed acyclic graphs, therefore only one directed edge can exist between two variables. The variables that an edge connects are defined with two object properties in the evidence ontology: pointsTo shows the variable that the edge is directed to, and comesFrom shows the variable that the edge is directed away.

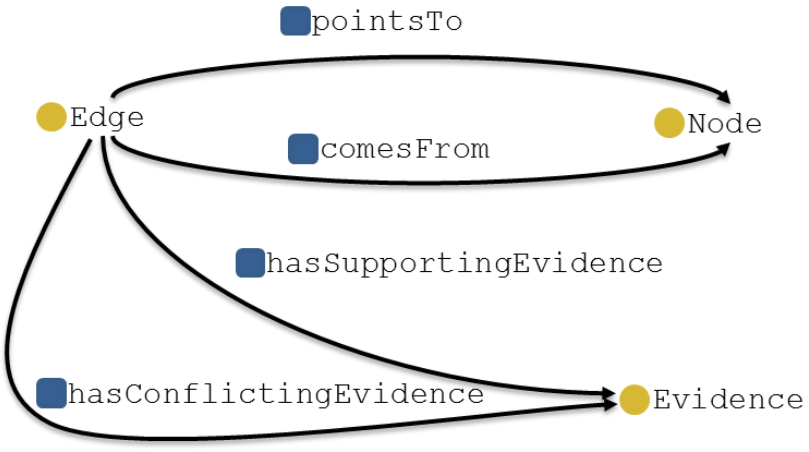


Figure 8.12 Object Properties related to Edge Class

Defining the direction of edges using pointsTo and comesFrom properties is sufficient for the ontology to infer the other object properties about the relation between two variables. For example, in order to define the directed relation A → B

between two variables A and B, we define an individual for the edge class, e.g. Edge1, and we define that Edge1 pointsTo B and comesFrom A (see Figure 8.13). The evidence ontology is able to infer the other object properties related to the relation between A and B (shown by dashed arrows in Figure 8.13) by using the following characteristics:

1. The inverse of the comesFrom property is the hasOutgoingEdge property. By using the inverse property, the ontology can infer that A hasOutgoingEdge Edge1 when it is defined that Edge1 comesFrom A. Similarly, the inverse of the pointsTo property is the hasIncomingEdge property.

2. The hasParent property can be inferred from the combination of the hasIncomingEdge and comesFrom properties. For example, if the variable B hasIncomingEdge Edge1 and Edge1 comesFrom the variable A, then the ontology can infer that the variable B hasParent A. Similarly, the ontology can infer the hasChild property from the combination of the hasOutgoingEdge and pointsTo properties.

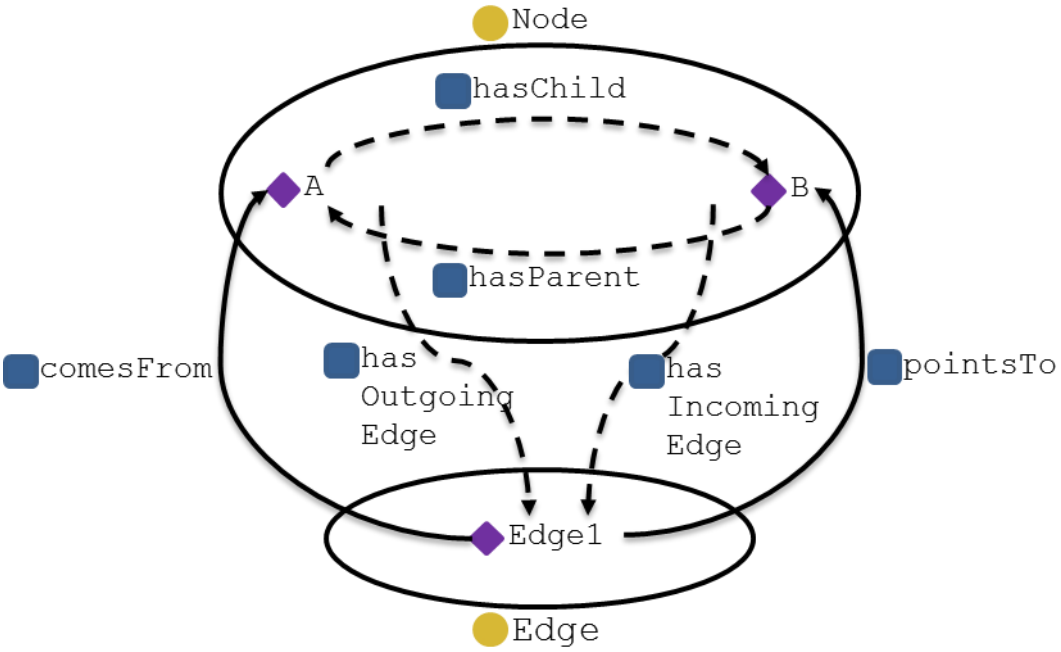


Figure 8.13 *A* → *B* modelled in Evidence Ontology

Each relation modelled in a BN should have supporting evidence in an evidence-based BN (see Figure 8.12). When the BN structure is built with domain experts,

evidence for the relations often comes from expert knowledge or scientific publications. The justification for including the relation is recorded as the statement of evidence. In purely data-driven BNs, the evidence comes from the dataset only. It may be difficult to describe knowledge supporting the relations in data-driven models apart from saying that the variables were correlated in the data (see Section 3.2 and 3.3). Conflicting evidence may also exist for a relation in the BN. For example, one publication may claim that two variables are independent, whereas another publication may find correlation between those variables. If these variables are connected by an edge in the BN structure, in accordance with the latter publication, the conflicting evidence from the former publication must also be recorded in order to have a comprehensive evidence base.

The meaning of a modelled relation is shown by its supporting evidence. In order to include an edge in a BN model, the BN developer must have evidence that the relation exist between two variables. Therefore, all edges in a BN must have supporting evidence in an evidence-based model, which may come from domain experts, publications or data. An edge does not require a separate description like a variable since evidence associated with the edge describes its meaning.

8.2.2.2 Evidence

Figure 8.14 shows the properties related to the evidence class. The evidence class organises the data about the type and statement of evidence related to the BN elements. There are 3 types of evidence in our ontology:

1. **Supporting Evidence:** This class contains the statement of evidence that supports the elements of a BN model.

2. **Conflicting Evidence:** Evidence in this class conflicts with the BN model. For example, two variables that are found to be independent in a scientific study may not be modelled as independent in the BN model. In this case, the results of the scientific study must be recorded as conflicting evidence in the evidence ontology. Conflicting expert opinions may also be recorded as conflicting evidence.

3. **Excluded BN Element:** Evidence about relevant variables or relations that are not included in the BN model is recorded within this class. The statement of evidence relevant to a excluded BN element and the justification for not including the element should be recorded. The excluded BN element class has 3 subclasses related to the type of the BN element excluded. The excluded child (parent) classes indicate that a variable in the BN may have an effect (cause) that is not modelled in the BN. The excluded relation class indicates that a direct relation may exist between two variables that are not directly linked in the BN.

Each type of evidence is a subclass of the evidence class. The subclasses can be inferred from the type of object property that refers to the evidence. For example, any evidence recorded by the hasSupportingEvidence property is inferred as a member of the SupportingEvidence class. Similarly, any evidence that is recorded by the hasConflictingEvidence or hasExcludedBNElement properties are inferred as the members of the ConflictingEvidence or ExludedBNElement classes respectively.

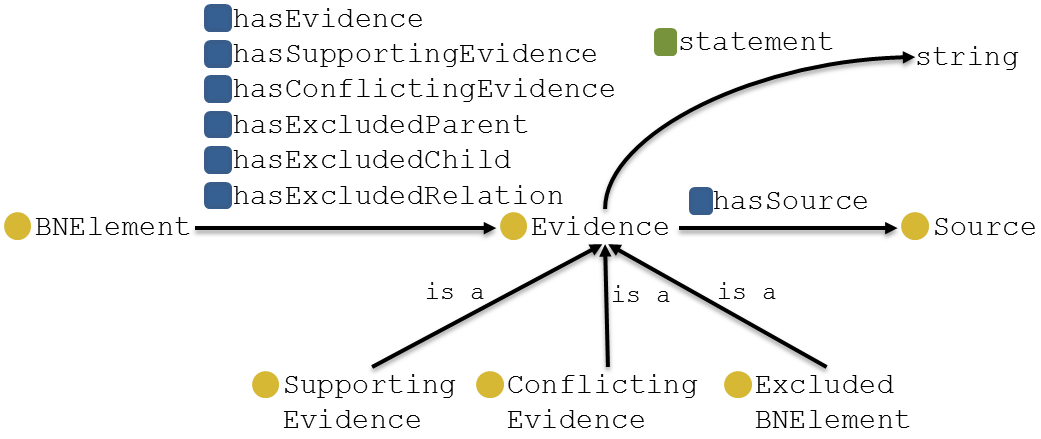


Figure 8.14 Object and Data Properties related to Evidence Class

A BN element can have multiple items of evidence, and an item of evidence can have multiple sources. For example, two publications that are stating similar results about a relation in the BN can be recorded as a single evidence item with two sources (publications). Two publications that are discussing different aspects of a relation can be recorded as two separate items of evidence with one source each. More examples about recording evidence and source are shown in Section 8.2.3. The following section describes the ontology class about the source of evidence.

8.2.2.3 Source

The source class contains information about the publication, expert opinion or data providing evidence. The subclasses, object and data properties of the source class are shown in Figure 8.15. The source class has three subclasses as shown below:

1. **Publication:** The members of this class are scientific publications. Information for referring these publications, such as digital object identifier (DOI) and PubMed identification number (PMID), can be stored in this class. A PMID is a unique identifier assigned that makes it convenient to find the publications in the PubMed database.

2. **Expert:** The domain experts' credentials and contact information can be recorded in this class.

3. **Data:** Evidence supporting or conflicting with model may come from data. The individuals of this class contain information about the details of the dataset including the sample size and the method of collecting the data.

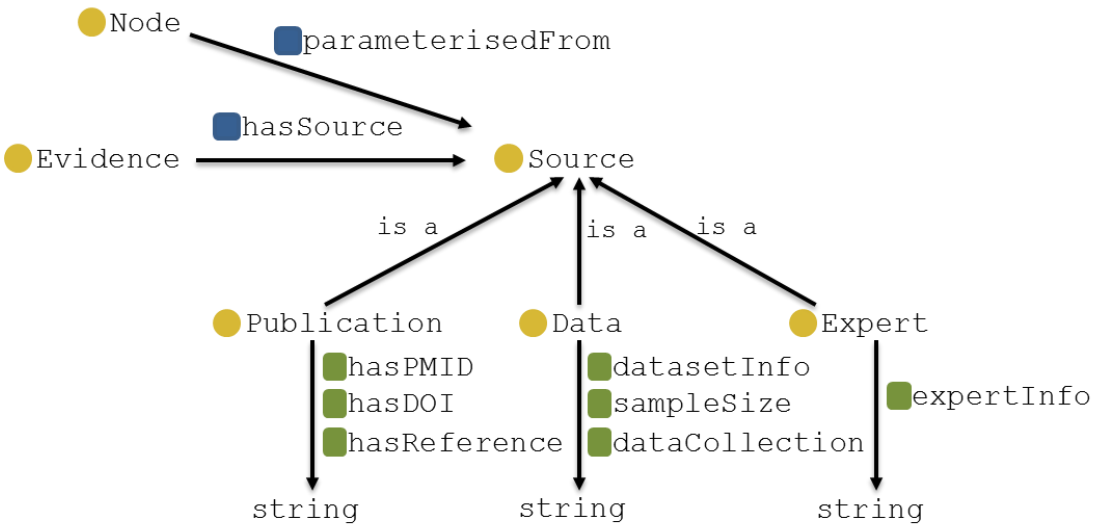


Figure 8.15 Object and Data Properties related to Source Class

8.2.3 Entering Data to Evidence Ontology

In this section, we illustrate how data is entered to the evidence ontology by using a simplified version of the ATC BN that has 3 variables and 2 edges (see Figure 8.16).

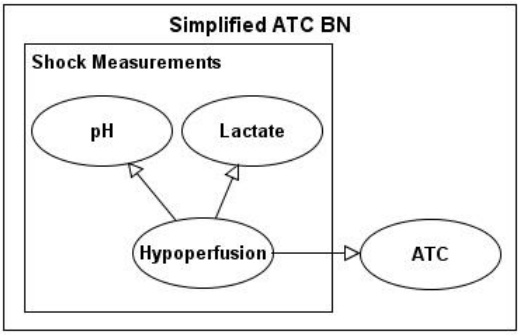


Figure 8.16 Simplified ATC BN

8.2.3.1 Fragment

The simplified ATC BN in Figure 8.16 contains a BN fragment called shock measurements. The state of hypoperfusion variable cannot be directly observed in clinical practice; it is estimated by several tests and observations including the lactate and pH levels in blood (see Chapter 5 for the description of the ATC BN). The shock measurements fragment models the relation between hypoperfusion and its measurements.

Table 8.1 Defining Shock Fragment

ShockMeasurements	
Type: BNElement Fragment MeasurementIdiom	Object Properties: within SimplifiedATCBN contains Hypoperfusion contains Lactate contains pH Data Properties: description "This part of the model estimates…"

Table 8.1 shows the classes, object and data properties related to the shock measurements fragment. The structure of the shock measurements fragment is modelled using a pre-defined BN structure called a measurement idiom (see Sections 5.3.3 and 8.2.2). We use the object property 'contains' to show the variables and fragments within this fragment, and 'within' to show the larger fragments that contain this fragment. The entire BN is also defined as a BN fragment (SimplifiedATCBN) that contains all of the other variables and fragment in the

evidence ontology. A free-text description of the fragment is recorded using a data property.

8.2.3.2 Variable

Table 8.2 shows the data related to the Hypoperfusion variable in the evidence ontology. This variable is within the Shock fragment, and it has three children: ATC, Lactate and pH. The parameters of the variable are learnt from the RLH dataset. The description of the variable and the names of its states are recorded by the 'description' and 'stateName' data properties. The edges between the Hypoperfusion variable and its children are modelled by the 'hasOutgoingEdge' object property that shows the edges that is directed away from the Hypoperfusion variable. For example, 'HypoToLactate' is a member of the edge class that represents the edge Hypoperfusion → Lactate. The evidence supporting the relation between Hypoperfusion and Lactate variables is defined under the edge individual. More information about the edge class and its individuals are shown in Section 8.2.3.3.

Table 8.2 Defining Hypoperfusion Variable

Hypoperfusion	
Type: BNElement Variable	Object Properties: within Shock hasChild ATC hasChild Lactate hasChild pH hasOutgoingEdge HypoToATC hasOutgoingEdge HypoToLactate hasOutgoingEdge HypoToPH parameterisedFrom RLHDataset Data Properties: description "The degree of inadequate oxygen…" stateName "None" stateName "Compensated" stateName "Uncompensated"

Table 8.3 shows the information recorded about the lactate variable. Since lactate is a continuous variable, we define its upper and lower bounds instead of state names. There is evidence that some factors affecting lactate are not included in this model.

This information is recorded in an item of evidence named 'LactateExParEv1' belonging to the evidence class (see Section 8.2.3.4).

Table 8.3 Defining Lactate Variable

Lactate	
Type: BNElement Variable	Object Properties: within Shock hasParent Hypoperfusion hasIncommingEdge HypoToLactate hasExcludedParent LactateExParEv1 parameterisedFrom RLHDataset Data Properties: description "The amount of lactate in…" hasLowerBound 0 hasUpperBound 100

8.2.3.3 Edge

HypoToLactate represents the edge Hypoperfusion → Lactate in the simplified ATC BN (see Table 8.4). The variables connected by this edge are modelled with 'comesFrom' and 'pointsTo' properties. There are two items of evidence supporting the modelling of this edge.

Table 8.4 Defining the Hypoperfusion → Lactate Edge

HypoToLactate	
Type: BNElement Edge	Object Properties: comesFrom Hypoperfusion pointsTo Lactate hasSupportingEvidence LactateSuppEv1 hasSupportingEvidence LactateSuppEv2

8.2.3.4 Evidence

Two publications indicated that lactate is an important marker of the degree of hypoperfusion. We recorded this as an item of evidence supporting the edge Hypoperfusion → Lactate in the ATC BN (see Table 8.5). The details of the publications are described under the source class in Section 8.2.3.5.

Table 8.5 Supporting Evidence 1

LactateSuppEv1	
Type: Evidence SupportingEvidence	Object Properties: hasSource RixenEtAl2005 hasSource VanDrommeEtAl2010 Data Properties: statement "Lactate is produced during anaerobic…"

Hypoperfusion and lactate levels were highly correlated in our dataset. We recorded this correlation as a second item of evidence supporting the edge Hypoperfusion → Lactate (see Table 8.6).

Table 8.6 Supporting Evidence 2

LactateSuppEv2	
Type: Evidence SupportingEvidence	Object Properties: hasSource RLHDataset Data Properties: statement "The likelihood of hypoperfusion increases with increasing lactate values in the RLH dataset…"

One publication indicated that several factors, including excessive alcohol use, can affect lactate levels independent from the effects of hypoperfusion but these factors were not included in the ATC BN. We recorded this as evidence for a parent variable for Hypoperfusion that is not included in the BN (see Table 8.7).

Table 8.7 Evidence about Excluded Parent

LactateExParEv1	
Type: Evidence ExcludedBNElement ExcludedParent	Object Properties: hasSource RixenEtAl2005 Data Properties: statement "…excessive alcohol use can increase the lactate levels …"

8.2.3.5 Source

Table 8.8 shows the details of the publication related to the items of evidence in Table 8.5 and Table 8.7. The publication is recorded as a member of the source class with its referencing details and PMID.

Table 8.8 Publication Source

RixenEtAl2005	
Type: Source Publication	Data Properties: hasPMID 16277731 hasRefDetails "D. Rixen and J. H. Siegel, Critical Care, vol. 9, no. 5, p. 441, 2005 …"

Table 8.9 shows the details of the database related to the item of evidence in Table 8.6. We recorded the description, sample size and method of collecting the dataset.

Table 8.9 Data Source

RLHDataset	
Type: Source Data	Data Properties: datasetInfo "A dataset of 600 patients who were treated at the…" dataCollection "Prospective Observational" sampleSize 600

8.2.4 Completeness Queries using SPARQL Query Language

The parts of a clinical BN that lacks evidence may be interesting to clinicians who review the BN. Therefore, an efficient way of assessing the completeness of evidence can be a useful feature for the evidence framework. SPARQL is a query language that can retrieve the data stored in an OWL ontology. In this section, we show several SPARQL queries to find the BN elements that have or lack evidence. It

is beyond the focus of this chapter to give a comprehensive description of SPARQL, a thorough introduction is given by Allemang & Hendler (2010).

The simplest form of query in SPARQL is the *SELECT* query that extracts the data from the ontology and presents them in a table format. *SELECT* query is followed by the *WHERE* block that limits the query by question pattern. For example, we can get a list of the edges that has supporting evidence by:

```
SELECT ?x
 WHERE {   ?x a :Edge.
           ?x :hasSupportingEvidence ?evidence.}
```

The first line of the *WHERE* construct above filters out the individuals that are not members of the Edge class. The second line filters out the individuals that do not have supporting evidence. When we apply this query to the evidence ontology of the simplified ATC BN (see Section 8.2.3), we get the following result:

```
:HypoToLactate
:HypoToLactate
```

Note that we get duplicate results since the edge Hypoperfusion ⟶ Lactate (named HypoToLactate in the ontology) has two items of supporting evidence (LactateSuppEv1 and LactateSuppEv2). We can use the *DISTINCT* keyword to filter out the duplicate results:

```
SELECT DISTINCT ?x
 WHERE {   ?x a :Edge.
           ?x :hasSupportingEvidence ?evidence.}
```

We get the following results after adding the *DISTINCT* keyword:

```
:HypoToLactate
```

The BN elements without evidence can be retrieved by using the *MINUS* keyword. The *UNSAID* keyword can be used as an alternative to *MINUS* keyword in SPARQL 1.1. The *MINUS* keyword is used within the *WHERE* construct.

```
SELECT DISTINCT ?x
 WHERE { ?x a :Edge.
   MINUS { ?x :hasSupportingEvidence ?evidence . } }
```

The SPARQL query above can be used to find the edges that do not have supporting evidence. The first line of the *WHERE* construct extracts all of the individuals that are edges. The second line filters out the individuals that have supporting evidence. For the simplified ATC BN example, it gives the following result:

```
:HypoToPH
:HypoToATC
```

The ASK query is another kind of SPARQL query that can answer Yes/No questions in the ontology. The answer for an ASK query is either 'True' or 'False'. For example, we can ask whether the Hypoperfusion → ATC edge has supporting evidence in the simplified ATC BN by:

```
ASK WHERE { atcbn:HypoToATC atcbn:hasSupportingEvidence
 ?evidence}
```

Since we haven't defined any evidence for this edge in Section 8.2.3, we get the following answer:

```
FALSE
```

8.3 Browsing Evidence

The evidence ontology is well suited for organising and querying evidence but it is not a convenient tool to browse evidence especially if the user is not proficient with the ontology language. Therefore, our ontology framework prepares a web page (HTML files) for browsing evidence after the data is entered to the evidence ontology. The web page is automatically generated from the ontology (a Protégé OWL file)[2].

The web page generator is not specific to a particular BN; it can generate the HTML files for browsing any BN model given that evidence is entered to Protégé OWL using the structure presented in Section 8.2.2. In the remainder of this section, we use the complete version of the ATC BN (see Section 5.3.3) as a case study to

[2] An early prototype of the web page generator was prepared as a part of an undergraduate graduate project at Queen Mary, University of London. The author continued the development of the browser and completed the current version.

demonstrate a web page generated from the evidence ontology. We call this web page the evidence browser for the ATC BN (see ATCBN, 2013).

The ATC BN has a total of 7 fragments (see Figure 8.17):

- The entire BN structure ('ATCBN')
- The mortality, injury, coagulopathy and shock fragments that are described in Section 5.3.3
- Two measurement idioms that models the relation between the latent variables and their measurements ('ShockMeasurements' and 'ATCMeasurements'). These measurement idioms fragments exist within the shock and coagulopathy fragments.

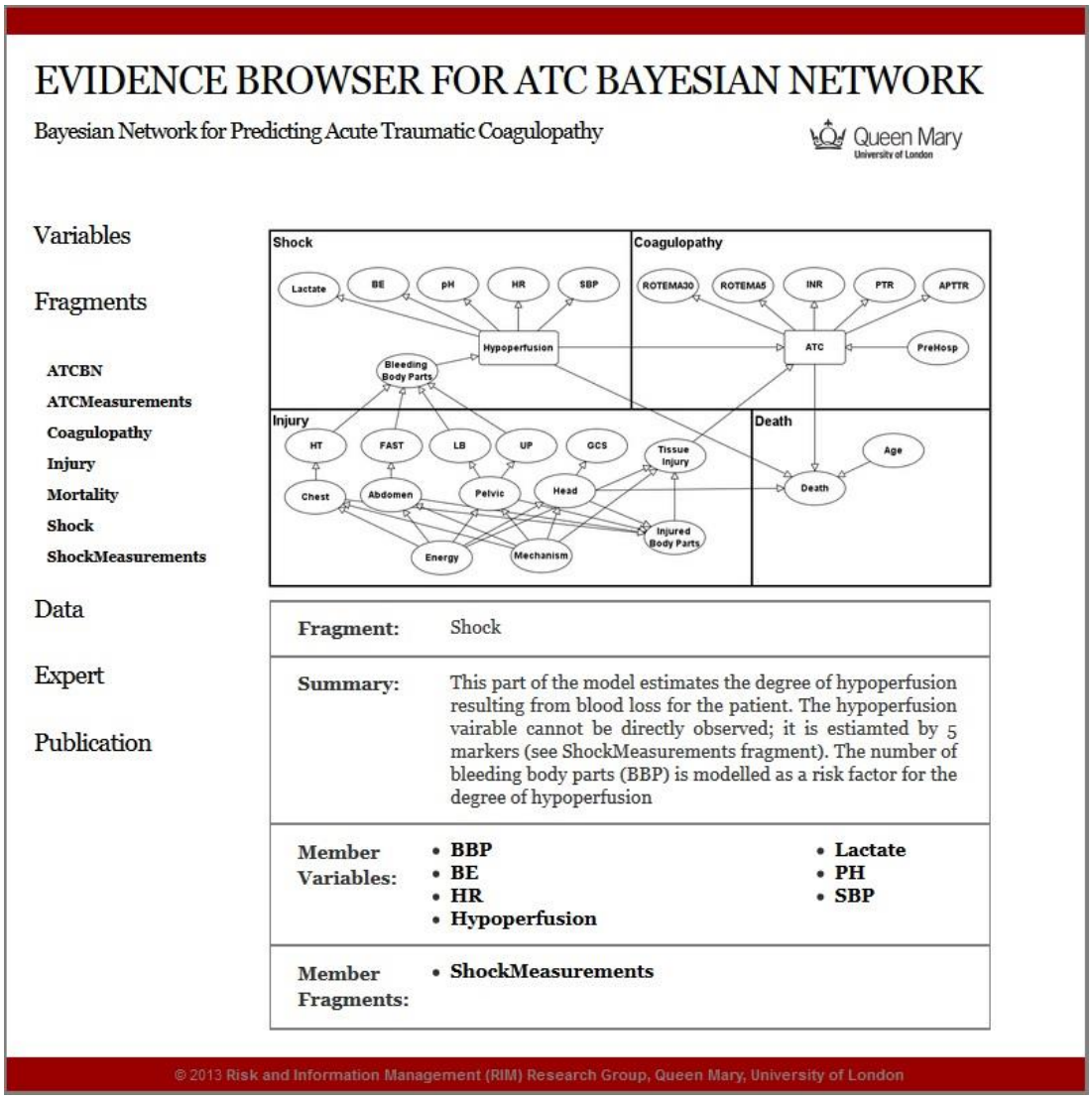


Figure 8.17 Shock Fragment shown in the Evidence Browser

Figure 8.17 is the presentation of the Shock fragment in the evidence browser. The browser shows a description of the fragment and links to other variables and fragments that are members of this fragment.

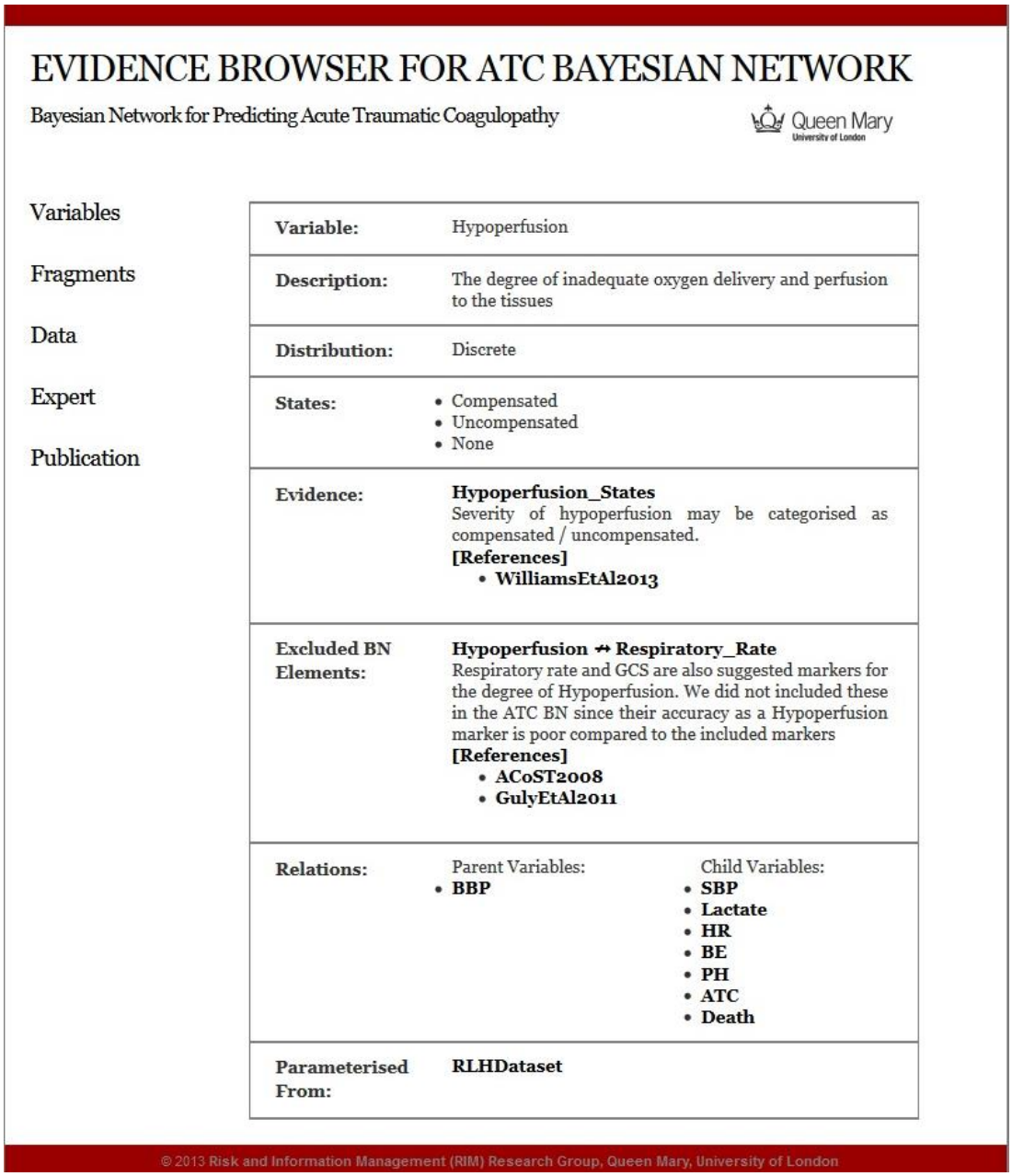


Figure 8.18 Hypoperfusion variable shown in the evidence browser

If the user clicks to the Hypoperfusion variable in Figure 8.17, the evidence browser shows the details of this variable as in Figure 8.18. The browser shows the description of the variable, type of its probability distribution, list of its states, and relevant evidence. In this example, there are two items of evidence relevant to the hypoperfusion variable. The first item, named 'Hypoperfusion States', shows

evidence for selecting the states of this variable. The second item, named 'Hypoperfusion ↛ Respiratory Rate', shows that the degree of Hypoperfusion can be also be measured with other markers that are not included in the ATC BN. These markers could be added as a child of the hypoperfusion variable. The reason for not adding these variables is also described.

The variables that are directly related to the hypoperfusion variable, as its child or parent, are also listed in Figure 8.18. Each variable in this list is linked to a page showing the evidence about the relation. For example, when the user clicks to ATC which is in the relations section in Figure 8.18, the evidence related to the Hypoperfusion ⟶ ATC relation is shown (see Figure 8.19).

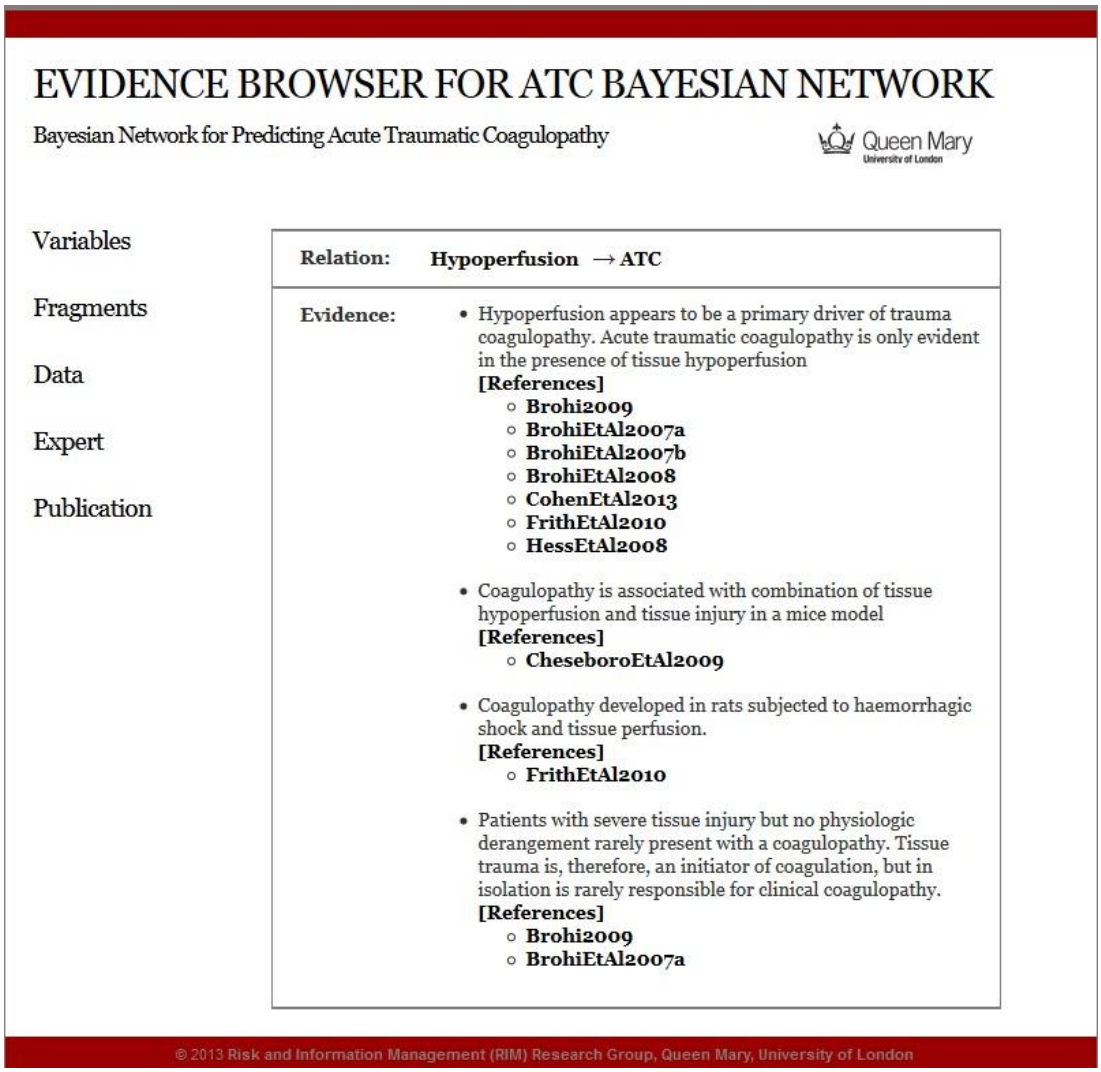


Figure 8.19 Relation between Hypoperfusion and ATC shown in the Evidence Browser

There are 4 items of evidence relevant to the relation Hypoperfusion ⟶ ATC (see Figure 8.19). The user can click any of the references in Figure 8.19 to examine the source of the evidence statements. For example, if the user clicks to BrohiEtAl2007a, the browser shows the details of this publication by connecting to the PubMed database (see Figure 8.20). The browser uses the PMID stored in our evidence ontology to find the publication in the PubMed database. If the source is a domain expert or a dataset, the browser shows the related information such as the institution and credentials of the expert or the description of the dataset.

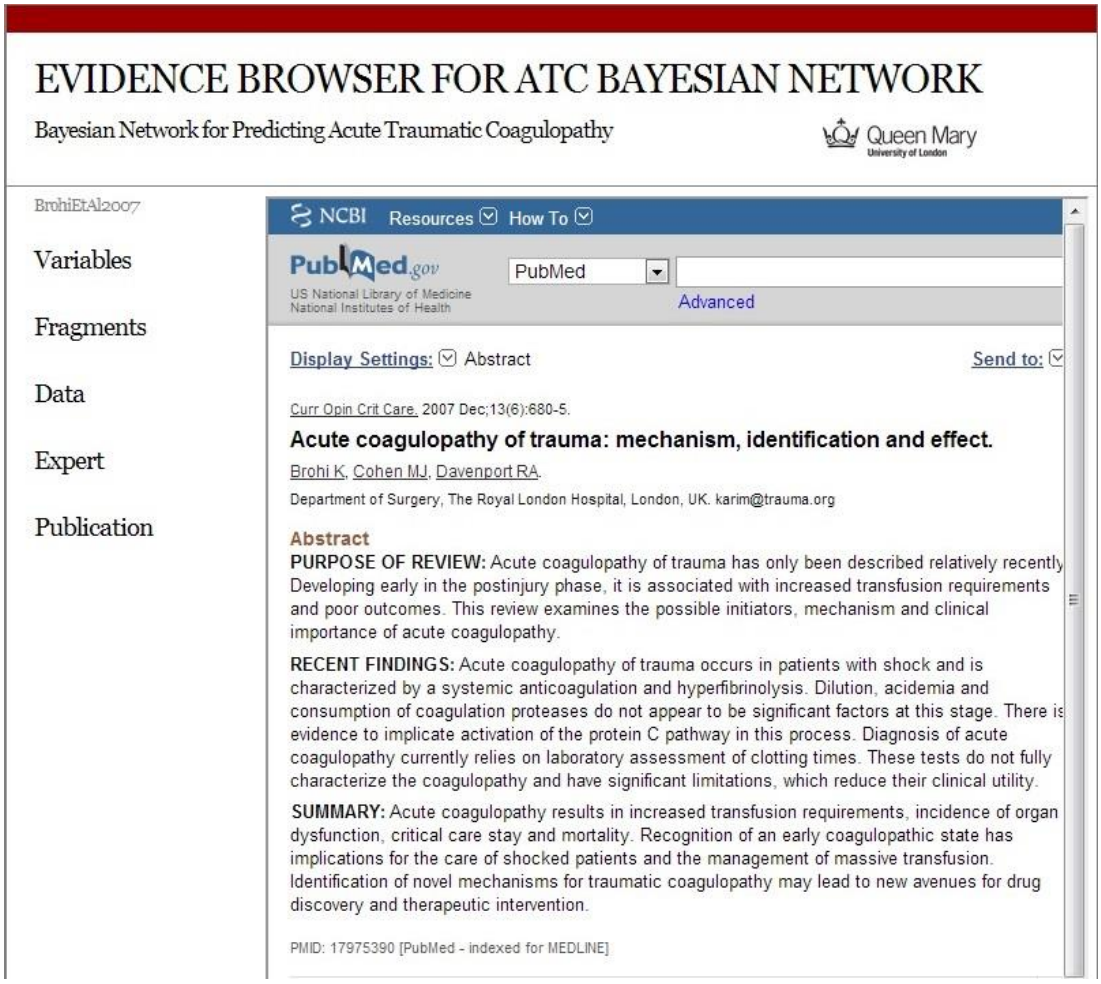


Figure 8.20 A Referred Publication shown in the Evidence Browser

8.4 Related Work

Several other frameworks have been proposed to describe knowledge behind a BN model. Antal et al. (2001) propose annotated BNs that describe the background knowledge for each node, edge and state in a BN by annotations. This representation

was primarily used for automated learning of BNs from the textual information in medical literature (Antal et al., 2004).

Helsper and van Der Gaag (2007) use ontologies to clarify the elicited knowledge used for building a BN. They use a semi-formal ontology language that includes tables, figures and natural language descriptions. Their ontology documents expert knowledge from static, dynamic and definitional perspectives. Causal and temporal relations are described using Rieger and Grinberg's representation (1977). Helsper and van der Gaag (2002) also propose a methodology to derive a BN structure from this ontology however the method was not used to develop the model in their case-study since the ontology is prepared retrospectively from an existing BN. Van der Gaag and Tabachneck-Schijf (2010) extend Helsper and van Der Gaag's ontology framework (2007) for describing BNs that are built for similar tasks.

While both Helsper and van Der Gaag's ontologies (2007) and Antal et al.'s annotated BNs (2001) aims to clarify knowledge behind BNs but these frameworks do not aim to show the link between the model and evidence. Helsper and van Der Gaag (2007) focus on describing knowledge elicited from experts in more detail. For example, their ontology has numerous types of causal and temporal relations that can distinguish between continuous and one-shot causal events. However, Helsper and van der Gaag do not show the link between the model and evidence. Knowledge for their ontology is elicited from experts. Evidence from publications or conflicting expert opinions is not shown. Annotated BN (Antal et al. 2004; 2001) lacks the structure for recording information regarding type and source of evidence which is a crucial feature for the clarity of an evidence-base.

Several studies have developed methodologies to use existing ontologies for automated construction of BNs. Devitt et al. (2006) propose a methodology for automated construction of a BN from ontologies in telecommunication networks domain. In this method, an ontology specific to the BN is derived from a more general ontology, and BN is automatically generated from the specific ontology. Similarly, Sadeghi et al. (2005) build an ontology more specific for the problem domain using the concepts from the unified medical language system (UMLS) (Bodenreider, 2004) and they learn a BN is based on this specific ontology. UMLS is a complicated medical terminology system but it lacks information about the causal

relations between clinical factors. Ishak et al. (2011) presents a set of rules that transform an ontology into a preliminary OOBN structure. Bucci et al. (2011) uses ontologies to build BNs on a predefined hierarchical structure in medical diagnosis domain. Fenz (2012) uses a semi-automated methodology to generate BNs from ontologies. In their methodology, the experts review the ontology and identify the nodes and states of the BN based on the classes and individuals in the ontology. Afterwards they identify the edges in the BN based on the object properties in the ontology. Fenz proposes a technique that uses weights defined in the ontology to parameterise NPTs in a similar approach to the parameterisation of ranked nodes (Fenton et al., 2007). An evidence-based BN requires justification for the relations modelled in the BN. Each edge in the BN should be supported with clinical studies, expert opinion or data. Many clinical ontologies are defined as clinical terminologies therefore they do not contain detailed information about causal and associational relations, and medical publications and datasets relevant to those relations. Therefore, the reviewed studies about automated BN construction from ontologies are not aligned with the aims of our evidence framework.

Another active field of research focuses on extending ontologies with BNs to represent uncertain knowledge. Ding and Peng (2004) propose additional mark-ups to OWL to represent probabilistic information. They present a set of rules that transforms ontology into a BN. Similarly, Yang and Calmet (2005) present a BN extension to OWL that is also able to cope with multinomial variables. Costa et al. (2008) extend OWL to express uncertainty using multi entity BNs (Laskey and Costa, 2005) which is a combination of first-order logic and Bayesian reasoning. Zheng et al. (2008) propose mark-ups to transform an ontology representing clinical concepts into a BN. The aims of these studies are also different from the aims of our evidence framework. These studies extend ontology languages, such as OWL, so that the ontologies are able to cope with uncertainties about class membership, object properties and their other features. The evidence framework uses ontology as a complementary tool to clarify knowledge and evidence behind a decision support model, and it does not require the ontology to deal with uncertainty.

8.5 Conclusion

This chapter proposed an evidence framework that complements clinical BNs by representing relevant knowledge and clinical evidence. The proposed framework consists of two parts: an ontology that stores evidence relevant to different elements of a BN, and a browser that presents the BN and evidence to clinicians. The ATC BN is used as a case-study to illustrate the evidence framework.

Our evidence framework is able to organise and present the evidence in more detail than it is possible with the existing BN representation or previously proposed annotation techniques. The evidence framework can store various types of evidence including evidence supporting or conflicting with the BN, as well as evidence justifying the exclusion of variables and relations from the BN. Although the evidence is stored within a complicated ontology structure, users can browse the evidence in a web page that is automatically generated from the ontology, without dealing with any of the underlying technical details.

The next step is to run a validation session with a group of clinicians to evaluate whether browsing evidence can improve their understanding of a BN model and its applicability in clinical practice. The evidence framework can also be improved to receive comments and suggestions from clinicians browsing the BN. Clinicians from different institutions can criticise evidence behind the BN and suggest modifications when new evidence becomes available.

The evidence ontology could be proposed as a standardised format for recording BN models. This would require the ontology to store information about NPTs. Another option could be to extend the currently available XML formats for representing BNs (Cozman, 1998) to include information about evidence using the structure presented in this chapter. The evidence framework could also be extended to match its data with the definitions from medical terminology systems such as UMLS and SNOMED (Bodenreider, 2004; Spackman et al., 1997). This would make it possible to have a universally consistent terminology for the evidence recorded in the framework. However, synchronising the evidence framework with the terminology systems is challenging due to the complexity of these systems.

Finally, the evidence framework could be integrated with the abstraction methodology presented in Chapter 7. Such integration would enable the abstraction methodology to show how each simplification in the BN structure relates to underlying evidence supporting the model.

Chapter 9

Summary and Future Directions

This chapter summarises the contributions of the thesis in three topics: combining evidence, assisting BN development and understanding evidence. The chapter also discusses the future directions of research.

9.1 Combining Evidence

One of the challenges of applying EBM for individual patients is to combine the relevant evidence about different aspects of the treatment from multiple sources. In this thesis, we proposed methodologies to develop BN models that combine evidence from previous publications, data and expert opinion based on clinical knowledge about the domain. The proposed methodologies contribute to EBM as they can provide decision support by combining evidence from multiple sources based on clinical knowledge. We claim novelty for the methodologies and BN models presented in Chapters 5 – 8.

Chapter 5 proposed a methodology for building decision support models that are consistent with clinical knowledge without being limited by the availability of data. The proposed methodology systematically used a combination of knowledge and data to develop and refine BN models. In this methodology, domain experts define the BN structure that explicitly models the important clinical variables even when they are not a part of the available data. The expert knowledge and data is used at two different stages: 1) to understand and model the behaviour of the latent variables, 2) to identify the potential improvements and refine the BN accordingly. We demonstrated the success of this methodology by using it for the development of

the ATC BN, which accurately predicts ATC, a fatal physiological disorder, in early stages of trauma care.

The challenges that Chapter 5 address are not specific to the trauma case study. In order to build a BN model that is consistent with domain knowledge, variables that are important for expressing causal knowledge must be included in the model whether or not they are a part of the available data. Our methodology allows inclusion of these variables into decision support as long as domain knowledge is provided by experts. Moreover, we proposed a systematic way of identifying the potential model improvements. Our approach can be applied to other domains where all important factors cannot be directly observed but domain knowledge and some data is available.

Chapter 6 proposed a methodology for building decision support models based on the evidence provided by meta-analyses. This methodology combines the evidence from a meta-analysis with expert knowledge and data to define the structure and parameters of a BN. This chapter also proposed a novel auxiliary Bayesian parameter learning model that combines data with the probabilities, from a meta-analysis, that are conditioned on fewer variables than the parameter being learned. We demonstrate the success of this methodology by applying it to a BN for predicting the short-term outcomes of a traumatic lower extremity with vascular injury. The results of our methodology are better than the results of purely data-based structure and parameter learning approaches, and an existing decision support model.

Both Chapters 5 and 6 proposed methodologies to combine the evidence from different sources for providing decision support. Yet, these methodologies address different aspects of this problem. Chapter 5 combined expert opinion and data for discovering knowledge about important variables that cannot be directly observed. Chapter 6 addressed the case where the data for some combinations of factors, in this case injuries, were small and inadequate to build a decision support model. This chapter combined evidence from systematic reviews and meta-analysis with data for these rare combinations. The methodologies proposed in Chapters 5 and 6 can be used together when, for example, a decision problem has both latent variables and inadequate data for some combinations of factors. The case-study in Chapter 6

contained several latent variables that were modelled by using ranked nodes. These latent variables could also be modelled with the methodology proposed in Chapter 5 instead of ranked nodes.

9.2 Assisting BN Development

Chapter 7 proposed a knowledge engineering methodology of simplifying a BN structure without losing the link to clinical knowledge supporting the BN structure. This chapter presented a sufficient set of operations that simplify a BN by merging or removing nodes, removing edges and collapsing states. Our methodology shows how each abstraction operation changes knowledge encoded in the BN. This chapter also proposed a graphical notation that shows the link between the initial and abstracted models by capturing the sequence and type of the abstraction operations applied. The proposed knowledge engineering problems addresses the problems of making ad-hoc simplifications in the BN structure without considering their effects to the BN's knowledge-base.

9.3 Understanding Evidence

Chapter 8 proposed a framework to organise and present evidence behind clinical BNs. The evidence framework can represent evidence related to different variables, relations and fragments of a BN. It also provides a detailed description of the BN elements and evidence to clarify the BN for other clinicians. The evidence framework is composed of an ontology that organises evidence, and a browser that presents evidence without showing the technical details of the ontology. The evidence framework is able to represent evidence in a more detailed and structured way than it is possible with the structure of a BN or with making annotations on the structure.

Both Chapters 7 and 8 aims to clarify knowledge encoded in a BN to its users. The methodology proposed in Chapter 7 addresses this issue at the stage of BN development. The methodology shows how simplifications of the BN structure affect knowledge behind the model. The evidence framework in Chapter 8 focuses on the

application stage. The framework aims to clarify knowledge and evidence behind a BN model that has already been developed and is being used. The evidence framework could complement the abstraction methodology at the development stage of a model. For example, while the abstraction methodology shows how encoded knowledge changes after a simplification, the evidence framework can enhance this by showing evidence relevant to these changes.

9.4 Future Directions

This thesis has showed that BNs can contribute to EBM by being used as a 1) evidence repository for organising evidence about clinical problems 2) risk calculator that quantifies uncertainties and calculates risks based on encoded evidence. The potential benefits for EBM can only be utilised if evidence-based BN models can be easily built by any BN developer who has access to clinical evidence. A BN developer must know the basic properties of BNs and relevant knowledge engineering methodologies in order to build a BN that correctly represents clinical knowledge and evidence. This thesis has contributed to the latter of these issues by proposing knowledge engineering methodologies for developing evidence-based BNs. BN developers also require tools that assist them in using correct modelling techniques and that provide them with state-of-the-art algorithms. Fenton and Neil's recent BN textbook (2013), and some commercial BN tools including AgenaRisk, has addressed some of these issues by focussing on the practical use of BNs and handling calculations details by the algorithms implemented in software. However, some techniques for BN development cannot be widely used in practical applications as they are not implemented in user friendly BN tools, whereas most traditional statistical techniques are readily implemented in software that is convenient and familiar to the medical community. More BN resources, textbooks and software must be targeted to practical applications in order to make BNs a mainstream modelling approach in clinical care. Further development of the ideas presented in this thesis could contribute to a wider use of BNs in clinical care.

The abstraction methodology proposed in Chapter 7 requires graphical operations with complicated calculations. The use of this methodology in practical applications could be improved if it could be implemented in an interactive development

environment (IDE) that handles the graphical operations. The IDE could assist a BN user in developing knowledge-based models and making abstractions without making undesirable changes in knowledge encoded in these models. Moreover, the IDE could show the equivalent ways of making an abstraction and recommend the option that adds the least amount of parameters to the BN. The result of each abstraction could be presented both graphically and textually. The evidence framework (see Chapter 8) could be integrated with the abstraction IDE to show evidence related to abstracted BN elements. A more systematic approach to knowledge-based modelling can be achieved by implementing the abstraction IDE to major BN software.

In Chapters 5 and 6, we proposed methodologies to combine data with knowledge and published studies to define the parameters of a BN. Data could also be integrated with other sources of evidence to build the BN structure. For example, it could be used to identify the parts of the BN that require more attention from the domain experts. The domain experts could review the BN structure more systematically by focusing on the edges that has weak correlations, and the unconnected variables that have strong correlations in the data. Although hybrid structure learning methods that combine knowledge with data exists, these methods are aligned for using expert knowledge to assist a data-driven learning. However, it could be beneficial for an evidence-based BN to use expert knowledge as the primary source of information for the BN structure, and the data as a way of reviewing the structure.

The evidence framework proposed in Chapter 8 could be expanded as a communication medium that aims to organise and disseminate state-of-the-art clinical evidence. The expanded framework could use a causal BN structure to give an overview of evidence about a particular medical topic. For example, a causal structure can show an overview of how various background factors, treatments and comorbidities are related with a disease. Since clinicians have limited amount of time to identify relevant evidence for individual patients, summarising evidence in a graphical BN structure could be beneficial for employing EBM in daily clinical practice (see Alper et al., 2004; Smith, 2013 for some examples of the challenges of identifying relevant evidence in daily practice). The meaning of each factor in the causal structure and the link with the publications would be provided by the evidence

framework. The expanded framework would also allow experts to suggest refinements on the causal structure when new evidence becomes available. Consensus methods such as Delphi could be used to update the causal structure when multiple expert opinions are available. Since the aim of the causal structure would be to give an overview of evidence, it does not have to be a complete – parameterised – BN. However, organising evidence in causal structures would make it easier to derive evidence-based decision support models from them. If causal knowledge about a subject is well documented, a BN developer would only need to focus on defining the parameters.

The BNs proposed in this paper provide predictions that are useful for the decision makers in trauma care. A useful next step would be to explain the predictions generated by the BN based on clinical evidence. For example, instead of plainly showing the probability of death, explaining why the model has calculated such probability of death can be more helpful to make a better risk assessment. Several techniques are available for explaining BN predictions by showing the influences and reasoning pathways of observed variables. By coupling the explanation techniques with the evidence framework, it could be possible to show how different items of evidence support a particular prediction of the model.

References

Abu-Hanna, A., Lucas, P.J.F., 2001. Prognostic models in medicine. Methods of Information in Medicine-Methodik der Information in der Medizin 40, 1–5.

Agena Ltd, 2013. AgenaRisk: Bayesian Network and Simulation Software for Risk Analysis and Decision Support [Online]. URL http://www.agenarisk.com (accessed 12.12.13).

Ahmed, B.A., Matheny, M.E., Rice, P.L., Clarke, J.R., Ogunyemi, O.I., 2009. A comparison of methods for assessing penetrating trauma on retrospective multi-center data. J Biomed Inform 42, 308–316.

Akhtar, J.I., Forse, R.A., 2010. Prognostic models: Are these models health fortune-telling tools?. Critical care medicine 38, 1605–1606.

Allemang, D., Hendler, J., 2011. Semantic web for the working ontologist: effective modeling in RDFS and OWL. Morgan Kaufmann.

Alper, B.S., Hand, J.A., Elliott, S.G., Kinkade, S., Hauan, M.J., Onion, D.K., Sklar, B.M., 2004. How much effort is needed to keep up with the literature relevant for primary care? J Med Libr Assoc 92, 429–437.

Altman, D.G., Vergouwe, Y., Royston, P., Moons, K.G., 2009. Prognosis and prognostic research: validating a prognostic model. BMJ: British Medical Journal 338, 1432–1435.

Alvarez, S.M., Poelstra, B.A., Burd, R.S., 2006. Evaluation of a Bayesian decision network for diagnosing pyloric stenosis. Journal of Pediatric Surgery 41, 155–161.

Andreassen, S., Jensen, F.V., Andersen, S.K., Falck, B., Kjærulff, U., Woldbye, M., Sørensen, A.R., Rosenfalck, A., Jensen, F., 1989. MUNIN—an expert EMG assistant. Computer-aided electromyography and expert systems 2, 255–277.

Antal, P., Fannes, G., Timmerman, D., Moreau, Y., De Moor, B., 2004. Using literature and data to learn Bayesian networks as clinical models of ovarian tumors. Artificial Intelligence in medicine 30, 257–281.

Antal, P., Mészáros, T., De Moor, B., Dobrowiecki, T., 2001. Annotated Bayesian Networks: a tool to integrate textual and probabilistic medical knowledge, in: Computer-Based Medical Systems, 2001. CBMS 2001. Proceedings. 14th IEEE Symposium On. pp. 177–182.

ATCBN, 2013. Acute Traumatic Coagulopathy Bayesian Network Evidence Browser [Online]. URL http://www.eecs.qmul.ac.uk/~barbaros/ATC_EBase/ (accessed 12.12.13) .

Attias, H., 1999. Inferring parameters and structure of latent variable models by variational Bayes, in: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence. pp. 21–30.

Baker, S.P., o' Neill, B., Haddon Jr, W., Long, W.B., others, 1974. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma 14, 187–196.

Bearfield, G., Marsh, W., 2005. Generalising event trees using bayesian networks with a case study of train derailment, in: Computer Safety, Reliability, and Security. Springer, pp. 52–66.

Beinlich, I.A., Suermondt, H.J., Chavez, R.M., Cooper, G.F., 1989. The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. Lecture Notes in Medical Informatics, 38, 247–256.Springer.

Berners-Lee, T., Hendler, J., Lassila, O., 2001. The semantic web. Scientific american 284, 28–37.

Bodenreider, O., 2004. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research 32, D267–D270.

Bonanni, F., Rhodes, M., Lucke, J.F., 1993. The futility of predictive scoring of mangled lower extremities. The Journal of Trauma and Acute Care Surgery 34, 99–104.

Bosse, M.J., MacKenzie, E.J., Kellam, J.F., Burgess, A.R., Webb, L.X., Swiontkowski, M.F., Sanders, R.W., Jones, A.L., McAndrew, M.P., Patterson, B.M., others, 2001. A prospective evaluation of the clinical utility of the lower-extremity injury-severity scores. The Journal of Bone & Joint Surgery 83, 3–14.

Bosse, M.J., McCarthy, M.L., Jones, A.L., Webb, L.X., Sims, S.H., Sanders, R.W., MacKenzie, E.J., 2005. The insensate foot following severe lower extremity trauma: an indication for amputation? The Journal of Bone & Joint Surgery 87, 2601–2608.

Boyd, C.R., Tolson, M.A., Copes, W.S., 1987. Evaluating trauma care: the TRISS method. The Journal of Trauma and Acute Care Surgery 27, 370–378.

Brier, G.W., 1950. Verification of forecasts expressed in terms of probability. Monthly weather review 78, 1–3.

Brohi, K., 2011. Prediction of acute traumatic coagulopathy and massive transfusion – Is this the best we can do? Resuscitation 82, 1128–1129.

Brohi, K., Singh, J., Heron, M., Coats, T., 2003. Acute traumatic coagulopathy. The Journal of trauma 54, 1127–1130.

Brotman, D.J., Walker, E., Lauer, M.S., O'Brien, R.G., 2005. In search of fewer independent risk factors. Archives of internal medicine 165, 138.

Bucci, G., Sandrucci, V., Vicario, E., 2011. Ontologies and Bayesian networks in medical diagnosis, in: System Sciences (HICSS), 2011 44th Hawaii International Conference On. pp. 1–8.

Buchan, I., Winn, J., Bishop, C., 2009. A unified modelling approach to data intensive healthcare, in: The Fourth Paradigm: Data-intensive Scientific Discovery. Microsoft Research, Redmond, WA, pp. 91–97.

Buchanan, B.G., Shortliffe, E.H., 1984. Rule Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project (The Addison-Wesley series in artificial intelligence). Addison-Wesley Longman Publishing Co., Inc.

Burnside, E.S., Rubin, D.L., Fine, J.P., Shachter, R.D., Sisney, G.A., Leung, W.K., 2006. Bayesian Network to Predict Breast Cancer Risk of Mammographic Microcalcifications and Reduce Number of Benign Biopsy Results: Initial Experience1. Radiology 240, 666–673.

Cano, A., Masegosa, A.R., Moral, S., 2011. A method for integrating expert knowledge when learning bayesian networks from data. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 41, 1382–1394.

Casscells, W., Schoenberger, A., Graboys, T.B., 1978. Interpretation by physicians of clinical laboratory results. The New England Journal of Medicine 299, 999.

Castelo, R., Kocka, T., 2003. On inclusion-driven learning of Bayesian networks. The Journal of Machine Learning Research 4, 527–574.

Celi, L.A., Hinske, L.C., Alterovitz, G., Szolovits, P., 2008. An artificial intelligence tool to predict fluid requirement in the intensive care unit: a proof-of-concept study. Crit Care 12, R151.

Champion, H.R., Sacco, W.J., Carnazzo, A.J., Copes, W., Fouty, W.J., 1981. Trauma score. Critical Care Medicine 9, 672–676.

Champion, H.R., Sacco, W.J., Copes, W.S., Gann, D.S., Gennarelli, T.A., Flanagan, M.E., 1989. A revision of the Trauma Score. The Journal of Trauma and Acute Care Surgery 29, 623–629.

Chang, K.C., Fung, R.M., 1990. Refinement and coarsening of Bayesian networks, in: Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence UAI'90. Elsevier Science Inc., Cambridge, MA, pp. 435–446.

Charitos, T., van der Gaag, L.C., Visscher, S., Schurink, K.A., Lucas, P.J., 2009. A dynamic Bayesian network for diagnosing ventilator-associated pneumonia in ICU patients. Expert Systems with Applications 36, 1249–1258.

Cheng, J., Greiner, R., Kelly, J., Bell, D., Liu, W., 2002. Learning Bayesian networks from data: An information-theory based approach. Artificial Intelligence 137, 43–90.

Chickering, D.M., 1996. Learning equivalence classes of Bayesian network structures, in: Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence. pp. 150–157.

Chickering, D.M., 2003. Optimal structure identification with greedy search. The Journal of Machine Learning Research 3, 507–554.

Chickering, D.M., Meek, C., 2002. Finding optimal bayesian networks, in: Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence. pp. 94–102.

Choi, A., Darwiche, A., 2006. An edge deletion semantics for belief propagation and its practical impact on approximation quality, in: Proceedings of the National Conference on Artificial Intelligence. p. 1107.

Civil, I.D., Schwab, C.W., 1988. The Abbreviated Injury Scale, 1985 revision: a condensed chart for clinical use. The Journal of trauma 28, 87.

Constantinou, A.C., Fenton, N.E., Neil, M., 2012. pi-football: A Bayesian network model for forecasting Association Football match outcomes. Knowledge-Based Systems 36, 322–339..

Constantinou, A.C., Fenton, N.E., Neil, M., 2013. Profiting from an inefficient association football gambling market: Prediction, risk and uncertainty using Bayesian networks. Knowledge-Based Systems 50, 60–86.

Cooper, G.F., 1997. A Simple Constraint-Based Algorithm for Efficiently Mining Observational Databases for Causal Relationships. Data Mining and Knowledge Discovery 1, 203–224.

Cooper, G.F., Herskovits, E., 1992. A Bayesian method for the induction of probabilistic networks from data. Machine learning 9, 309–347.

Coupé, V.M., Peek, N., Ottenkamp, J., Habbema, J.D.F., 1999. Using sensitivity analysis for efficient quantification of a belief network. Artificial Intelligence in Medicine 17, 223–247.

Coupé, V.M., Van Der Gaag, L.C., Habbema, J.D.F., 2000. Sensitivity analysis: an aid for belief-network quantification. Knowledge Engineering Review 15, 215–232.

Cozman, F.G., 1998. The Interchange Format for Bayesian Networks [Online]. URL http://http://www.cs.cmu.edu/~fgcozman/Research/InterchangeFormat/ (accessed 12.12.13).

Cristancho, S.M., Vanstone, M., Lingard, L., LeBel, M.-E., Ott, M., 2013. When surgeons face intraoperative challenges: a naturalistic model of surgical decision making. The American Journal of Surgery 205, 156–162.

Crump, C., Silvers, C.T., Wilson, B., Schlachta-Fairchild, L., Lingley-Papadopoulos, C.A., Ashley, J.S., 2011. Using Principal Component Analysis to Aid Bayesian Network Development for Prediction of Critical Care Patient Outcomes. The Journal of Trauma: Injury, Infection, and Critical Care 71, 1841–1849.

Cruz-Ramírez, N., Acosta-Mesa, H.-G., Barrientos-Martínez, R.-E., Nava-Fernández, L.-A., 2006. How good are the Bayesian information criterion and the minimum description length principle for model selection? A Bayesian network analysis, in: MICAI 2006: Advances in Artificial Intelligence. Springer, pp. 494–504.

Cruz-Ramírez, N., Acosta-Mesa, H.G., Carrillo-Calvet, H., Alonso Nava-Fernández, L., Barrientos-Martínez, R.E., 2007. Diagnosis of breast cancer using Bayesian networks: A case study. Computers in Biology and Medicine 37, 1553–1564.

Da Costa, P.C.G., Laskey, K.B., Laskey, K.J., 2008. PR-OWL: A Bayesian ontology language for the semantic web, in: Uncertainty Reasoning for the Semantic Web I. Springer, pp. 88–107.

Daly, R., Shen, Q., Aitken, S., 2011. Learning Bayesian networks: approaches and issues. Knowledge Engineering Review 26, 99–157.

Dash, D., Druzdzel, M.J., 2003. Robust independence testing for constraint-based learning of causal structure, in: Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc., Acapulco, Mexico, pp. 167–174.

Davenport, R., Manson, J., De'Ath, H., Platton, S., Coates, A., Allard, S., Hart, D., Pearse, R., Pasi, K.J., MacCallum, P., others, 2011. Functional Definition and Characterisation of Acute Traumatic Coagulopathy. Critical care medicine 39, 2652–2658.

Devitt, A., Danev, B., Matusikova, K., 2006. Ontology-driven Automatic Construction of Bayesian Networks for Telecommunication Network Management, in: 2nd Int. Workshop: Formal Ontologies Meet Industry (FOMI 2006).

Dıez, F.J., Druzdzel, M.J., 2006. Canonical probabilistic models for knowledge engineering. Technical Report CISIAD-06-01, UNED, Madrid, Spain.

Díez, F.J., Mira, J., Iturralde, E., Zubillaga, S., 1997. DIAVAL, a Bayesian expert system for echocardiography. Artificial Intelligence in Medicine 10, 59–73.

Ding, Z., Peng, Y., 2004. A probabilistic extension to ontology language OWL, in: System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference On. p. 10–pp.

Druzdel, M.J., Van Der Gaag, L.C., 2000. Building probabilistic networks:" Where do the numbers come from?" IEEE Transactions on knowledge and data engineering 12, 481–486.

Druzdzel, M.J., Díez, F.J., 2003. Combining knowledge from different sources in causal probabilistic models. The Journal of Machine Learning Research 4, 295–316.

Druzdzel, M.J., Flynn, R.R., 2009. Decision Support Systems, in: Encyclopedia of Library and Information Sciences, Third Edition. Taylor & Francis, pp. 1458–1466.

Durham, R.M., Mistry, B.M., Mazuski, J.E., Shapiro, M., Jacobs, D., 1996. Outcome and utility of scoring systems in the management of the mangled extremity. The American journal of surgery 172, 569–574.

Egger, M., Smith, G.D., Schneider, M., 2001. Systematic reviews of observational studies. Systematic Reviews in Health Care: Meta-Analysis in Context, Second Edition 211–227.

Elidan, G., Friedman, N., 2006. Learning hidden variable networks: The information bottleneck approach. Journal of Machine Learning Research 6, 81–127.

Evidence-Based Medicine Working Group, 1992. Evidence-based medicine. A new approach to teaching the practice of medicine. Jama 268, 2420–2425.

Feelders, A., Van der Gaag, L.C., 2006. Learning Bayesian network parameters under order constraints. International Journal of Approximate Reasoning 42, 37–53.

Fenton, N., 2011. Science and law: Improve statistics in court. Nature 479, 36–37.

Fenton, N., 2012. A simple story illustrating why pure machine learning (without expert input) may be doomed to fail and totally unnecessary [Online] URL http://www.eecs.qmul.ac.uk/~norman/papers/ml_simple_example.pdf (accessed 12.12.13).

Fenton, N., Neil, M., 2010. Comparing risks of alternative medical diagnosis using Bayesian arguments. Journal of Biomedical Informatics 43, 485–495.

Fenton, N., Neil, M., 2012a. Chapter 8 - Building and Eliciting Node Probability Tables, in: Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press.

Fenton, N., Neil, M., 2012b. Chapter 9 - Numeric Variables and Continous Distributions, in: Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press.

Fenton, N., Neil, M., 2012c. Chapter 2 - The Need for Causal Explanatory Models in Risk Assessment, in: Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press.

Fenton, N., Neil, M., Lagnado, D.A., 2013. A General Structure for Legal Arguments About Evidence Using Bayesian Networks. Cognitive Science 37, 61–102.

Fenton, N.E., Neil, M., Caballero, J.G., 2007. Using ranked nodes to model qualitative judgments in Bayesian networks. Knowledge and Data Engineering, IEEE Transactions on 19, 1420–1432.

Fenz, S., 2012. An ontology-based approach for constructing Bayesian networks. Data & Knowledge Engineering 73, 73–88.

Flin, R., Youngson, G., Yule, S., 2007. How do surgeons make intraoperative decisions? Qual Saf Health Care 16, 235–239.

Flores, J.M., Nicholson, A.E., Brunskill, A., Korb, K.B., Mascaro, S., 2011. Incorporating expert knowledge when learning Bayesian network structure: A medical case study. Artificial intelligence in medicine 53, 181–204.

Friedman, N., 1997. Learning Belief Networks in the Presence of Missing Values and Hidden Variables, in: Proceedings of the Fourteenth International Conference on Machine Learning. Morgan Kaufmann Publishers, pp. 125–133.

Friedman, N., 1998. The Bayesian structural EM algorithm, in: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence. pp. 129–138.

Friedman, N., 2004. Inferring Cellular Networks Using Probabilistic Graphical Models. Science 303, 799–805.

Frith, D., Goslings, J., Gaarder, C., Maegele, M., Cohen, M., Allard, S., Johansson, P., Stanworth, S., Thiemermann, C., Brohi, K., 2010. Definition and drivers of acute traumatic coagulopathy: clinical and experimental investigations. Journal of Thrombosis and Haemostasis 8, 1919–1925.

Gabbe, B.J., Cameron, P.A., Finch, C.F., 2003. Is the Revised Trauma Score Still Useful? ANZ Journal of Surgery 73, 944–948.

Gelman, A., 2008. Objections to Bayesian statistics. Bayesian Analysis 3, 445–449.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical science 457–472.

Gennarelli, T.A., Wodzin, E., 2008. Abbreviated Injury Scale 2005: Update 2008. Association for the Advancement of Automative Medicine.

Gigerenzer, G., 2003. Reckoning with risk: learning to live with uncertainty. Penguin UK.

Glover, F., Laguna, M., 1997. Tabu search. Encyclopaedia of Operations Research and Management Science, 671-701.Springer.

Goodman, S.N., 1999. Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy. Ann Intern Med 130, 995–1004.

Gregory, R.T., Gould, R.J., Peclet, M., Wagner, J.S., Gilbert, D.A., Wheeler, J.R., Snyder, S.O., Gayle, R.G., Schwab, C.W., 1985. The mangled extremity syndrome (MES): a severity grading system for multisystem injury of the extremity. The Journal of trauma 25, 1147.

Guyatt, G., Cook, D., Haynes, B., 2004. Evidence based medicine has come a long way: the second decade will be as exciting as the first. BMJ: British Medical Journal 329, 990.

Guyatt, G.H., Oxman, A.D., Vist, G.E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., Schünemann, H.J., 2008. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 336, 924–926.

Hadorn, D.C., Baker, D., Hodges, J.S., Hicks, N., 1996. Rating the quality of evidence for clinical practice guidelines. Journal of Clinical Epidemiology 49, 749–754.

Harbour, R., Miller, J., 2001. A new system for grading recommendations in evidence based guidelines. BMJ 323, 334–336.

Harris, I.A., Young, J.M., Rae, H., Jalaludin, B.B., Solomon, M.J., 2008. Predictors of general health after major trauma. The Journal of Trauma and Acute Care Surgery 64, 969–974.

Hanna, G., Phillips, L., Priest, O., Ni, Z., 2010. Improving the safety of nasogastric feeding tube insertion [Online]. URL:http://www.birmingham.ac.uk/Documents/

college-mds/haps/projects/cfhep/psrp/finalreports/PS048Improvingthesafetyofnasogastricfeedingtubeinsertion REVISEDHannaetal.pdf. (accessed 12.12.13)

Haynes, R.B., 1993. Where's the meat in clinical journals. ACP Journal club 119, A22–A23.

Haynes, R.B., Devereaux, P.J., Guyatt, G.H., 2002. Physicians' and patients' choices in evidence based practice: Evidence does not make decisions, people do. BMJ: British Medical Journal 324, 1350.

Heckerman, D., 1990. Probabilistic similarity networks. Networks 20, 607–636.

Heckerman, D., Geiger, D., Chickering, D.M., 1995. Learning Bayesian networks: The combination of knowledge and statistical data. Machine learning 20, 197–243.

Heckerman, D.E., Horvitz, E.J., Nathwani, B.N., 1989. The Pathfinder System, in: Proceedings of the Annual Symposium on Computer Application in Medical Care. p. 203.

Heckerman, D.E., Nathwani, B.N., 1992. Toward normative expert systems: Part II. Probability-based representations for efficient knowledge acquisition and inference. Methods of information in medicine 31, 106–116.

Helsper, E.M., van der Gaag, L.C., 2002. Building Bayesian networks through ontologies, in: ECAI. p. 15th.

Helsper, E.M., van Der Gaag, L.C., 2007. Ontologies for probabilistic networks: a case study in the oesophageal-cancer domain. The Knowledge Engineering Review 22, 67–86.

Helsper, E.M., Van Der Gaag, L.C., Feelders, A.J., Loeffen, W.L.A., Geenen, P.L., Elbers, A.R.W., 2005. Bringing order into bayesian-network construction, in: Proceedings of the 3rd International Conference on Knowledge Capture. pp. 121–128.

Henrion, M., 1987. Some practical issues in constructing belief networks, in: Proceedings of the Third Conference on Uncertainty in Artificial Intelligence UAI'87. Elsevier Science, Amsterdam, Netherlands, pp. 161–173.

Higgins, J., Thompson, S.G., Spiegelhalter, D.J., 2009. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society) 172, 137–159.

Hill, A.B., 1951. The clinical trial. British Medical Bulletin 7, 278–282.

Himes, B.E., Dai, Y., Kohane, I.S., Weiss, S.T., Ramoni, M.F., 2009. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. J Am Med Inform Assoc 16, 371–379.

Holmes, J.H., Peek, N., 2007. Intelligent data analysis in biomedicine. Journal of biomedical informatics 40, 605–608.

Honaker, J., Gary King, Blackwell, M., 2013. Amelia II: A Program for Missing Data.

Horton, R., 2000. Common sense and figures: the rhetoric of validity in medicine (Bradford Hill Memorial Lecture 1999). Statistics in medicine 19, 3149–3164.

Hosmer, D.W., Lemeshow, S., 1980. Goodness of fit tests for the multiple logistic regression model. Communications in statistics. Theory and methods 9, 1043–1069.

Howe Jr, H.R., Poole Jr, G.V., Hansen, K.J., Clark, T., Plonk, G.W., Koman, L.A., Pennell, T.C., 1987. Salvage of lower extremities following combined orthopedic and vascular trauma. A predictive salvage index. The American surgeon 53, 205.

Ishak, M.B., Leray, P., Amor, N.B., 2011. Ontology-based generation of object oriented Bayesian networks, in: Proceedings of the 8th Bayesian Modelling Applications Workshop. pp. 9–17.

Jacklin, R., Sevdalis, N., Darzi, A., Vincent, C., 2008. Mapping surgical practice decision making: an interview study to evaluate decisions in surgical care. The American Journal of Surgery 195, 689–696.

Jenks, S., Volkers, N., 1992. Razors and Refrigerators and Reindeer—Oh My! Journal of the National Cancer Institute 84, 1863–1863.

Johansen, K.A.J., Daines, M., Howey, T., Helfet, D., Hansen Jr, S.T., 1990. Objective criteria accurately predict amputation following lower extremity trauma. The Journal of Trauma and Acute Care Surgery 30, 568–573.

Kahneman, D., 2011. Thinking, fast and slow. Farrar, Straus and Giroux.

Katz, M.H., 2003. Multivariable analysis: a primer for readers of medical research. Ann Intern Med 138, 644–650.

Khan, O.Z., Poupart, P., John-mark, M.A., 2011. Automated Refinement of Bayes Networks' Parameters based on Test Ordering Constraints, in: Proceedings of the 25th Annual ConferenceAdvances in Neural Information Processing Systems. pp. 2591–2599.

Klein, G., 2008. Naturalistic decision making. Human Factors: The Journal of the Human Factors and Ergonomics Society 50, 456–460.

Klein, G.A., Calderwood, R., Clinton-Cirocco, A., 1986. Rapid decision making on the fire ground, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. pp. 576–580.

Klein, G.A., Orasanu, J.E., Calderwood, R.E., Zsambok, C.E., 1993. Decision making in action: Models and methods., in: This Book Is an Outcome of a Workshop Held in Dayton, OH, Sep 25–27, 1989.

Kline, J.A., Novobilski, A.J., Kabrhel, C., Richman, P.B., Courtney, D.M., 2005. Derivation and Validation of a Bayesian Network to Predict Pretest Probability of Venous Thromboembolism. Annals of Emergency Medicine 45, 282–290.

Knublauch, H., Fergerson, R., Noy, N., Musen, M., 2004. The Protégé OWL plugin: An open development environment for semantic web applications. The Semantic Web–ISWC 2004 229–243.

Koller, D., Friedman, N., 2009a. Skill: Knowledge Engineering, in: Probabilistic Graphical Models Principles and Techniques. MIT Press, Cambridge (Mass.), pp. 64–68.

Koller, D., Friedman, N., 2009b. Probabilistic graphical models principles and techniques. MIT Press, Cambridge (Mass.).

Koller, D., Friedman, N., 2009c. Bayesian Parameter Estimation, in: Probabilistic Graphical Models Principles and Techniques. MIT Press, Cambridge (Mass.), pp. 733–754.

Koller, D., Friedman, N., 2009d. Bayesian Learning with Incomplete Data, in: Probabilistic Graphical Models Principles and Techniques. MIT Press, Cambridge (Mass.), pp. 897–908.

Koller, D., Pfeffer, A., 1997. Object-oriented Bayesian networks, in: Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc., Providence, RI, pp. 302–313.

Korb, K.B., Nicholson, A.E., 2004a. Modelling and Elicitation, in: Bayesian Artificial Intelligence, Computer Science and Data Analysis. Chapman & Hall/CRC, Boca Raton, pp. 241–247.

Korb, K.B., Nicholson, A.E., 2004b. Bayesian artificial intelligence. cRc Press.

Korompilias, A.V., Beris, A.E., Lykissas, M.G., Vekris, M.D., Kontogeorgakos, V.A., Soucacos, P.N., 2009. The mangled extremity and attempt for limb salvage. Journal of Orthopaedic Surgery and Research 4, 1–6.

Lacave, C., Díez, F.J., 2003. Knowledge Acquisition in PROSTANET – A Bayesian Network for Diagnosing Prostate Cancer, in: Palade, V., Howlett, R.J., Jain, L. (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 1345–1350.

Lam, W., Bacchus, F., 1994. Learning Bayesian belief networks: An approach based on the MDL principle. Computational intelligence 10, 269–293.

Lappenschaar, M., Hommersom, A., Lucas, P.J., Lagro, J., Visscher, S., 2013. Multilevel Bayesian networks for the analysis of hierarchical health care data. Artificial intelligence in medicine.

Laskey, K.B., Costa, P.C., 2005. Of klingons and starships: Bayesian logic for the 23rd century, in: Uncertainty in Artificial Intelligence: Proceedings of the Twenty-first Conference.

Laskey, K.B., Mahoney, S.., 1997. Network fragments: Representing knowledge for constructing probabilistic models, in: Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97). Morgan Kaufmann Publishers, Providence, RI, pp. 334–341.

Laskey, K.B., Mahoney, S.., 2000. Network Engineering for Agile Belief Network Models. IEEE Transactions on Knowledge and Data Engineering 12, 487–498.

Lauritzen, S.L., 1995. The EM algorithm for graphical association models with missing data. Computational Statistics & Data Analysis 19, 191–201.

Lauritzen, S.L., Spiegelhalter, D.J., 1988. Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society. Series B (Methodological) 157–224.

Lipshitz, R., Klein, G., Orasanu, J., Salas, E., 2001. Taking stock of naturalistic decision making. Journal of Behavioral Decision Making 14, 331–352.

Lucas, P.J., de Bruijn, N.C., Schurink, K., Hoepelman, A., 2000. A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU. Artificial Intelligence in medicine 19, 251–279.

Lucas, P.J., van der Gaag, L.C., Abu-Hanna, A., 2004. Bayesian networks in biomedicine and health-care. Artificial Intelligence in medicine 30, 201–214.

Lucas, P.J.F., 1995. Logic engineering in medicine. The Knowledge Engineering Review 10, 153–179.

Lucas, P.J.F., Boot, H., Taal, B., 1998. Computer-based decision support in the management of primary gastric non-Hodgkin lymphoma. Methods of Information in Medicine-Methodik der Information in der Medizin 37 (3), 206–219UU-CS.

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: Evolution, critique and future directions. Statistics in medicine 28, 3049–3067.

Ly, T.V., Travison, T.G., Castillo, R.C., Bosse, M.J., MacKenzie, E.J., 2008. Ability of lower-extremity injury severity scores to predict functional outcome after limb salvage. The Journal of Bone & Joint Surgery 90, 1738–1743.

MacKenzie, E.J., Bosse, M.J., Pollak, A.N., Webb, L.X., Swiontkowski, M.F., Kellam, J.F., Smith, D.G., Sanders, R.W., Jones, A.L., Starr, A.J., 2005. Long-term persistence of disability following severe lower-limb traumaResults of a seven-year follow-up. The Journal of Bone & Joint Surgery 87, 1801–1809.

Margaritis, D., 2003. Learning Bayesian network model structure from data. University of Pittsburgh [PhD Thesis].

Margaritis, D., 2004. Distribution-free learning of graphical model structure in continuous domains. Technical report TR-ISU-CS-04-06, Department of Computer Science, Iowa State University.

Marquez, D., Neil, M., Fenton, N., 2010. Improved reliability modeling using Bayesian networks and dynamic discretization. Reliability Engineering & System Safety 95, 412–425.

Marsh, W., Bearfield, G., 2004. Using Bayesian networks to model accident causation in the UK railway industry, in: International Conference on Probabilistic Safety Assessment and Management, PSAM7, Berlin.

Marshall, J.C., 2006. Surgical Decision-Making: Integrating Evidence, Inference, and Experience. Surgical Clinics of North America 86, 201–215.

Maskery, S.M., Hu, H., Hooke, J., Shriver, C.D., Liebman, M.N., 2008. A Bayesian derived network of breast pathology co-occurrence. Journal of Biomedical Informatics 41, 242–250.

McGeachie, M., Ramoni, R.L.B., Mychaleckyj, J.C., Furie, K.L., Dreyfuss, J.M., Liu, Y., Herrington, D., Guo, X., Lima, J.A., Post, W., Rotter, J.I., Rich, S., Sale, M., Ramoni, M.F., 2009. Integrative predictive model of coronary artery calcification in atherosclerosis. Circulation 120, 2448–2454.

McNamara, M.G., Heckman, J.D., Corley, F.G., 1994. Severe open fractures of the lower extremity: a retrospective evaluation of the Mangled Extremity Severity Score (MESS). Journal of orthopaedic trauma 8, 81–87.

Medlock, S., Ravelli, A.C.J., Tamminga, P., Mol, B.W.M., Abu-Hanna, A., 2011. Prediction of Mortality in Very Premature Infants: A Systematic Review of Prediction Models. PLoS ONE 6, e23441.

Meek, C., Heckerman, D., 1997. Structure and parameter learning for causal independence and causal interaction models, in: Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence. pp. 366–375.

Miller, R.A., Pople Jr, H.E., Myers, J.D., 1982. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. The New England journal of medicine 307, 468.

Mitra, B., Cameron, P.A., Mori, A., Maini, A., Fitzgerald, M., Paul, E., Street, A., 2011. Early prediction of acute traumatic coagulopathy. Resuscitation 82, 1208–1213.

Moons, K.G.M., Altman, D.G., Vergouwe, Y., Royston, P., 2009a. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ: British Medical Journal 338, 1487–1490.

Moons, K.G.M., Royston, P., Vergouwe, Y., Grobbee, D.E., Altman, D.G., 2009b. Prognosis and prognostic research: what, why, and how? BMJ 338, b375–b375.

Nadkarni, S., Shenoy, P.P., 2004. A causal mapping approach to constructing Bayesian networks. Decision Support Systems 38, 259–281.

Nathwani, B.N., Heckerman, D.E., Horvitz, E.J., Lincoln, T.L., 1990. Integrated expert systems and videodisc in surgical pathology: an overview. Human pathology 21, 11–27.

Needham, C.J., Bradford, J.R., Bulpitt, A.J., Westhead, D.R., 2007. A Primer on Learning in Bayesian Networks for Computational Biology. PLoS Comput Biol 3, e129.

Neil, M., Fenton, N.E., Nielsen, L., 2000. Building large-scale Bayesian networks. The Knowledge Engineering Review 15, 257–284.

Neil, M., Hager, D., Andersen, L.B., 2009. Modeling operational risk in financial institutions using hybrid dynamic Bayesian networks. Journal of Operational Risk 4, 3–33.

Neil, M., Tailor, M., Marquez, D., 2007. Inference in hybrid Bayesian networks using dynamic discretization. Statistics and Computing 17, 219–233.

Neumann, J., Fox, P.T., Turner, R., Lohmann, G., 2010. Learning partially directed functional networks from meta-analysis imaging data. Neuroimage 49, 1372–1384.

Nielsen, T.D., Jensen, F.V., 2007. Bayesian networks and decision graphs. Springer.

Nissan, A., Protic, M., Bilchik, A., Eberhardt, J., Peoples, G.E., Stojadinovic, A., 2010. Predictive model of outcome of targeted nodal assessment in colorectal cancer. Annals of surgery 251, 265–274.

Ni, Z., Phillips, L.D., Hanna, G.B., 2011. Evidence Synthesis Using Bayesian Belief Networks, in: Evidence Synthesis in Healthcare. Springer, pp. 155–168.

Nordmann, J.-P., Berdeaux, G., 2007. Use of a Bayesian Network to Predict the Nighttime Intraocular Pressure Peak from Daytime Measurements. Clinical Therapeutics 29, 1751–1760.

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T., 2006. Uncertain judgements: Eliciting experts' probabilities. Wiley.

Ogunsanya, O.V., 2012. Decision support using Bayesian networks for clinical decision making. Queen Mary, University of London [PhD Thesis].

Onisko, A., Druzdzel, M.J., Wasyluk, H., 1998. A probabilistic causal model for diagnosis of liver disorders, in: Proceedings of the Workshop Held in Malbork, Poland, Malbork, Poland.

Orsanu, J., Fischer, U., 1997. Finding decisions in natural environments- The view from the cockpit. Naturalistic decision making(A 97-39376 10-53), Mahwah, NJ, Lawrence Erlbaum Associates, Inc., Publishers, 1997, 343–357.

Patel, V.L., Shortliffe, E.H., Stefanelli, M., Szolovits, P., Berthold, M.R., Bellazzi, R., Abu-Hanna, A., 2009. The coming of age of artificial intelligence in medicine. Artificial intelligence in medicine 46, 5.

Patil, R.S., 1981. Causal representation of patient illness for electrolyte and acid-base diagnosis (PhD). MIT, Cambirdge, MA.

Patil, R.S., Szolovits, P., Schwartz, W.B., 1981. Causal understanding of patient illness in medical diagnosis, in: Proceedings of the Seventh International Joint Conference on Artificial Intelligence. Presented at the Seventh International Joint Conference on Artificial Intelligence, Vancouver, BC, pp. 893–899.

Pauley, K., Flin, R., Yule, S., Youngson, G., 2011. Surgeons' intraoperative decision making and risk management. The American Journal of Surgery 202, 375–381.

Pearl, J., 1988. Probabilistic reasoning in intelligent systems : networks of plausible inference. Morgan Kaufmann Publishers, San Francisco, Calif.

Pearl, J., 2000. Causality: models, reasoning and inference. Cambridge Univ Press.

Pearl, J., Verma, T.S., 1991. A theory of inferred causation. Studies in Logic and the Foundations of Mathematics 134, 791–811.

Perel, P., Prieto-Merino, D., Shakur, H., Clayton, T., Lecky, F., Bouamra, O., Russell, R., Faulkner, M., Steyerberg, E.W., Roberts, I., 2012. Predicting early death in patients with traumatic bleeding: development and validation of prognostic model. BMJ: British Medical Journal 345.

Perkins, Z., Glasgow, Tai, N., 2012. A systematic review of prognostic factors related to secondary amputation in patients with lower limb vascular trauma requiring surgical repair (No. CRD42012002720). PROSPERO-Database. Registration no: CRD42012002720 .

Perkins, Z., Yet, B., Glasgow, S., Brohi, K., Marsh, W., Tai, N., 2013. Early Prediction of Acute Traumatic Coagulopathy using Admission Clinical Variables. Shock 40, S25.

Pradhan, M., Provan, G.M., Middleton, B., Henrion, M., 1994. Knowledge Engineering for Large Belief Networks, in: Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence (UAI-94). Presented at the Tenth international conference on Uncertainty in artificial intelligence (UAI-94), Seattle, Washington, pp. 484–490.

R Core Development Team, 2013. R Statistical Software [Online]. URL:http://www.r-project.org (accessed 12.12.13).

Rawlins, M., 2008. De testimonio: on the evidence for decisions about the use of therapeutic interventions. Clinical Medicine 8, 579–588.

Renooij, S., 2001. Probability elicitation for belief networks: issues to consider. The Knowledge Engineering Review 16, 255–269.

Renooij, S., 2010. Bayesian network sensitivity to arc-removal, in: Proceedings of the Fifth European Workshop on Probabilistic Graphical Models. Presented at the Fifth European Workshop on Probabilistic Graphical Models, Helsinki, Finland, pp. 233–240.

Rieger, C., Grinberg, M., 1977. The declarative representation and procedural simulation of causality in physical mechanisms, in: Proceedings of the 5th International Joint Conference on Artificial intelligence-Volume 1. pp. 250–256.

Robertson, P.A., 1991. Prediction of amputation after severe lower limb trauma. J Bone Joint Surg Br 73, 816–818.

Royal College of Nursing, 2012. Royal College of Nursing Evidence to the NHS Pay Review Body 2013-14 (No. 004 336).

Royston, P., Moons, K.G.M., Altman, D.G., Vergouwe, Y., 2009. Prognosis and prognostic research: Developing a prognostic model. BMJ 338, 1373–1377.

Russell, R.J., Hodgetts, T.J., McLeod, J., Starkey, K., Mahoney, P., Harrison, K., Bell, E., 2010. The role of trauma scoring in developing trauma clinical governance in the Defence Medical Services. Philosophical Transactions of the Royal Society B: Biological Sciences 366, 171–191.

Russell, W.L., Sailors, D.M., Whittle, T.B., Fisher Jr, D.F., Burns, R.P., 1991. Limb salvage versus traumatic amputation. A decision based on a seven-part predictive index. Annals of surgery 213, 473.

Sackett, D.L., Rosenberg, W.M., Gray, J.A., Haynes, R.B., Richardson, W.S., 1996. Evidence based medicine: what it is and what it isn't. BMJ: British Medical Journal 312, 71.

Sadeghi, S., Barzi, A., Smith, J.W., 2005. Ontology driven construction of a knowledgebase for Bayesian decision models based on UMLS. Studies in Health Technology and Informatics 116, 223–228.

Sakellaropoulos, G.C., Nikiforidis, G.C., 1999. Development of a Bayesian network for the prognosis of head injuries using graphical model selection techniques. Methods of Information in Medicine-Methodik der Information in der Medizin 38, 37–42.

Savage, N., 2012. Better medicine through machine learning. Communications of the ACM 55, 17–19.

Schurink, C.A., Visscher, S., Lucas, P.J., van Leeuwen, H.J., Buskens, E., Hoff, R.G., Hoepelman, A.I., Bonten, M.J., 2007. A Bayesian decision-support system for diagnosing ventilator-associated pneumonia. Intensive care medicine 33, 1379–1386.

Schurink, C.A.M., Lucas, P.J.F., Hoepelman, I.M., Bonten, M.J.M., 2005. Computer-assisted decision support for the diagnosis and treatment of infectious diseases in intensive care units. The Lancet infectious diseases 5, 305–312.

Schwarz, G., 1978. Estimating the dimension of a model. The annals of statistics 6, 461–464.

Scutari, M., 2010. Learning Bayesian networks with the bnlearn R package. Journal of Statistical Software 35, 1–29.

Segaran, T., Evans, C., Taylor, J., 2009. Programming the semantic web. O'Reilly Media.

Shachter, R.D., 1986. Evaluating influence diagrams. Operations research 34, 871–882.

Shachter, R.D., 1988. Probabilistic inference and influence diagrams. Operations Research 36, 589–604.

Shachter, R.D., 1990. Evidence Absorption and Propagation through Evidence Reversals, in: Proceedings of the Fifth Annual Conference on Uncertainty in Artificial Intelligence. Windsor, ON, pp. 173–190.

Shipp, M.A., Ross, K.N., Tamayo, P., Weng, A.P., Kutok, J.L., Aguiar, R.C.T., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G.S., Ray, T.S., Koval, M.A., Last, K.W., Norton, A., Lister, T.A., Mesirov, J., Neuberg, D.S., Lander, E.S., Aster, J.C., Golub, T.R., 2002. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 8, 68–74.

Shortliffe, E., 1976. Computer-based medical consultations: MYCIN. Elsevier.

Sierra, B., Larrañaga, P., 1998. Predicting survival in malignant skin melanoma using Bayesian networks automatically induced by genetic algorithms. An empirical comparison between different approaches. Artificial Intelligence in Medicine 14, 215–230.

Smith, G.C.S., Pell, J.P., 2003. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. BMJ 327, 1459–1461.

Smith, R., 2013. "10,000 more GPs needed" for NHS to cope with increasing workload [Online]. URL: http://www.telegraph.co.uk/health/healthnews/9873540/10000-more-GPs-needed-for-NHS-to-cope-with-increasing-workload.html (accessed 12.12.13).

Sox, H.C., Higgins, M.C., 1988. Medical decision making. Royal Society of Medicine.

Spackman, K.A., Campbell, K.E., CÃ, R., 1997. SNOMED RT: a reference terminology for health care., in: Proceedings of the AMIA Annual Fall Symposium. p. 640.

Spirtes, P., Glymour, C., Scheines, R., 2001. Causation, Prediction, and Search, 2nd Edition, 1st ed. The MIT Press.

Srinivas, S., 1994. A probabilistic approach to hierarchical model-based diagnosis, in: Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc., Seattle, WA, pp. 538–545.

Straus, S.E., Richardson, W.S., Glasziou, P., Haynes, R.B., 2005. Evidence-based medicine: how to practice and teach EBM.

Südkamp, N., Haas, N., Flory, P.J., Tscherne, H., Berger, A., 1989. Criteria for amputation, reconstruction and replantation of extremities in multiple trauma patients]. Der Chirurg; Zeitschrift für alle Gebiete der operativen Medizen 60, 774.

Sutton, A.J., Abrams, K.R., 2001. Bayesian methods in meta-analysis and evidence synthesis. Statistical Methods in Medical Research 10, 277–303.

Swiontkowski, M.F., Engelberg, R., Martin, D.P., Agel, J., 1999. Short Musculoskeletal Function Assessment Questionnaire: Validity, Reliability, and Responsiveness*. The Journal of Bone & Joint Surgery 81, 1245–60.

Swiontkowski, M.F., MacKenzie, E.J., Bosse, M.J., Jones, A.L., Travison, T., for the LEAP Study Group, 2002. Factors Influencing the Decision to Amputate or Reconstruct after High-Energy Lower Extremity Trauma. Journal of Trauma-Injury Infection 52, 641–649.

Teasdale, G., Jennett, B., 1974. Assessment of coma and impaired consciousness: a practical scale. The Lancet 304, 81–84.

Tonelli, M.R., 1999. In defense of expert opinion. Academic Medicine 74, 1187–92.

Tong, Y., Ji, Q., 2008. Learning Bayesian networks with qualitative constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008. pp. 1–8.

Tsamardinos, I., Brown, L.E., Aliferis, C.F., 2006. The max-min hill-climbing Bayesian network structure learning algorithm. Machine learning 65, 31–78.

Tversky, A., Kahneman, D., 1974. Heuristics and biases: Judgement under uncertainty. Science 185, 1124–1130.

Van Der Gaag, L.C., Bolt, J., Loeffen, W., Elbers, A., 2010. Modelling patterns of evidence in Bayesian networks: a case-study in classical swine fever, in: Computational Intelligence for Knowledge-Based Systems Design. Springer, pp. 675–684.

Van der Gaag, L.C., Renooij, S., Witteman, C.L.M., Aleman, B.M.P., Taal, B.G., 2002. Probabilities for a probabilistic network: a case study in oesophageal cancer. Artificial Intelligence in medicine 25, 123–148.

Van der Gaag, L.C., Tabachneck-Schijf, H.J.M., 2010. Library-style ontologies to support varying model views. International journal of approximate reasoning 51, 196–208.

Van der Heijden, M., Lucas, P.J., Lijnse, B., Heijdra, Y.F., Schermer, T.R., 2013. An Autonomous Mobile System for the Management of COPD. Journal of Biomedical Informatics, 46 (3), 458–469..

Van Engelen, R.A., 1997. Approximating Bayesian belief networks by arc removal. Pattern Analysis and Machine Intelligence, IEEE Transactions on 19, 916–920.

Velikova, M., Lucas, P.J., Samulski, M., Karssemeijer, N., 2013. On the interplay of machine learning and background knowledge in image interpretation by Bayesian networks. Artificial intelligence in medicine, 57 (1), 73–86..

Velikova, M., Samulski, M., Lucas, P.J.F., Karssemeijer, N., 2009. Improved mammographic CAD performance using multi-view information: a Bayesian network framework. Phys Med Biol 54, 1131–1147.

Verduijn, M., Rosseel, P.M., Peek, N., de Jonge, E., de Mol, B.A., 2007. Prognostic Bayesian networks II: an application in the domain of cardiac surgery. Prognostic Methods in Cardiac Surgery and Postoperative Intensive Care 33.

Vickers, A.J., 2006. Whose data set is it anyway? Sharing raw data from randomized trials. Trials 7, 15.

Visscher, S., Kruisheer, E.M., Schurink, C.A., Lucas, P.J., Bonten, M.J., 2008. Predicting pathogens causing ventilator-associated pneumonia using a Bayesian network model. Journal of antimicrobial chemotherapy 62, 184–188.

W3C, 2013. OWL 2 Web Ontology Language [Online]. URL: http://www.w3.org/TR/owl-features/ (accessed: 12/12/13)

Wallace, C., Korb, K.B., Dai, H., 1996. Causal discovery via MML, in: ICML. pp. 516–524.

Wallace, C.S., Korb, K.B., 1999. Learning linear causal models by MML sampling, in: Causal Models and Intelligent Data Management. Springer, pp. 89–111.

Wang, X.-H., Zheng, B., Good, W.F., King, J.L., Chang, Y.-H., 1999. Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network. International Journal of Medical Informatics 54, 115–126.

Ware Jr, J.E., Sherbourne, C.D., 1992. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Medical care 473–483.

Wasyluk, H., Onisko, A., Druzdzel, M.J., 2001. Support of diagnosis of liver disorders based on a causal Bayesian network model. Medical Science Monitor 7, 327–332.

Way, L.W., Stewart, L., Gantert, W., Liu, K., Lee, C.M., Whang, K., Hunter, J.G., 2003. Causes and Prevention of Laparoscopic Bile Duct Injuries: Analysis of 252 Cases From a Human Factors and Cognitive Psychology Perspective. Annals of Surgery 237, 460–469.

Wegener, S.T., Castillo, R.C., Haythornthwaite, J., MacKenzie, E.J., Bosse, M.J., 2011. Psychological distress mediates the effect of pain on function. Pain 152, 1349–1357.

Weigel, A.P., Liniger, M.A., Appenzeller, C., 2007. The discrete Brier and ranked probability skill scores. Monthly Weather Review 135, 118–124.

Weiss, S.M., Kulikowski, C.A., Amarel, S., Safir, A., 1978. A model-based method for computer-aided medical decision-making. Artificial intelligence 11, 145–172.

Wellman, M.P., Liu, C., 1994. State-space abstraction for anytime evaluation of probabilistic networks, in: Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence (UAI-94). Seattle, WA, pp. 567–574.

World Health Organization, 2012. WHO | Global tuberculosis report 2012 [Online]. WHO. URL http://www.who.int/tb/publications/global_report/en/ (accessed 12.12.13).

Wu, X., Poh, K.L., 2000. Decision-model construction with multilevel influence diagrams. The Knowledge Engineering Review 15, 233–256.

Wyatt, J.C., Altman, D.G., 1995. Commentary: Prognostic models: clinically useful or quickly forgotten? BMJ 311, 1539–1541.

Yang, Y., Calmet, J., 2005. OntoBayes: An Ontology-Driven Uncertainty Model, in: International Conference on Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce. Presented at the International Conference on Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, pp. 457–463.

Yet, B., Bastani, K., Raharjo, H., Lifvergren, S., Marsh, W., Bergman, B., 2013a. Decision support system for Warfarin therapy management using Bayesian networks. Decision Support Systems 55, 488–498.

Yet, B., Marsh, W., Perkins, Z., Tai, N., Fenton, N., 2013b. Predicting Latent Variables with Knowledge and Data: A Case Study in Trauma Care, in: Intelligence Application Workshops: Part I: Big Data Meet Complex Models. p. 49.

Yet, B., Perkins, Z., Marsh, W., Fenton, N., 2011. Towards a Method of Building Causal Bayesian Networks for Prognostic Decision Support. Probabilistic Problem Solving in BioMedicine 107–120.

Yet, B., Perkins, Z.B., Fenton, N.E., Tai, N.R., William R Marsh, D., 2013c. Not just data: A method for improving prediction with knowledge. Journal of biomedical informatics. DOI: 10.1016/j.jbi.2013.10.012.

Yet B., Marsh W., 2014. Compatible and Incompatible Abstractions in Bayesian Networks. Knowledge-Based Systems. DOI: 10.1016/j.knosys. 2014.02.020

Yu, V.L., Fagan, L.M., Wraith, S.M., Clancey, W.J., Scott, A.C., Hannigan, J., Blum, R.L., Buchanan, B.G., Cohen, S.N., 1979. Antimicrobial selection by a computer. A blinded evaluation by infectious diseases experts. JAMA: the journal of the American Medical Association 242, 1279.

Zhang, N.L., 2004. Hierarchical latent class models for cluster analysis. The Journal of Machine Learning Research 5, 697–723.

Zheng, H.-T., Kang, B.-Y., Kim, H.-G., 2008. An Ontology-Based Bayesian Network Approach for Representing Uncertainty in Clinical Practice Guidelines, in: Uncertainty Reasoning for the Semantic Web I: ISWC International Workshops, URSW 2005-2007, Revised Selected and Invited Papers. Springer-Verlag, pp. 161–173.

Zhou, Y., Fenton, N., Neil, M., 2013a. Bayesian Network Approach to Multinomial Parameter Learning using Data and Expert Judgements [Online]. URL: http://www.eecs.qmul.ac.uk/~yz300/publication/Yun-IJAR-2013.pdf (accessed 12.12.13).

Zhou, Y., Fenton, N., Neil, M., Zhu, C., 2013b. Incorporating Expert Judgement into Bayesian Network Machine Learning, in: Proceedings of the 23th International Joint Conference on Artificial Intelligence (IJCAI2013). Presented at the 23th

International Joint Conference on Artificial Intelligence (IJCAI2013), Beijing, China.

Ziliak, S.T., McCloskey, D.N., 2008. The cult of statistical significance: How the standard error costs us jobs, justice, and lives. University of Michigan Press.

Zupan, B., Holmes, J.H., Bellazzi, R., 2006. Knowledge-based data analysis and interpretation. Artificial Intelligence in Medicine 37, 163–165.