Bayesian network models for the management of ventilator-associated pneumonia

SIKS Dissertation Series No. 2008-23
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems and the University of Utrecht

The research described in this thesis has been funded by the Netherlands Organisation for Scientific Research (NWO grant number 634.000.026) under the ToKeN programme, as part of the TimeBayes project.

Design cover: AARGH! designers, Arnhem, Netherlands
Printing: PrintPartners Ipskamp B.V., Enschede, Netherlands

ISBN 978-90-393-4874-1

Bayesian network models for the management of ventilator-associated pneumonia

Bayesiaanse netwerkmodellen voor de diagnose en behandeling van beademingsgerelateerde longontsteking

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. J.C. Stoof,
ingevolge het besluit van het college voor promoties in het openbaar te
verdedigen op dinsdag 30 september 2008 des middags te 12.45 uur

door

Stefan Visscher

geboren op 31 maart 1980, te Utrecht

Promotor:
Prof.dr. M.J.M. Bonten

Co-promotoren:
Dr. P.J.F. Lucas
Dr. C.A.M. Schurink

Contents

[1]

General Introduction

Health care is a diverse, dynamic and rapidly developing area. Considerable progress has been made in both the biomedical sciences as well as in the fields of biomedical informatics during the past two decades, which is now increasingly affecting clinical practice. Medicine has, therefore, become a hugely complex area, – some say – beyond the full understanding of any human being.

Not even too long ago, decisions on how to treat patients were mainly based on authority, sometimes referred to as *eminence-based* medicine; nowadays, *evidence-based* medicine has become the standard, where decisions are based as much as possible on weighing all the, possibly contradictory, evidence available from scientific research. This approach offers a better foundation for objective decision-making in clinical practice than relying on authority. Evidence is derived from randomised controlled clinical trials, or from observational studies, frequently combined in meta-analyses.

Nowadays, computers with their associated information technology, play a central role in patient management and medical science. Their role in health care will only further increase with the widespread introduction of electronic patient records. For clinical medicine, this development offers great challenges and opportunities. In particular, computers may assist in better understanding disease processes and supporting clinicians in patient management. Clinical decision support and the analysis of patient data to increase our understanding of disease processes are the two central topics of this thesis. These topics will be briefly introduced in the subsequent sections.

1.1 Decision making in medicine

The art of medicine includes a considerable amount of decision making. For instance, multiple decisions are made when a patient is brought into the emergency room:

1. identifying the most likely cause of a disorder using a selection of tests, for example blood tests or medical imaging, in addition to history taking and physical exam which yields a *diagnosis*;
2. choosing the optimal, hopefully curative, *treatment* of the disorder of the patient;
3. determining the *prognosis* for the patient given the disorder and the selected treatment.

The entire process is visualised in Figure 1.1. This figure illustrates that a treatment choice is not only based on the results of the diagnostic process, but also on the possible effects of particular treatment options.

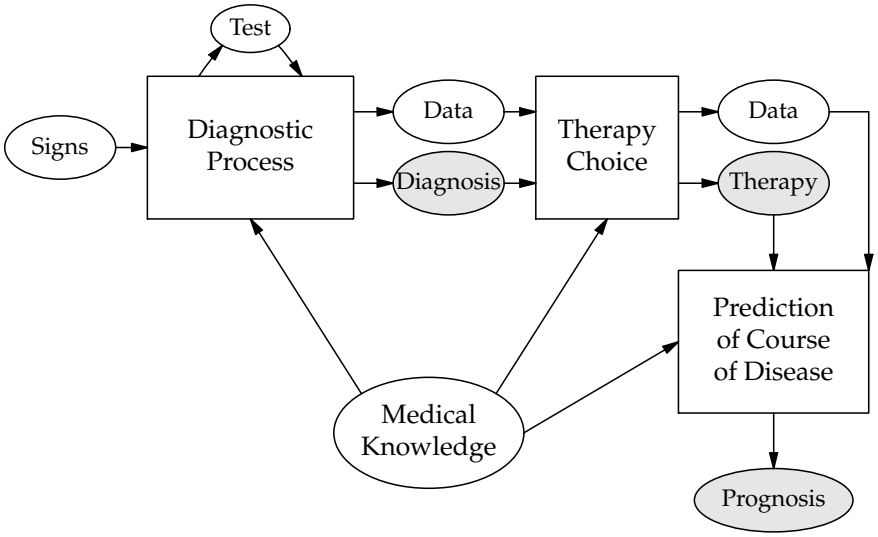


Figure 1.1: Overview of the various subprocesses of clinical decision making.

Establishing a diagnosis can be simple and fast, depending on the underlying disease and the time between a test and the availability of its results, called *diagnostic delay*. There may also be some time between presentation of the disease and the actual moment that a physician orders the right diagnostic tests, which is called the *physicians delay*. For completeness, the time between disease onset and the moment that a patient seeks medical advice is called *patient delay*. Problems in diagnosis frequently relate to the characteristics – sensitivity and specificity – of a test. Ideally, both sensitivity (probability of a positive test result given that the disease is present) and specificity (probability of a negative test result in absence of the disease) are 100%. More important and informative for physicians are the positive predictive value (PPV: probability of presence of the disease, given a positive test result) and negative predictive value (NPV: probability of absence of the disease, given a negative test result) [1].

In contrast to sensitivity and specificity, PPV and NPV also depend on the prevalence of the disease in the population tested. Unfortunately, no tests have these optimal characteristics (=100%), thereby hampering medical decision making.

In diagnostic research, test characteristics (such as sensitivity, specificity, positive and negative predictive values) are usually determined by comparison to a 'gold standard'. The *gold standard*, by definition, is a diagnostic test or benchmark that is regarded as definitive. However, this gold standard is not always available. For instance, histopathological proof of inflammation of lung tissue and bacterial invasion is the gold standard for diagnosing bacterial pneumonia, but this procedure can only be performed after biopsy, surgery or autopsy. As these procedures are harmful to the patient, and not essential in the selection of the treatment, they are normally not done. For such diseases, a surrogate gold standard is frequently created, such as a combination of tests or expert judgements, called the *reference standard* [2]. In the absence of a gold standard, a randomised study comparing a diagnostic strategy with and without using a new diagnostic test or algorithm might be the alternative, possibly best approach.

After establishing a diagnosis, the next step is to consider all treatment options and to identify the most appropriate one for that specific patient. For instance, some antibiotics should be avoided in old (or very young) patients or in those with renal impairment or hematological diseases. In case of infection, treatment should be started as soon as possible. In some infections, diagnostic results can be awaited [3]. However, in critically ill patients with a presumed bacterial infection, antimicrobial treatment must be started as soon as possible, even before identification of the etiological microorganisms.

Obviously, physicians have to deal repeatedly with uncertainty and multiple options in the process of diagnosing and treating their patients. The incorporation of all these options and uncertainties, which sometimes must be realised in short periods of time, can be challenging. This is an area where computerised systems might be beneficial in assisting physicians in performing these difficult tasks.

1.2 Computer-assisted decision making

In as early as the 1960s, computer programs were developed to investigate the applicability of a computerised version of Bayes' theorem to assist clinicians in decision making [4, 5, 6, 7]. A frequently cited medical application is the 'acute abdominal pain program' developed by De Dombal et al., a program capable of diagnosing causes of abdominal pain, such as perforated peptic ulcer [5, 8]. However, the specific form of Bayes' theorem used in these programs is unsuitable for other forms of clinical decision support, such as treatment selection and prediction. In addition, it is not possible to incorporate clinical knowledge in this restricted form of Bayes' theorem. More about this probabilistic approach to clinical decision making is said in the next section.

As clinical knowledge lies at the heart of clinical decision making, as illustrated in Figure 1.1, researchers have come up with the idea that knowledge should also be explicitly incorporated into computer systems that support decision making. This idea has given rise to the area of knowledge engineering, a field within artificial intelligence that is concerned with the development of knowledge-based systems. Such systems are computer programs that contain large amounts of knowledge and that include reasoning mechanisms to manipulate the knowledge so that, based on factual observations, the system is capable of providing solutions to the actual real-world problem at hand.

The knowledge an expert has learnt, especially from experience, is normally highly dependent of the domain concerned [9]. The phrase 'expert system' was introduced at the beginning of the 1980s to denote knowledge-based systems that contain expert knowledge about specialised tasks. In the early years of expert systems, acquiring enough high-quality knowledge to build a robust and useful system was noticed to be a very tedious and often expensive activity. The process of knowledge acquisition, i.e., the elicitation of knowledge from human experts, was quickly identified as a major bottleneck in building an expert system, and the phrase 'expert system' fell into disuse. However, during the last decade, the phrase 'expert system' is again increasingly mentioned in literature on applications to describe knowledge-based systems incorporating expert knowledge. The main reason is that new, more versatile methods for building such systems have become available. Expert systems now exist in various forms, for different areas of expertise and, in the field of medicine, usually aim at achieving a better efficacy in managing disease in patients.

When building expert systems, one has to account for the dynamic nature of knowledge, as it may need to be updated by new knowledge. Therefore, knowledge base and inference engine should be kept separate, which allows updating the knowledge base without having to touch upon the inference engine. There are several formal languages to represent knowledge in a knowledge-base. In the most optimal situation, a formalism should offer a good trade-off, on the one hand, as a tool that is sufficiently expressive to encode relevant domain knowledge having a well-understood semantics, and, on the other hand, support fast computation or reasoning with the encoded knowledge. Certainly in the medical field it is valuable if the meaning of the knowledge to be represented is easy to comprehend by the user. Various formalisms to represent knowledge have been proposed, but most of them can be seen as either a variant or subset of logic or as a formalism to represent and reason with uncertainty [9]. Logics are, for example, used to represent ontologies, in particular description logics are popular in this area, whereas Bayesian networks are seen as modern, powerful methods to represent and reason with uncertainty. All formalisms have their strengths and weaknesses, and differences in expressive power [10].

For the domain of study of this thesis, one of the earliest expert systems, MYCIN, should be mentioned [11]. MYCIN is generally considered as being the first expert system in medicine and was developed at Stanford University

in the 1970s by Shortliffe. From a modern point of view, medical knowledge in the system was representation in a restricted, rule-like logic, augmented by an uncertainty calculus. The aim of the system was to assist internal-medicine specialists in diagnosing and treating certain blood-stream infections. These infections may have a diagnostic delay of up to 48 hours, which is unacceptably long in critically ill patients. MYCIN was developed to assist making decisions using incomplete information; in particular junior or less-experienced physicians were taken as the target group. Although MYCIN seemed promising, it was in the end only used as a research prototype and never put into practical use.

A perceived limitation of the system was in particular the restricted way in which uncertainty was handled, which was done by an ad-hoc method for reasoning with uncertainty, called the *certainty-factor model* [12]. However, the methods underlying MYCIN were able to solve problems which could not be handled by the early, restricted form of Bayes' theorem. Since the end of the 1990s, uncertainty reasoning in expert systems is done using more powerful probabilistic methods, where in particular statistical independence information is explored to specify probability distributions compactly, and to reason with probability distributions efficiently. These modern methods now surpass both the early uses of probability theory and the uncertainty logic of MYCIN.

1.3 Probabilistic reasoning

In principle, probability theory, as axiomatised by Kolmokorov at the beginning of the 20th century, contains all the necessary elements for reasoning with uncertainty. However, there were two major disadvantages linked to the practical use of probability theory: probabilistic specifications are in general exponential in size, and, computation of relevant probabilities from a specification is nondeterministic polynomial-time (NP) hard. The latter means that for some probability distributions computation could take exponential time in terms of the number of random variables involved. Actually, the abovementioned work from the the 1960s to the 1970s can been seen as early, partial solutions to this problem ([13]).

1.3.1 Bayes' theorem

Suppose, for example, that one wishes to compute the probability $\Pr(D \mid S)$, were D stands for the presence (d) or absence ($\bar{d}$) of a disorder D, and S for a set of signs and symptoms; D and S are discrete random variables. By *Bayes' theorem*, this probability can be computed as follows [14]:

$$\Pr(D \mid S) = \frac{\Pr(S \mid D)\Pr(D)}{\Pr(S)}$$

The problem now is that the probability $\Pr(S \mid D)$ is exponential in size in the number of random variables in S, and, thus, very difficult to specify.

Let in the following $S = \{S_1, \ldots, S_n\}$. By assuming that S_i is conditionally independent of S_j given D, with $i \neq j$, it is possible to write

$$\Pr(S \mid D) = \prod_{i=1}^{n} \Pr(S_i \mid D)$$

and now it is possible to specify the exponential number of probabilities $\Pr(S \mid D)$ by a linear number of probabilities of the form $\Pr(S_i \mid D)$, for $i = 1, \ldots, n$. The probabilities $\Pr(S_i \mid D)$ can be interpreted as stating that 'disorder D may cause the occurrence of sign S_i'. The resulting, simplified form of Bayes' theorem is often referred to as *naive Bayes*. Although easy to understand, it is normally not the case that signs and symptoms are conditionally independent given a disorder: even when the disorder is known, the signs and symptoms may be correlated. In addition, the naive form of Bayes' theorem is unsuitable for representing clinical decision making where diagnostic reasoning, treatment selection and prediction of outcome are combined. Bayesian networks, on the other hand, which are discussed in the next section, allow specifying unrestricted joint probability distributions in a compact manner, yet support computation of any probability without making unrealistic independence assumptions as in naive Bayes.

1.3.2 Bayesian networks

Probabilistic graphical models are graphs[1] in which vertices represent random variables, and absence of edges or arcs, dependent on the type of graphical model, represents conditional independence assumptions. A joint probability distribution is defined over a set of random variables. The possibility to represent independence assumptions is why this formalism allows one to specify a joint probability distribution compactly. Two types of probabilistic graphical models are in particular popular in research. In the first one, undirected graphs are used, whereas in the second type of graphical model directed graphs are used.

Undirected graphical models, also called *Markov networks*, employ a simple definition of independence: if the undirected paths from vertices in the sets of vertices A and B are always blocked by a vertex from a third set, let us say C, then the corresponding sets of random variables in the associated joint probability distribution, A and B, are conditionally independent given the third set of associated random, C.

Directed graphical models, usually called *Bayesian networks* or *belief networks* (BNs), support a more complicated notion of independence as they take into account the directionality of the edges (which are then called arcs). Starting with a diagram indicating the causal relationships between domain variables, it is possible to interpret the diagram as a Bayesian network. Once the

[1] Informally, a graph is a diagram consisting of points and lines or arrows that connect the points; a line and arrow always connect two points. Points are also called *vertices*, and lines are often called *edges*, whereas arrows are called *arcs*.

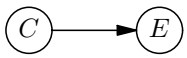


Figure 1.2: A simple example of a causal relationship.

probability distribution has been supplied, it can be used for calculating probability distributions of the unobserved random variables given the observed random variables. As a Markov network, a Bayesian network consists of two parts: a qualitative part; an acyclic directed graph, i.e. a directed graph which does not contain directed paths where it is possible to end at a vertex where the path was started, and a quantitative part; the joint probability distribution. The qualitative part reflects all the dependences present in the quantitative part, i.e., the joint probability distribution, and encodes the causal relations among the domain variables in the form of the acyclic, directed graph.

Formally, a *Bayesian network*, BN for short, is defined as a pair $\mathcal{B} = (G, \Pr)$, where $G = (\mathbf{V}(G), \mathbf{A}(G))$ is a directed acyclic graph with a set of vertices $\mathbf{V}(G) = \{V_1, \ldots, V_n\}$, corresponding one to one to random variables, here denoted by the same indexed letters, and a set of arcs $\mathbf{A}(G) \subseteq \mathbf{V}(G) \times \mathbf{V}(G)$, and $\Pr$ is a joint probability distribution $\Pr(V_1, \ldots, V_n)$ representing statistical dependences and independences among the variables, respecting the independences represented in the graph, as follows:

$$\Pr(V_1, \ldots, V_n) = \prod_{i=1}^{n} \Pr(V_i \mid \pi(V_i)),$$

where $\pi(V_i)$ stands for the variables corresponding to the parents of vertex V_i. For example, in Figure 1.2, C is the parent of E, which in turn is the child of C. The joint probability distribution $\Pr(E, C)$ can now be decomposed as follows:

$$\Pr(E, C) = \Pr(E \mid C) \Pr(C)$$

1.3.3 Causality and conditional independence

Causal knowledge is of major importance in finding the right structure of a Bayesian network. It is well known that humans are well capable of distinguishing cause from effect when talking about a domain of concern. Experts are encouraged to specify variables that are directly relevant causally to a variable and do influence that variable directly. This is indicated in diagrammatic form as a directed graph, where there is an arc drawn from the cause to the effect. For example, in Figure 1.2, it is indicated that cause C influences effect E. Often, the directions of the arcs reflect the direction of causal influence, as perceived by the decision maker.

However, it is not always the case that every directed arc in a Bayesian network denotes causality, as a Bayesian network specifies conditional independence assumptions, and not causal knowledge. For example, as mentioned

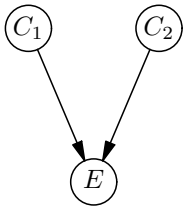


Figure 1.3: Two causes and a common effect.

above, the dependence between C and E in Figure 1.2, corresponds to the joint probability distribution

$$\Pr(E, C) = \Pr(E \mid C) \Pr(C)$$

but as it also holds, by the axioms of probability theory, that

$$\Pr(C, E) = \Pr(C \mid E) \Pr(E)$$

the arc $C \rightarrow E$ might also be reversed, i.e., $E \rightarrow C$, yielding an equivalent Bayesian network, as $\Pr(E, C) = \Pr(C, E)$.

The only graphical structure where this reversal would violate equivalence is in the case of a so-called *v-structure*, also called collider, where there are two or more causal random variables, e.g., C_1 and C_2, and one effect random variable, e.g., E. Figure 1.3 illustrates this situation. Here, the arcs cannot be reversed, and this structure can, therefore, be given a causal interpretation. The standard causal interpretation is that C_1 and C_2 are common causes of the effect E.

After making sure that available causal knowledge is correctly represented in a Bayesian network, checks with respect to possible violations of conditional independence of the associated random variables are usually made. Here use is made of the the *Markov condition*, that states that any vertex is conditionally independent of its non-descendants (the vertices that cannot be reached by a directed path started at the given vertex) given its parents. Hence, the parents block the influence of vertices that have paths going to the parents. Considering this condition may give rise to the insight that an intermediate vertex in the causal structure of the system is missing, and, subsequently, the model may be refined. The probabilistic consequences of causal structure, in terms of patterns of dependences, are so strong that an expert seeking to fulfill the Markov condition, in fact, often ends up finding the right causal model of the domain.

1.4 Decision trees

A decision tree, or classification tree, is a data structure which efficiently organises an ordered series of questions and can be used to represent clinical guidelines. Upon traversing the decision tree, the user is presented with questions. The answers determine which further questions should be useful. Essentially,

a decision tree defines a set of paths from the root node to the leaves. The answers to the questions on the non-leaf or internal nodes determine which branch is followed. When a leaf is reached, then there are no more questions and the result, i.e. a classification, is given as the output of the tree. Decision trees are excellent tools for choosing between several courses of action, such as treatment planning [15]. However, they are unsuitable for explicitly reasoning with uncertainty, which is why decision trees are only used occasionally in the research underlying this thesis.

1.5 Data mining and machine learning

In computer science, the field of data mining involves mining collections of data to find and explore relationships between the attributes making up the data. By definition, data mining is the (semi or fully) automated process of discovering patterns in data. The patterns discovered must be meaningful in that they lead to some advantage, such as an economic advantage [16]. Mining data is based on exploitation of various techniques from fields such as statistics, pattern recognition, numerical mathematics and machine learning [17, 18].

Machine learning, i.e., the ability of a machine to improve its performance based on previous results, can be seen as a form of data mining: computerised methods are used to gain new or additional knowledge in a particular area of interest. Learning from data may well become more important in the future, as large amounts of data are being collected in areas like, for example, biology and medicine. Biological data, such as micro-array data (usually binary, representing the absence or presence of specific genes or gene expression) [19] can be very complex and, subsequently, there is an obvious need for new techniques to be developed.

Health-care institutions also collect large amounts of information about patients' clinical status and interventions performed by physicians and nursing staff, such as therapies. The use of these 'raw' process data is still limited to patient-management purposes [20]. However, there is increasing interest to analyse these data collections for scientific purposes. As the complex decision-making processes in health-care often include uncertainty, Bayesian-network models might be useful to support decision-making in real-life practice.

So far, Bayesian networks have been used as models for uncertainty reasoning in decision-support systems for clinical domains ([21]); they have been less popular as tools for the analysis of clinical data, despite the availability of a wide range of Bayesian-network structure and parameter learning techniques [22]. This is somewhat surprising, as the statistical nature of Bayesian networks would render them – in principle – as useful as data-analytical tools as logistic regression or other multivariate analysis techniques. In the research underlying this thesis, Bayesian networks have not only be adopted as the principal technique for decision making, but also for the purpose of data analysis. However, standard statistical techniques have also been used.

In case of mining data that have a temporal nature, other descriptive statics, such as rates, are required. When it concerns patient data, time series of patients are frequently used, where time runs from time of admission of the patient to the hospital to the day the same patient is discharged. A *time series* is a sequence of data points, measured typically at successive times, spaced at (often uniform) time intervals:

$$\langle X_t \mid t = 0, \ldots, n_p \rangle$$

where $t = n_p$ is the time of discharge of patient p. Temporal Bayesian networks are increasingly used to model such data.

When analysing data researchers often face the reality of missing values, that may result from many reasons. When it concerns continuous data, several techniques allow for a fair substitute value to be *imputed*, such as the mean of all values that are available [23]. Another method that offers an alternative for missing data is the expectation maximisation (EM) algorithm [24].

In health care, several scoring systems are in use to assess the condition of the patient. A typical example is the Apache score, which is used to assess the severity of illness of patients [25]. For the computation of these scores, several symptoms and other factors are needed. In real life, though, not all factors of a scoring system may be known at a particular moment; and these scores may not accurately reflect the patients' clinical condition [26]. A significant advantage of using Bayesian networks, is that they rely on their prior probabilities and can produce reliable posteriors, even when data is missing.

1.6 Ventilator-associated pneumonia

Accurately diagnosing infections is desirable for critically ill patients. Patients who depend on respiratory support are extremely vulnerable to develop *ventilator-associated pneumonia*, or VAP for short. VAP is a type of pneumonia that occurs in patients that are mechanically ventilated for more than 48 hours in intensive care units (ICUs), usually after previous colonisation of the upper respiratory tract and is known to be associated with increased morbidity and mortality. There is no readily available, accurate gold standard for diagnosing VAP. Important signs and symptoms that may indicate the presence of VAP include: *body temperature*, amount and colour of *sputum production*, interpretation of the *chest X-ray*, the duration of *mechanical ventilation*, amount of *leukocytes*, i.e. white blood cells, and pO_2/FiO_2, i.e. the ratio between the amount of oxygen in the arterial blood and the fractional inspired oxygen concentration [27].

The development of VAP is a time-based process. Durations of *mechanical ventilation* and *hospitalisation* represent a certain notion of time. The probability of colonisation of the respiratory tract with potentially pathogenic microorganisms is low during the first 48 hours of mechanical ventilation. After a period of two days, this probability increases and, as a result, the probability of developing VAP increases accordingly. To explore dynamics in the development of VAP, patient data should be analysed from admission to discharge by means

of a time series. Few studies have addressed the predictive value of changing infection parameters in time. By comparing data of patients that did and did not develop VAP, possible differences in dynamics could be explored and discovered.

1.7 Aim of this thesis

The purpose of the research described in this thesis was to develop Bayesian network models for the analysis of patient data, as well as to use such a model as a clinical decision-support system for assisting clinicians in the diagnosis and treatment of VAP in mechanically ventilated ICU patients. Both construction of the models and use of such a decision-support system were expected to be guided by clinical expertise and patient data collected within the hospital. For this research, we were particularly interested in patient data collected routinely by the electronic patient record system, EPR for short. Although not yet used in all hospital wards, EPR has become relatively common in ICUs. The exploitation of routinely collected ICU data, therefore, defined the context of the research.

Typical for EPR data is that it includes both clinical and laboratory data that is time oriented. In this thesis methods, techniques and tools are developed for using these temporal data to support clinical decision making. To this end, graphical statistical models, in particular Bayesian networks, have been chosen as starting point, since these formalisms enable the deduction of uncertain knowledge, both on the basis of knowledge from a domain and on the basis of factual data. Moreover, Bayesian networks provide the opportunity to represent both temporal and atemporal relationships. The validity of such associations should then be clinically judged. In this project specific attention has been paid to the structure and content of both atemporal and temporal Bayesian networks, the learning of networks from data and to the role of expert knowledge in this.

A Bayesian decision-support system (BDSS) for diagnosing and treating VAP was developed that, in principle, can be used by intensivists. A previously collected database of data from mechanically ventilated ICU patients was used for improving and assessing the performance of the BDSS.

The research has built upon the work that has been described previously in the thesis of Karin Schurink [28].

1.8 Thesis outline

This thesis consists of three parts, namely Part I in which the diagnosis of VAP plays a central role, whereas in Part II dynamics of diagnostic parameters in the ICU environment are further analysed. Predicting causative pathogens and the selection of antibiotic treatment for VAP is described in Part III.

The chapters are organised as follows.

In Chapter 2, endotracheal colonisation patterns of ICU patients were investigated. The results of these colonisation dynamics were incorporated in a previously constructed Bayesian-network model, to improve predictions about the etiology of endotracheal colonisation in ICU patients. Furthermore, effects of previously administered antibiotics on these colonisation patterns were studied for further model improvement.

In Chapter 3, the test characteristics of the previously developed Bayesian-network model for diagnosing VAP are determined. This includes the use of machine learning techniques to optimise the performance of the Bayesian-network model.

In Chapter 4, the dynamics of the signs and symptoms in mechanically ventilated ICU patients are explored, where a distinction was made between patients diagnosed with and without VAP to determine differences in dynamic patterns.

Temporal Bayesian networks are introduced and constructed for the purpose of exploring (a)temporal associations in data of patients with and without VAP in Chapter 5.

For the purpose of starting targeted antimicrobial treatment, the performance of the Bayesian network model in predicting pathogens causing VAP is assessed in Chapter 6.

The practical use of the closed-world assumption to the selection of optimal antimicrobial treatment is investigated in Chapter 7. In addition, different ways of modelling susceptibility were explored and a generalised version of the noisy threshold was introduced in Chapter 8.

What has been achieved by the research described in this thesis is summarised in Chapter 9. In addition, limitations of the approaches taken and possibilities for further research are discussed.

Part I

Diagnosis of VAP

[2]

Colonisation Dynamics of the Respiratory Tract in ICU Patients

2.1 Introduction

Respiratory tract colonisation with nosocomial bacteria predisposes critically ill patients to life threatening infections, such as ventilator-associated pneumonia (VAP) [29]. Approximately one quarter of ICU patients develop VAP, which makes it the most frequently occurring ICU-acquired infection, responsible for the majority of antibiotic prescriptions in ICU [30]. In patients admitted to the ICU directly from the community, the upper respiratory tract flora changes from so-called early-onset bacteria (such as *S. pneumoniae*, *H. influenzae* and *S. aureus*) to typical late-onset (or nosocomial) pathogens (such as enteric Gram-negative bacteria, *P. aeruginosa* and *Acinetobacter* species). The mechanisms underlying this ecological change remain largely unknown, though a role of antibiotics herein seems plausible. Only a proportion of the colonised patients will develop VAP, usually after a gradual change from asymptomatic colonisation to clinical infection.

Bacterial colonisation of the respiratory tract frequently persists, even when a patient receives antimicrobial treatment, and even though the colonising bacteria are, in vitro, susceptible to the antibiotics. The capacity of bacteria to continue to grow in the presence of antibiotics has been demonstrated long ago [31], *in-vitro*, and such bacteria were called 'persister' cells. Such persister cells appear specialised survivors [32, 33].

A cohort of mechanically-ventilated ICU patients was used to investigate the phenomenon of persistent respiratory tract colonisation and the effects of systemic antibiotics hereon, by analysing endotracheal aspirate cultures performed during ICU stay. We hypothesised that antibiotics with presumed

This chapter is published in Intensive Care Medicine, 2008.

efficacy, based on in-vitro susceptibility testing, reduce the likelihood of persistence of respiratory tract colonisation, compared to antibiotics presumed to be ineffective or when no antibiotics were administered at all. Of note, we did not investigate the effects of topical antibiotics. Previously, we described the construction and evaluation of a computerised decision-support system for diagnosing VAP [34]. This computer system, based on Bayesian statistics, computes the likelihood that a mechanically-ventilated ICU patient suffers from VAP. The next step in model-development is to have the model predict the most likely pathogen causing VAP. The analyses described in the current manuscript were needed as input for optimising the predictive performance of the computer system.

2.2 Materials and methods

2.2.1 Patients

All patients admitted to two adult ICUs (a ten-bed medical ICU and an eight-bed neurosurgical ICU) of the University Medical Center Utrecht (UMCU) between January 1st 2000 and January 1st 2003 were included in this observational cohort study. Patients with cystic fibrosis were excluded, as they frequently are chronically colonised by Pseudomonas aeruginosa [35]. The UMCU is a 1062-bed tertiary care hospital. In both ICUs all relevant clinical, laboratory and microbiological variables are registered on a daily basis in computerised patient data management systems. The institutional review board waived the necessity of informed consent for this observational study.

2.2.2 Definitions

Respiratory tract colonisation (from here often referred to as 'colonisation') status is based on microbiological culture results of endotracheal aspirates, which were obtained as part of daily patient care, without specific surveillance protocol. Samples were analysed semi-quantitatively (three quadrant streak method), and culture positivity was defined as any growth [36]. Absence of a pathogen was considered 'negative'. Distinction was made between patients who were 'colonised on admission' and patients who 'acquired colonisation during ICU stay'.

Colonisation was analysed for the following bacterial species: *Pseudomonas aeruginosa*, *Acinetobacter* species, *Enterobacteriaceae*, *Staphylococcus aureus*, *Haemophilus influenzae* and *Streptococcus pneumoniae*. The group of *Enterobacteriaceae* contains multiple species (spp.) including *Escherichia coli*, *Klebsiella pneumoniae*, *Klebsiella* spp. , *Proteus Mirabilis*, *Citrobacter* spp., *Enterobacter cloacae*, *Enterobacter* spp., *Morganella* spp. and *Serratia* spp..

2.2.3 Bacterial colonisation and persistence

Colonisation on admission was defined as a positive culture obtained $\leq$48 hours after ICU admission. Acquired endotracheal colonisation with a certain pathogen was defined as colonisation demonstrated >48 hours after ICU admission and preceded by a negative culture for that specific pathogen. The first day on which a specific pathogen was detected was considered 'the first day of colonisation'. When no culture results were obtained within 48 hours after ICU admission, the microbiological status of a patient on admission was considered 'unknown'. As a patient can get colonised by multiple pathogens during ICU stay, it is possible to categorise a patient into more than one category of colonisation (for example: a patient is colonised with *S. aureus* on admission and acquires *P. aeruginosa* during ICU stay). For each group of microorganisms prevalence of colonisation on admission was determined, as well as the incidence and first day of acquired colonisation.

Persistent colonisation was defined as a time period, starting at the first day of colonisation, in which at least two successive culture results were positive for the same bacterial species. After a positive culture, a single negative culture followed by positive cultures with same bacterial species was also considered persistent colonisation. The period between the last positive culture and ICU-discharge, during which no other culture was obtained, again was considered persistent colonisation. 'Non-persistent colonisation' was defined as conversion of a positive colonisation status to a negative status (for that pathogen) after a period of persistent colonisation. The interval between two successive cultures was considered a sub-period of colonisation. Thus, multiple sub-periods of colonisation per patient could occur. Two successive negative sputum cultures followed by a single positive culture, again followed by a negative culture, was considered 'non-persistent colonisation'. Duration of bacterial persistence was analysed for all episodes of bacterial colonisation.

2.2.4 Antibiotics

For each pathogen, susceptibility to antibiotics was based on in-vitro susceptibility data as determined in our hospital's medical microbiology laboratory. Breakpoints for susceptibility were based on CLSI recommendations [Clinical and Laboratory Standards Institute, Performance Standards for Antimicrobial Susceptibility Testing: 16th Informational Supplement]. "No antibiotics" was considered ineffective treatment. The effects of antibiotics on bacterial persistence were analysed for all episodes of colonisation by determining antibiotic exposure during four-day periods before the moment of culture taking. Exposure was defined as administration of presumed effective antibiotics (based on susceptibility data) during two or more days in this four-day period.

Naturally, only patients with at least two cultures obtained during ICU admission were eligible for this analysis. In case of multiple colonisation periods of an individual patient with the same bacteria, only the first episode of bacterial colonisation was analysed. Furthermore, repetitive periods of colonisation

in an individual patient with similar antibiotic exposure were excluded.

The likelihood of bacterial persistence was determined for situations with and without exposure to effective antibiotics. Absence of antibiotics was grouped with exposure to ineffective antibiotics. In a subsequent case-control analysis, each patient that acquired bacterial colonisation with a certain pathogen was matched to three patients that had not acquired colonisation with that particular pathogen during ICU stay. The day of matching was the first day of colonisation for the "case" patient and a day representing similar length of stay in ICU for the "control" patient (See Figure 1 for an example). Antibiotic exposure in the days before matching was defined as described above. Antibiotic exposure between patients with acquired colonisation and their controls was evaluated.

2.2.5 Statistical analysis

SPSS version 14.0 [37] was used for statistical analysis and Gnuplot version 4.0 [38] for creating graphs. Student t-test and Mann Whitney-U tests were used when appropriate and significance level less than 0.05 was considered statistically significant. A Kaplan-Meier survival analysis was performed to assess the duration of persistence over time. Colonised patients were eligible for the analysis and patients colonised when discharged were censored, i.e., 'withdrawn alive'.

2.3 Results

2.3.1 Patients

The total study cohort consists of 1,410 admissions (17,709 ICU days), with 3,289 cultures of endotracheal aspirates performed in 1,012 admissions (3.25 cultures per admission) and bacterial growth from at least one endotracheal aspirate was documented for 715 admissions (71%; 2,111 cultures). The characteristics of this patient population have been described previously [34]. For the 715 patients included in this study, the median duration of ICU stay was 11 days (IQR 5–21 days). The median ICU stay of patients without documented bacterial growth from endotracheal aspirates was 6 days (IQR 3–10 days) ($p<0.05$). For clarity, patients without cultures taken were not included in this study. To determine whether a longer stay in ICU was associated with a lower or higher number of samples taken, the first 50 days of ICU stay were categorised in consecutive 5-day groups. The average percentage of days on which cultures were taken was 20% and varied from 17% on days 41–45 to 22% on days 26–30. Therefore, longer length of stay did not increase the likelihood of sampling.

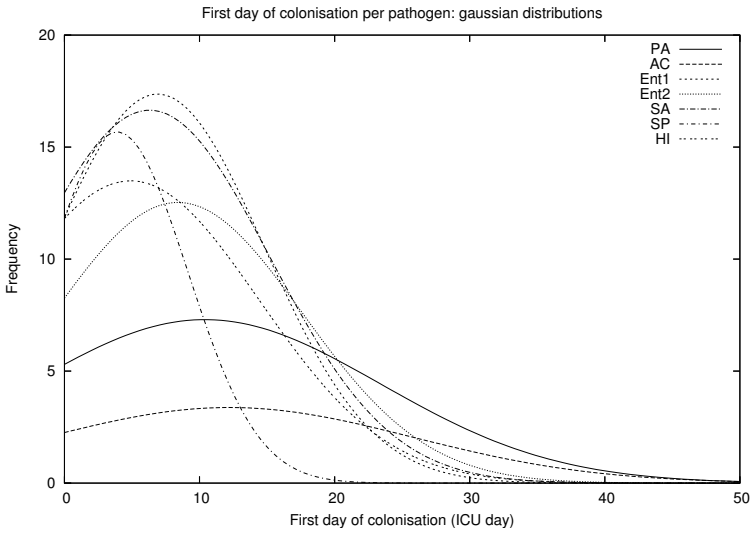


Figure 2.1: First day of colonisation.

2.3.2 Episodes of colonisation

Colonisation with Enterobacteriaceae occurred most frequently and was demonstrated in 413 patients followed by *S. aureus* (258 patients, of which 140 (54%) admitted to the medical ICU and 118 (46%) to the neurosurgical ICU), *P. aeruginosa* (180 patients), *S. pneumoniae* (153 patients), *H. influenzae* (142 patients) and *Acinetobacter* species (73 patients) (See Table 2.1).

Colonisation on admission could not be evaluated for 295 patients, because of absence of cultures obtained within 48 hours after admission. Colonisation on admission was analysed in 311 patients and was most frequently observed for Enterobacteriaceae, that is, in 33%, whereas colonisation on admission with *Acinetobacter* species was documented in 4% of these cases. Overall, the median first day of acquired colonisation was day seven (IQR 4–14 days) after ICU admission, which differed considerably between pathogens. Acquired colonisation with *H. influenzae* and *S. pneumoniae* was documented after a median of four and five days in ICU, respectively, whereas the median first day of colonisation with *Acinetobacter* species and *P. aeruginosa* was eleven days. The Enterobacteriaceae group and *S. aureus* fell in-between the early-onset and late-onset colonisers (See Table 2.1). Figure 2.1 shows the (Gaussian-fitted) distribution of the first day patients were colonised, specified per pathogen group.

The 'burden of colonisation', as depicted in Figure 2.2 by fitted Gaussian functions, was represented by the number of patients colonised per day following ICU-admission, was highest for Enterobacteriaceae.

Once established, duration of colonisation also differed considerably between pathogens. Bacterial persistence occurred in 85% of patients colonised

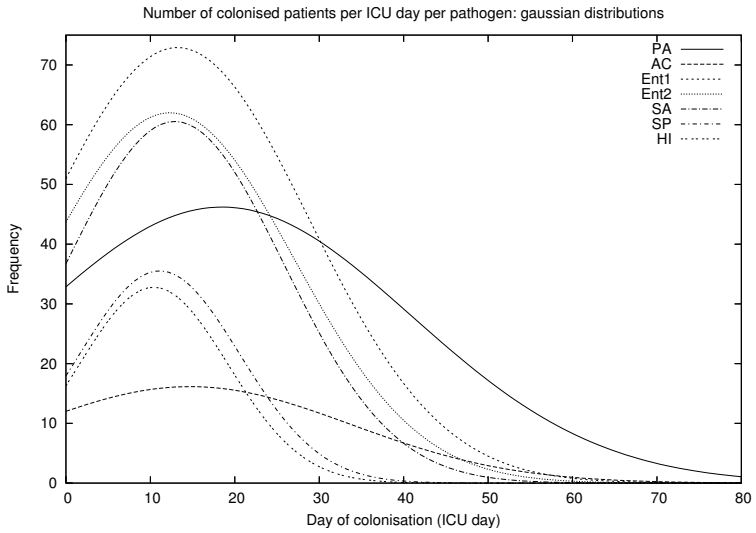


Figure 2.2: Burden of colonisation.

with *P. aeruginosa*, which was a significantly higher proportion as compared to all other bacteria or bacteria groups (Table 1). Lowest proportions of patients with bacterial persistence were found for *H. influenzae* and *S. pneumoniae*. The median duration of bacterial persistence for patients colonised with *P. aeruginosa* was eight days (IQR 4–16 days), which was significantly longer than persistence of any of the other pathogens, as shown in Table 2.1. Figure 2.3 shows the (Gaussian-fitted) distribution of duration of persistence and the Kaplan-Meier curve (Figure 2.4) represents the cumulative probability of a patients persistently colonised by a pathogen during ICU stay.

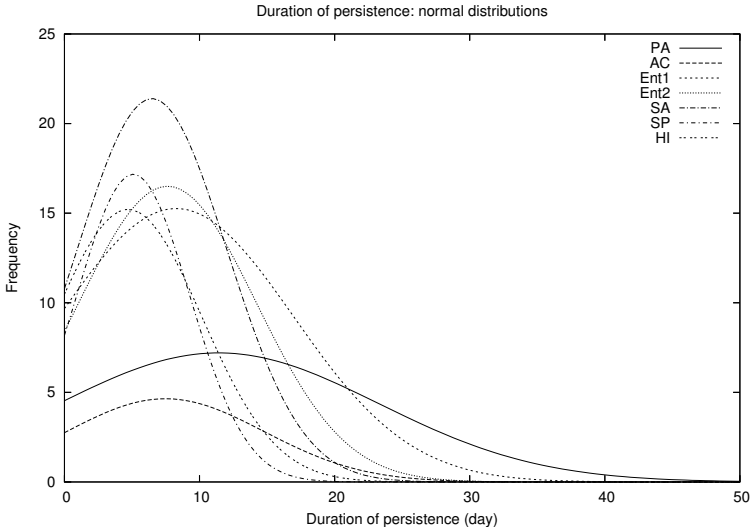


Figure 2.3: Duration of bacterial persistence.

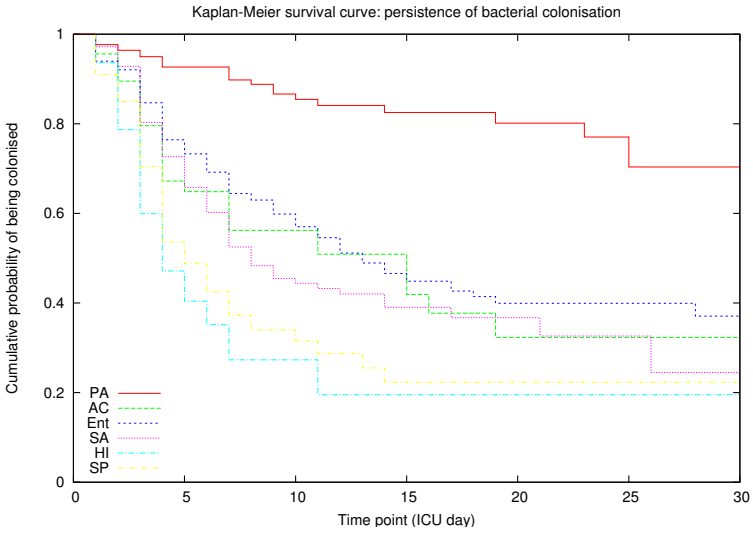


Figure 2.4: Kaplan-Meier curve: duration of bacterial persistence.

Table 2.1: Colonisation characteristics.

		Colonisation on Admission N=311**	**Acquired Colonisation** N=715***			**Duration of Bacterial Persistence**			
Pathogen	**N** (col. pts.)	n (% of total N)	n (% of total N)	Median first day [IQR]	Difference in first day of acquired colonisation compared to PA M-W U test: p-value	BP	NBP	Median [IQR]	Difference in duration of persistence compared to PA M-W U test: p-value
HI	142	63 (20%)	62 (9%)	4 [3- 6]	0.001*	61 (43%)	81 (57%)	4 [2- 5]	0.001*
SP	153	86 (28%)	57 (8%)	5 [4- 9]	0.001*	72 (47%)	81 (53%)	4 [2- 6]	0.001*
SA	258	95 (31%)	117 (16%)	6 [4-13]	0.001*	154 (60%)	104 (40%)	5 [3- 8]	0.001*
Ent	413	104 (33%)	240 (34%)	6 [4-11]	0.001*	262 (64%)	151 (37%)	6 [3-10]	0.001*
PA	180	57 (18%)	97 (14%)	11 [7-24]	x	153 (85%)	27 (15%)	8 [4-16]	x
AC	73	11 (4%)	55 (8%)	11 [6-17]	0.552	41 (56%)	32 (44%)	5 [3-12]	0.008*

* $p<0.05$

* Patients can be colonised with more than 1 pathogen.

** Total number of admissions analysed for acquired colonisation.

BP: bacterial persistent; NBP: bacterial non-persistence; M-W U test: Mann-Whitney U-test; IQR: interquartile range; col: colonised; pts: patients. PA: *Pseudomonas aeruginosa*; AC: *Acinetobacter* species; Ent: Enterobacteriaceae; SA: *Staphylococcus aureus*; SP: *Streptococcus pneumoniae* and HI: *Haemophilus influenzae*.

2.3.3 Antibiotics and antibiotic susceptibility

In the 715 patients included in this analysis antibiotics were administered on 7,102 (61%) of 11,689 patient days. Amoxicillin clavulanic acid (58% of patients), gentamicin (36%), ceftriaxone (31%) and ciprofloxacin (28%) were prescribed most frequently (See Table 2.2). When expressed in proportions of all patient days the same four antibiotics were prescribed most frequently. Median durations of antibiotic therapy ranged from two to five days. Of note, many patients were discharged with antibiotic therapy continued on regular wards.

Non-susceptibility levels to antibiotics of the first isolates per patient were fairly low (See Table 2.3). For instance, all *S. pneumoniae* were susceptible to penicillin, all *S. aureus* to flucloxacillin and all Gram-negative bacteria to imipenem. Resistance development during colonisation only occurred sporadically (data not shown).

2.3.4 Effects of antibiotics on colonisation persistence

Systemic antibiotic treatment (with antibiotics for which pathogens were in vitro susceptible) was associated with reduced persistence of all pathogens except for *P. aeruginosa* and *Acinetobacter* species (See Table 2.4). The highest relative risk (RR) of non-persistence during effective antibiotic treatment was 3.1 (95% CI 1.4–6.6) for *H. influenzae*. In this analysis, the effect of effective antibiotics was compared to ineffective antibiotics, being mainly 'no antibiotics' for *H. influenzae*. Significant associations between effective antibiotics and non-persistence of colonisation were also observed for *S. pneumoniae* (2.1, 95% CI 1.2–3.6), *S. aureus* (1.6, 95% CI 1.3–2.0), and Enterobacteriaceae (1.5, 95% CI 1.3–1.7).

In the case-control analysis patients that had acquired colonisation with either *H. influenzae*, *S. pneumoniae*, *S. aureus*, Enterobacteriaceae had had less exposure to effective antibiotics in the days before acquisition was demonstrated (See Table 2.5). No such association could be demonstrated for *P. aeruginosa* and *Acinetobacter* species.

2.4 Discussion

The major findings of this observational study are that endotracheal colonisation dynamics and the effects of systemic antibiotics hereon differ among bacterial species. This implies that evaluation of bacterial persistence as a marker of ongoing infection or as an endpoint of prevention studies should be pathogen-specific. Early-onset pathogens (such as *H. influenzae*, *S. pneumoniae* and *S. aureus*) appear to be much more sensitive to intravenously administered antibiotics (with documented in-vitro susceptibility) than typical late-onset pathogens (such as *P. aeruginosa* and *Acinetobacter* species).

Table 2.2: Number of patients receiving systemic antibiotic treatment during ICU stay.
IQR: interquartile range. NA: not applicable

Antimicrobial agent	Patients (%) total n = 715	Patient days (%) total n = 11689	Duration of antimicrobial administration per patient: median days [IQR]
Benzyl Penicillin	29 (4%)	152 (1%)	2 [1-7]
Cephalotin	31 (4%)	79 (1%)	2 [1-3]
Ceftazidime	100 (14%)	509 (4%)	3 [1-8]
Ceftriaxone	223 (31%)	996 (9%)	3 [1-6]
Cefuroxime	5 (1%)	15 (0%)	2 [1-5]
Ciprofloxacin	199 (28%)	1040 (9%)	3 [1-8]
Amoxicillin + Clavulanic acid	412 (58%)	2220 (19%)	5 [3-7]
Clindamycin	64 (9%)	311 (3%)	3 [1-8]
Cotrimoxazole	135 (2%)	609 (5%)	3 [1-7]
Erythromycin	70 (10%)	381 (3%)	4 [2-7]
Flucloxacillin	148 (21%)	639 (5%)	2 [1-6]
Gentamicin	255 (36%)	887 (8%)	3 [1-5]
Imipenem	26 (4%)	127 (1%)	4 [2-8]
Meropenem	93 (13%)	497 (4%)	3 [1-9]
Metronidazole	33 (5%)	118 (1%)	2 [1-5]
Piperacillin	102 (14%)	587 (5%)	5 [1-9]
Tazobactam	73 (10%)	401 (3%)	5 [1-9]
Tobramycin	128 (18%)	760 (7%)	5 [1-9]
Vancomycin	68 (10%)	286 (2%)	3 [1-6]
None	NA	4587 (39%)	NA

Table 2.3: Frequency of pathogens (only first isolate per patient) with antibiotic susceptibility based on in vitro susceptibility testing.
PA: *Pseudomonas aeruginosa*; AC: *Acinetobacter* species; Ent: Enterobacteriaceae; SA: *Staphylococcus aureus*; SP: *Streptococcus pneumoniae* and HI: *Haemophilus influenzae*.

	Frequency and percentage of first isolates being susceptible					
pathogen	**PA**	**AC**	**Ent**	**SA**	**SP**	**HI**
patients (n)	180	73	413	258	153	142
Antimicrobial						
Benzyl Penicillin					153 (100%)	
Ceftazidime	157 (87%)	69 (94%)	390 (94%)			
Ceftriaxone			389 (94%)			
Cefuroxime			244 (92%)			
Ciprofloxacin	165 (93%)		400 (97%)			
Amoxicillin + Clavulanic acid			272 (66%)		153 (100%)	139 (98%)
Clindamycin				255 (99%)		
Cotrimoxazole			372 (90%)	248 (96%)		
Flucloxacillin				258 (100%)	153 (100%)	
Gentamicin	178 (99%)	68 (93%)	395 (96%)			
Imipenem/ Meropenem	180 (100%)	180 (100%)	413 (100%)			
Piperacillin/ Tazobactam	169 (94%)	68 (93%)	397 (96%)			
Tobramycin	172 (96%)					
Vancomycin				258 (100%)		

Table 2.4: Effects of antimicrobial treatment on bacterial persistence.
BP: bacterial persistent; NBP: bacterial non-persistent; AB: antibiotics; RR: relative risk.
*p<0.05

	Effects of AB on Bacterial Persistence		
Pathogen	*N* sub-periods of BP (% effective AB)	*N* sub-periods of NBP (% effective AB)	RR [95%CI] NBP when receiving effective AB treatment
H. influenzae	21 (24%)	81 (73%)	3.1 [1.4-6.6]*
S. pneumoniae	21 (38%)	81 (79%)	2.1 [1.2-3.6]*
S. aureus	124 (48%)	104 (78%)	1.6 [1.3-2.0]*
Enterobacteriaceae	289 (52%)	151 (77%)	1.5 [1.3-1.7]*
P. aeruginosa	262 (33%)	27 (37%)	1.1 [0.7-1.9]
Acinetobacter spp.	57 (47%)	32 (25%)	0.5 [0.3-1.0]

Table 2.5: Case-control analysis: Effects of antibiotics on acquisition of colonisation.
AB: antibiotics; OR: odds ratio.
* p<0.05

	risk of acquiring colonisation		
Pathogen	*N* cases (% effective AB)	*N* controls (% effective AB)	OR [95%CI] of acquiring colonisation when receiving ineffective AB treatment as compared to receiving effective treatment
S. pneumoniae	57 (30%)	171 (66%)	4.6 [2.4-8.8]*
S. aureus	117 (26%)	351 (55%)	3.5 [2.2-5.6]*
H. influenzae	62 (28%)	186 (55%)	3.1 [1.7-5.9]*
Enterobacteriaceae	240 (39%)	720 (55%)	1.9 [1.4-2.6]*
P. aeruginosa	97 (12%)	291 (21%)	1.8 [0.9-3.6]
Acinetobacter spp.	55 (22%)	165 (23%)	1.0 [0.5-2.2]

Strengths of our study include the large patient cohort with detailed data on colonisation, antibiotic susceptibility of pathogens and antibiotic administration. The latter allowed us to specifically analyse the effects of antibiotics for which the pathogen was susceptible, based on in vitro susceptibility testing. Limitations, though, include the pragmatic nature of microbiological data collection, as we had to rely on samples obtained for clinical reasons and not on a prospectively implemented surveillance schedule. Nevertheless, on average 3 cultures were obtained for each of the 715 patients included in this analysis.

Our data confirm previously reported findings on the different time-frames of bacterial colonisation in mechanically-ventilated patients [39, 40]. Colonisation on admission was more frequently observed for early-onset pathogens and acquired colonisation most frequently occurred with typical late-onset pathogens. Our findings strongly suggest that there are differences in the effects of intravenous antibiotics on different bacterial species. This implies that prophylactic use of systemic antibiotics will mainly affect endotracheal colonisation with early-onset pathogens. The risk of acquiring colonisation with *S. aureus*, *S. pneumoniae* and *H. influenzae* was reduced by approximately 50% when effective antibiotics are administered. Accordingly, patients with low risk of respiratory tract colonisation with Gram- negative bacteria on ICU-admission, such as those with little co-morbidity and directly admitted to the ICU without long-term hospitalisation (e.g., patients with trauma, those needing complex surgery or patients with acute neurological diseases) could potentially benefit from prophylactic antibiotic use. To the best of our knowledge this concept has been empirically evaluated only once. Sirvent and co-workers randomised 100 mechanically-ventilated patients with structural coma to a short course of prophylaxis with cefuroxime (two dosages of 1,500 mg 12 h apart after intubation) or no prophylaxis [41]. If colonised during intubation, most patients were colonised with early-onset pathogens. Cefuroxime prophylaxis was associated with a reduction in the development of microbiologically-confirmed ventilator-associated pneumonia (VAP) from 48% in control patients to 23% in the study group. Sixty-four percent of the VAP episodes were caused by methicillin-sensitive *S. aureus*, *H. influenzae* and *S. pneumoniae*. In another study previous (short-term) antibiotic treatment in mechanically ventilated patients with head injury appeared protective against initial tracheobronchial colonisation with early-onset pathogens, but increased the risk for subsequent lower airway colonisation with late-onset pathogens [42]. The other consequence of our findings is that the approach of systemic antibiotic prophylaxis is less likely to be effective for late-onset pathogens, even with antibiotics for which these pathogens are, in-vitro, susceptible.

The reason for the observed difference between early-onset and late-onset pathogens, though, remains to be determined. Pathogen-specific characteristics, such as the ability of bio-film formation of *P. aeruginosa* or patient-specific characteristics, such as the severity of underlying disease or immune paralysis could all be involved, but further - more detailed - studies are needed to elucidate this matter.

[3]

Bayesian Network Models for Diagnosing Ventilator-associated Pneumonia

3.1 Introduction

Ventilator-associated pneumonia (VAP) is the most frequently occurring nosocomial infection among mechanically ventilated patients in Intensive Care Units (ICUs). Reported incidence rates of VAP have ranged from 5% to 67%, depending on the selection of patients and the criteria used to establish its diagnosis [43].

In the absence of a clinically available gold standard, VAP is usually diagnosed upon a combination of criteria, such as systemic signs of infection, abnormalities on chest radiograph and microbiological identification of pathogens. However, each of these criteria combines high sensitivity with low specificity. In an attempt to raise diagnostic accuracy, algorithms, such as the Centers of Disease Control definition for nosocomial pneumonia and the Clinical Pulmonary Infection Score (CPIS) have been developed [44, 45]. Furthermore, invasive diagnostic techniques, such as bronchoalveolar lavage (BAL) and protected specimen brush (PSB), and quantitative analysis of microbiological cultures have been proposed to better distinguish between colonisation and infection of the respiratory tract [46, 47, 48]. Nevertheless, bronchoscopy includes a certain, though probably small, risk for complications and quantitative culturing is expensive and labour intensive. As a result, these techniques have not become standard of care in most ICUs.

Nowadays, patient records have become fully computerised and linked to hospital information systems in many ICUs. These systems contain all relevant variables on which intensive care physicians base their clinical decision. In such a setting computerised decision-support systems may assist in

This chapter is based on work published in Intensive Care Medicine, 2007.

diagnosing VAP. For these purposes, a Bayesian approach seems most attractive, because these models are able to deal with uncertainties, like missing values, can be improved by self-learning and can be linked to computerised patient data systems [49]. We recently developed such a Bayesian decision-support system (BDSS) for diagnosing VAP [50]. A BDSS consists of a Bayesian network and an inference engine, containing a mechanism for reasoning with, for example, patient data for the purpose of diagnosing a certain disorder. Formally, a Bayesian network $\mathcal{B} = (G, \Pr)$ is a directed acyclic graph $G = (V(G), A(G))$ with set of vertices $V(G) = \{V_1, \ldots, V_n\}$, representing stochastic variables, and a set of arcs $A(G) \subseteq V(G) \times V(G)$, representing statistical dependences and independences among the variables. On the set of stochastic variables, a joint probability distribution $\Pr(V_1, \ldots, V_n)$ is defined that is factorised respecting the (in)dependences represented in the graph:

$$\Pr(V_1, \ldots, V_n) = \prod_{i=1}^{n} \Pr(V_i \mid \pi(V_i))$$

where $\pi(V_i)$ stands for the variables corresponding to the parents of vertex V_i. The aim of the present study was to evaluate the performance of the BDSS, both in the daily routine assessment of these patients and in the subgroup of patients with clinically suspected VAP. Its use for daily assessment (approach 1) might support physicians to withhold antibiotics on days that the likelihood for VAP is low. The assessment of the BDSS on the days of clinical suspicion of VAP (approach 2), might increase the diagnostic accuracy, thereby reducing the likelihood that antimicrobial therapy is withheld for patients with VAP.

3.2 Methods

3.2.1 Setting

All patients admitted to two ICUs (a ten-bed medical ICU and an eight-bed neurosurgical ICU) of the University Medical Center Utrecht (UMCU) between January 1st 2000 and January 1st 2003 were included in this prospective cohort study. The UMCU is a 1062-bed tertiary care hospital. As both ICUs are equipped with patient data management systems all relevant clinical, laboratory and microbiological variables are registered on a daily basis. Indications for antibiotic use and interpretation of chest radiographs by radiologists were manually added to the database. Patients not receiving mechanical ventilation and patients with chronic home mechanical ventilation were excluded from analysis. As no intervention was evaluated and daily care was not influenced by the study, the institutional review board waived the necessity of informed consent.

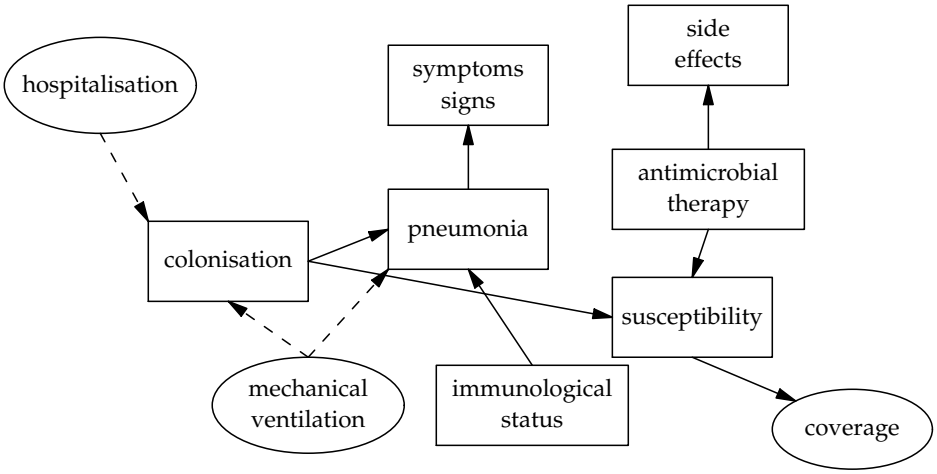


Figure 3.1: Global structure of the Bayesian network.

3.2.2 Bayesian decision-support system

A Bayesian network (BN) is a graphical representation of a process (in our case VAP) involving a set of variables that is based on probability theory and consists of nodes (containing tables with probabilities) and arrows, indicating the causal relationship between the nodes. Bayesian networks are often used in fields where decision-making occurs under uncertainty, like in health care. The Bayesian network underlying the BDSS for diagnosing VAP incorporates (conditional) probability distributions that were initially assessed subjectively, that is, either estimated or obtained from scientific literature (See global structure of the Bayesian network in Figure 3.1). Subsequently, these estimates were validated upon the dataset of this study and updated where needed using machine learning techniques [51]. The probability distributions for the model were derived from a randomly selected subset of the data and the diagnostic performance of the new model was assessed by testing it on the remainder. This process was repeated several times to rule out chance and prevent for overfitting, and the model with the average performance was considered our new model.

The diagnostic part (shown in detail in Figure 3.5) – as provided by two infectious disease specialists – of the BDSS uses the variables body temperature, blood leukocyte count, infiltrative abnormalities on chest X-ray ('Radiological signs'), pO_2/FiO_2 ratio, sputum production, sputum colour and duration of mechanical ventilation. The definitions of these variables are summarised in Table 3.1. Depending on these variables and probability distributions, the BDSS calculates a likelihood of VAP that ranges from 0 to 100%. With the exception of information regarding chest radiographs and tracheal aspirate cultures, which had to be added manually, all information required for the BN was automatically retrieved from patient information systems.

Table 3.1: Definitions of variables used in the BDSS. Each variable relates to the period within 24 hours before time of VAP likelihood prediction.

Variable	Values	Classifications	Details
Body temperature	$<36.0^{\circ}C$ or $>38.5^{\circ}C$	'abnormal'	
	otherwise	'normal'	
Sputum production [45]	$\geq$14 points	'abnormal'	Scoring points of sputum production per ICU day: 'no' = +0 'little' = +1 'moderate' = +2 'much' = +3
	otherwise	'normal'	
Sputum colour [52]	'yellow' or 'green'	'purulent'	
	otherwise	'non-purulent'	
pO_2/FiO_2 ratio	$\leq$205 mmHg or absolute decrease of $>$35 mmHg compared to the preceding day	'abnormal'	
	otherwise	'normal'	
Antipyretic drugs	acetaminophen or non-steroid anti-inflammatory drugs, steroids	'yes'	
	otherwise	'no'	
Chest X-ray	localised or diffuse infiltrate	'abnormal'	
	otherwise	'normal'	
Blood leukocyte count	$<4 \cdot 10^9/l$ or $>11 \cdot 10^9/l$	'abnormal'	
	otherwise	'normal'	
Mechanical ventilation	0 – 48 hours	'0–48h'	Excluded
	48 – 96 hours	'48–96h'	
	96 – 144 hours	'96–144h'	
	$>$ 144 hours	'$>$144h'	

3.2.3 Learning probabilities

One of the attractive features of Bayesian networks is that it is possible to combine information from various sources, for example starting with defining a probability distribution from one source, and then refining it using data [18].

There are two aspects to learning Bayesian networks from data:

- *Structure learning*, meaning that the structure of the network is learnt from data. This amounts to extracting information on statistical dependence and independence among variables from the data.
- *Parameter learning*, which means that the probability distribution is estimated from the data.

These two aspects of learning are to some extent intertwined, as the relevance of a structure can only be assessed in terms of a probability distribution of the structure, or a substructure, given the data. However, the usual order of affairs is that first the structure is learnt, after which the probability distribution of a given Bayesian network structure is learnt given data. The joint probability distribution of each of the models was estimated from the dataset using *Dirichlet learning*:

$$\Pr(V_i \mid \pi(V_i), D) = \frac{n}{n+n_0}\widehat{\Pr}_D(V_i \mid \pi(V_i)) + \frac{n_0}{n+n_0}\Theta_i \tag{3.1}$$

where $\widehat{\Pr}_D(V_i \mid \pi(V_i))$ are local probabilities computed for each variable V_i in the model, conditioned on the parent variables $\pi(V_i)$; the prior Θ_i was assumed to be a uniform probability distribution with equivalence sample size $n_0 = 5$.

In machine learning, usually conditional probabilities are learnt from data, without accounting for available prior knowledge, for example knowledge provided by domain experts. In this approach, we try to combine both sources, as expert knowledge, in our opinion, should not be excluded from learning process to make the qualitative part of a knowledge base as representative for a specific domain as possible.

3.2.4 Naive Bayesian networks

Since the 1960s several research efforts on automated reasoning with uncertainty for diagnostic problem solving in a clinical setting were undertaken. The systems constructed in this period were based to a large extent on applications of a simplified version of Bayes' Theorem; the technique used in these early systems is called the naive-Bayesian method.

A *diagnosis* of a problem, given evidence (signs and symptoms in medicine) $s \subseteq \mathcal{S}$ about the problem, is seen as a solution $d \subseteq \mathcal{D}$ that *best* explains s. In a probabilistic setting 'best' is interpreted as 'yields the highest posterior probability'. In terms of probability theory, this corresponds to computing

$$d^* = \arg\max_{d \subseteq \mathcal{D}} \Pr(d \mid s)$$

Generally Bayes' theorem is used for computing this probability:

$$\textit{posterior of hypothesis} = \frac{\textit{conditional likelihood} \cdot \textit{prior of hypothesis}}{\textit{prior of evidence}}$$

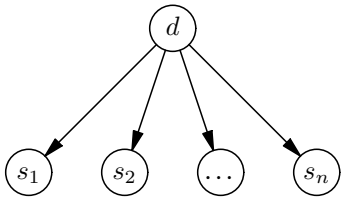


Figure 3.2: An example of a naive Bayesian network

or as a formula:

$$\Pr(d \mid s) = \frac{\Pr(s \mid d)\Pr(d)}{\Pr(s)} \tag{3.2}$$

In the typical diagnostic model, d is a single variable that can takes two or more values. It is called the diagnostic variable or *class variable*. By adopting specific independence assumptions, computation of $\Pr(d \mid s)$ can be considerably simplified. The assumptions follow from the directed graph shown in Figure 3.2. It is composed of a root node, i.e. a node with outgoing arcs only, and several leaf nodes, i.e. nodes with incoming arcs only. The diagnostic variable d corresponds to the root, whereas the evidence variables are leafs. The network expresses that each evidence variable s_i is conditionally independent of another evidence variable s_j, $i \neq j$, given the diagnostic variable d.
In this example, given a set of symptoms $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$, one wants to determine whether these symptoms give rise to a particular disease (d). In the naive Bayesian networks, using the independence assumption just mentioned, the general equation given above can be simplified to:

$$\Pr(d \mid s_1, s_2, \ldots, s_n) \propto \prod_{i=1}^{n} \Pr(s_i \mid d) \cdot \Pr(d) \tag{3.3}$$

3.2.5 Tree-augmented Bayesian networks

To represent dependence, with the simplicity of naive Bayes, we use the so-called tree-augmented Bayes which is defined by the following conditions:

- Each attribute has the class attribute, i.e. the root node as a parent.
- Attributes may have one other attribute as a parent.

By means of figure 3.2, the latter condition means that if there is an arc from s_i to s_j, and index $i \neq j$, the two symptoms are not independent given the class, i.e. the disease. Instead the influence of s_j on the class probabilities depends on the value of s_i. Figure 3.3 shows an example of a tree-augmented Bayesian network (TAN). A property of a TAN is that if we remove arcs from the class-node to the evidence-nodes, the evidence-nodes are all in the same tree-structure.

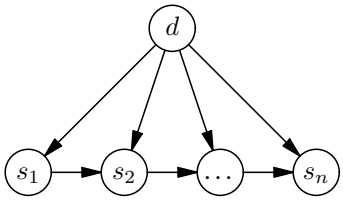


Figure 3.3: An example of an augmented Bayesian network.

3.2.6 Forest-augmented Bayesian networks

When TANs are constructing, arcs between several nodes were added, in others words: several dependencies are now present between variables. It is imaginable that there exist dependencies that are not really present. To alleviate this problem a Forest-Augmented Bayesian Network (FAN) algorithm was developed. This algorithm is a modified version of a previously-described algorithm for structure learning [51]. It computes mutual information for all pairs of evidence present in the database given the class vertex, i.e., VAP. A minimal-cost spanning forest, i.e., generalising the concept of a spanning tree, is constructed upon an undirected graph with costs attached to each edge. Next, the undirected forest is transformed into a directed forest by choosing a root vertex for every tree in the forest, and by adding an outward direction to the branches encountered on the paths from the root to every other vertex in the tree. The resulting directed forest is transformed into a connected directed graph by adding an arc (directed edge) from the class vertex to every evidence vertex in the forest [53].

3.2.7 Data used for learning

Table 3.2 gives a description of the clinical data of patients who were mechanically-ventilated for >48 hours, that were used for the machine learning process. Cross-validation was used for evaluation, i.e. our data set was split up: one partition was used to learn the conditional probabilities, and the classification performance of the model was tested on the remainder partition. This process was repeated ten times to rule out chance and to prevent for overfitting, and the median best performing model was considered our new model.

3.2.8 Reference standard of VAP

In the absence of a reference standard for diagnosing VAP, a diagnostic decision tree was developed to categorise all clinically suspected episodes of VAP (See Figure 3.4). In this diagnostic decision tree aspects of different proposed diagnostic algorithms were combined [44, 45, 54]. Categorisation was performed retrospectively through independent adjudication by two reviewers taking all relevant clinical, microbiological and radiological criteria. Episodes of clinical

Table 3.2: Data description. mech: mechanical; abn: abnormal; IQR: interquartile range.

Diagnosis	**VAP** n = 157 patient days	**no VAP** n = 9265 patient days
abn. temperature	61%	32%
antipyretics	84%	72%
mech. ventilation (median [IQR])	7 [4-15] days	7 [3-17] days
abn. leukocytes	65%	49%
abn. sputum amount	76%	42%
abn. sputum colour	74%	41%
abn. pO_2/FiO_2 ratio	67%	41%
positive X-chest	75%	24%

suspicion of VAP (defined as days on which clinicians had prescribed antibiotics for presumed respiratory infections or infections without evident focus) were evaluated at the time of clinical suspicion as well as in the following three days. Disagreement in categorisation between both reviewers was resolved through discussion and consensus was achieved on all cases.

On the day of prescription of antibiotics fulfillment of criteria of clinical suspicion was determined (Figure 2). The criteria of possible VAP, probable VAP, definite VAP or definite absence of VAP were verified at day 3 in all patients with clinical suspicion of VAP. All episodes with *clinical suspicion*, *possible*, *probable* and *definite* VAP were considered true VAP.

3.2.9 Evaluation

The performance of each candidate network was evaluated using cross-validation, i.e, the database was split up in equal parts, and the performance of each network was determined by evaluating the results for each of the parts, after its underlying joint probability distribution was learnt (using Dirichlet learning) from the other parts. Cross-validation offers a good balance between the bias and variance of learning results.

We used a Receiver Operating Characteristic (ROC) curve to denote the diagnostic performance of each resulting FAN and TAN network. A ROC curve is used to assess the quality of the discriminatory power of a test using the sensitivity, i.e. the number of patients for whom both IDS and DSS diagnosed VAP, and the specificity, i.e. the number of patients for whom both IDS and DSS did not diagnose VAP. The optimal point on the curve, i.e. the point closest to (0,1) identifies the best trade-off between the sensitivity and specificity. An AUC of 1 represents a perfect performance, whereas an AUC of 0.5 represents a worthless test, i.e. not better than 'flipping a coin'. A rough guide for classifying the accuracy of diagnostic performance is the traditional academic

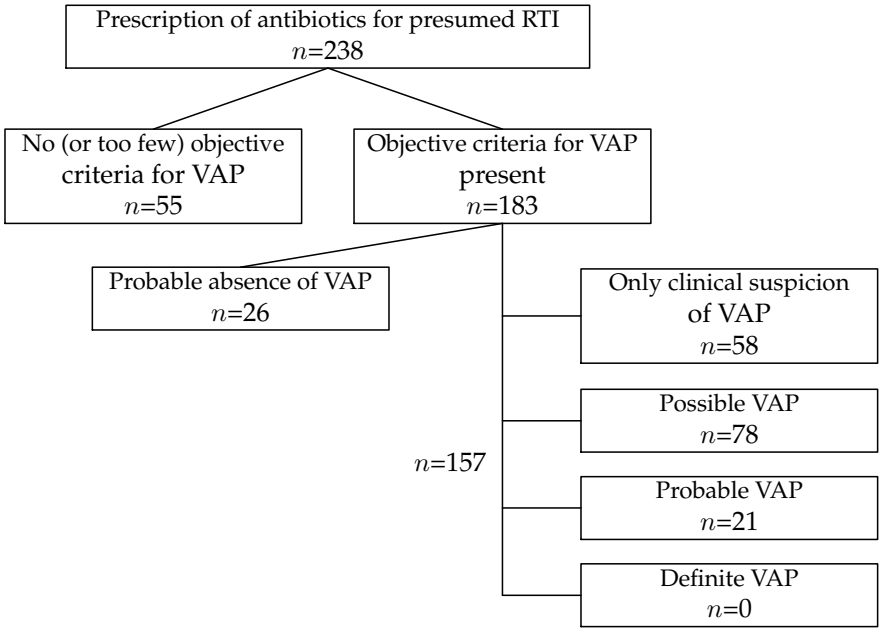


Figure 3.4: Decision tree for diagnosing VAP: number of episodes of presumed respiratory tract infections (RTI) and number of episodes according to the definitions of VAP.

point system [55] :

- $0.90 < \text{AUC} \leq 1.00$ = excellent;
- $0.80 < \text{AUC} \leq 0.90$ = good;
- $0.70 < \text{AUC} \leq 0.80$ = fair;
- $0.60 < \text{AUC} \leq 0.70$ = poor;
- $0.50 < \text{AUC} \leq 0.60$ = fail.

3.2.10 Statistical analysis

For data analysis two approaches were pursued. Test characteristics of the BDSS were determined for a situation in which all VAP-positive and all VAP-negative patient days (n=9422) were included, as well as for the situation in which only the days that antibiotics were prescribed because of presumed respiratory tract infection (n=238) were included. In the latter analysis, VAP diagnosis was confirmed, according to our reference standard, on 157 days.

Daily predictions of the likelihood of VAP by the BDSS were retrospectively compared to the reference standard of VAP. Data were expressed as absolute numbers with percentages and as means or medians with standard deviation or ranges. Independent variables were compared by Mann-Whitney U-test or t-test, when appropriate, and Wilcoxon signed rank test was used for non-independent variables. A probability value less than 0.05 was considered

Table 3.3: Patient characteristics.

Number of admissions	909
Number of patients	872
Number of patient days	9422
Age (mean, $\pm$ SD)	54 $\pm$18
Male sex (n,%)	507 (58)
APACHE	n=752
mean $\pm$ SD	24 $\pm$ 7
ICU mortality (%)	25.9
Day of admission ICU (mean $\pm$ SD)	13.5 $\pm$ 14
median [range] (days)	8.6 [2–109]
Days with antibiotics (n,%)	5558 (59)
Diagnosis on admission:	
Trauma	21
Neurotrauma	110
Intracerebral bleeding	162
Other neurological disease	85
Sepsis	58
Pneumonia	154
Post-operative	93
Auto-intoxication	9
Cardiac arrest	49
Medical diseases	145
Other	23

statistically significant. Test characteristics of the BDSS were analysed using Receiver Operator Characteristic (ROC) curves and by calculation of the Likelihood Ratio (LR) [56]. SPSS statistical software for Windows (version 12) was used to perform these analyses.

3.3 Results

In all, 872 patients (909 admissions) were included with a total of 9422 patient days (See Table 3.3). Antibiotics were prescribed for presumed respiratory tract infection on 238 days (Figure 2). Based on the diagnostic decision tree, 157 episodes (66%) of VAP were diagnosed, subdivided in *clinical suspicion* (n=58), *possible* (n=78) and *probable* (n=21). There were no cases of definite VAP. The overall incidence of VAP was 18% with an average incidence per patient-day of 1.7%.

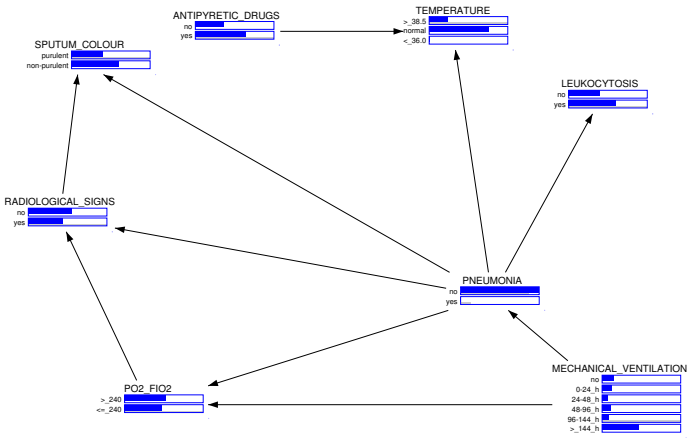


Figure 3.5: Expert model: diagnostic part.

3.3.1 FAN learning

Results and performance for three of the models are discussed here. They were:

(1) An expert-based Bayesian network (expert model) with subjective estimates of probability distributions (see Figure 3.5)

(2) The conditional naive Bayesian network (this can be seen as the base-line model) (see Figure 3.6)

(3) An extended naive Bayesian network model, inspired by expert knowledge (see Figure 3.7). The association between the administration of antipyretics, i.e., fever-suppressing drugs and body temperature was added.

Figure 3.8 shows a ROC curve for each of the three models. Area Under the Curve (AUC) was computed for each model and is summarised in Table 3.4. The performance of the network as elicited by the infectious-disease experts, expressed as AUC, is 0.83 whereas model performances of model 2 and 3 were 0.85 and 0.87, respectively. Thus, performance of these networks is considered 'good' according to this system. However, model 3 (shown in Figure 3.7) outperforms the other models.

3.3.2 Parameter learning

The results of different folds in performing cross-validation for the best performing diagnostic network structure (as shown in Figure 3.7) that was embedded in the whole Bayesian network, are shown in Table 3.5.

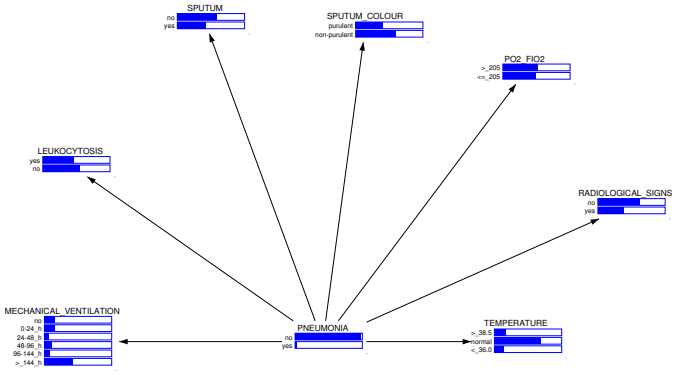


Figure 3.6: Conditional naive model.

Table 3.4: ROC analysis.

Model		AUC
1 (Fig. 3.5)	Expert	0.83
2 (Fig. 3.6)	Naive	0.85
3 (Fig. 3.7)	Extended Naive	0.87

3.3.3 Performance of the final model

Using the Bayesian network model underlying the BDSS with structure as depicted in Figure 3.7) and median performance, i.e., fold 6 in Table 3.5, test characteristics were the following. BDSS correctly identified the days that VAP was diagnosed. Median likelihood prediction of VAP for these 157 days was 77% [Interquartile range (IQR)= 56–91%], as compared to 14% [IQR 5–42%] for the 9265 days with absence of VAP ($p<0.001$). Likelihood predictions differed between the different levels of diagnostic certainty of VAP, being: 72% [IQR 39–90%] for *clinical suspicion*, 89% [IQR 66–92%] for *possible* and 88% [IQR 47–96%] for *probable* VAP. The ROC curve for BDSS predictions and VAP (no or yes, 157 episodes) had an area under the curve (AUC) of 0.857 [0.827–0.888 95% CI] (Figure 3), with an optimal cut-off point for diagnosing VAP of 46% (i.e., a prediction >46% would be considered VAP). This cut-off point had a sensitivity and specificity of 80%, with a positive predictive value (PPV) of 6.1% and a negative predictive value (NPV) of 99.6%. For this cut-off point the likelihood ratio for a positive test (LR+) was 4.0 . The AUC was somewhat higher for episodes fulfilling criteria of *possible* (AUC=0.884 [0.842–0.925 95% CI], optimal cutoff=52.6%) and *probable* VAP (AUC=0.875 [0.804–0.945 95% CI], optimal cut-off=53.0%), respectively. For episodes fulfilling criteria of *clinical suspicion* the

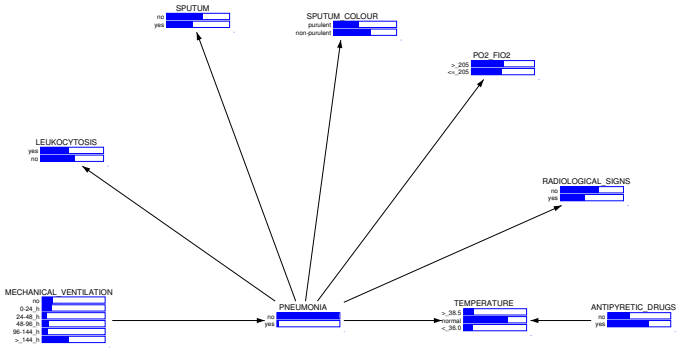


Figure 3.7: Extended naive model.

AUC was 0.818 [0.765–0.871 95% CI], with an optimal cut-off point of 40.6%.

A matched-cohort analysis was performed to evaluate to what extent development of VAP was associated with changes in BDSS predictions in time. For this, a patient with VAP was matched to 3 control patients, who had not developed VAP, with a similar duration of mechanical ventilation on the days of matching, similar gender and similar ICU ward. BDSS predictions on the day of matching (day 0) and days 1, 2 and 3 were compared between cases and controls. In the days before VAP was diagnosed median BDSS likelihood predictions increased from 28% on day 3 to 35% on day 2, 35% on day 1 and eventually 77% on day of VAP. In contrast, only small changes in likelihood predictions were observed in control patients (median values ranging from 16% to 20%) (Figure 4).

In the second approach (238 patient days of antibiotic prescription, with 157 episodes of VAP), the optimal cut-off point for VAP increased to 78%, with sensitivity and specificity of 79%, PPV of 87% and NPV of 66%. In this analysis the AUC was 0.846 [0.794–0.899 95% CI], which hardly changed when using only episodes of *possible* (0.853 [0.791–0.916 95% CI]) or *probable* VAP (0.875 [0.804–0.945 95% CI]). The LR for a positive test in this approach was 3.8 .

3.4 Conclusions and discussion

A Bayesian decision-support system, linked to computerised patient information systems in the ICU, accurately predicted absence and presence of VAP. Two approaches for application were analysed. When used on a daily basis, BDSS predictions had extremely high negative predictive values, and might be a helpful tool to reduce unnecessary antibiotic use. Yet, positive predictive values were, because of a low pre-test probability, also low. When restricted to the days that physicians prescribed antibiotics for presumed respiratory tract

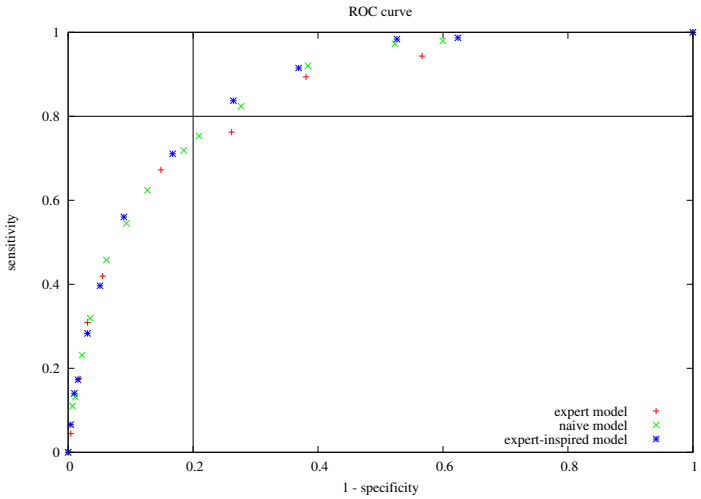


Figure 3.8: ROC curves.

infections, BDSS predictions had excellent test characteristics and a positive predictive value of 87%. Computerised decision-support systems might become accurate, relatively cheap, patient friendly and safe adjuncts for intensive care medicine.

The construction of our model has been based, as much as possible, on the available scientific evidence on the pathophysiology of VAP. Duration of exposure (ICU-stay and mechanical ventilation) and colonisation of the upper respiratory tract were considered the most important risk factors for VAP [57, 58]. In contrast, gastric colonisation and contamination of ventilatory circuits were not included in the model, as there is cumulating evidence that gastric colonisation is less relevant than colonisation of the upper respiratory tract [59, 60] and that frequently changing ventilatory circuits, in attempt to minimise the risks of contaminated equipment, has not been demonstrated to prevent VAP [61].

To the best of our knowledge, other BDSSs to diagnose VAP have not been described or evaluated before. Previously described BDSSs for infectious diseases considered choosing empirical therapy for bacterial infections [62], predicting the pathogens of bacteraemia originating from the urinary tract [63] and from other sites [64] and diagnosing community-acquired pneumonia (CAP) [65, 66]. This BDSS for diagnosing CAP was retrospectively evaluated in 41.371 patients that had visited an emergency department and for which a discharge diagnosis was known. Five hundred fifty three patients were diagnosed with CAP. The BDSS had a sensitivity for diagnosing CAP of 95%, with a specificity of 96.5%, an area under the receiver operating characteristic curve of 0.98, and a positive predictive value of 26.8%. This model is now being used

Table 3.5: Results for the different folds of the cross-validation parameter learning process sorted by increasing area under the curve(AUC).
CI = confidence interval; 'Cut-off' represents the point on the ROC curve that maximises sensitivity (Sens.) and specificity (Spec.).

Folds	AUC (95% CI)	Cut-off	Sens.	Spec.
1	0.691 (0.640–0.741)	0.188	0.595	0.584
2	0.766 (0.707–0.826)	0.980	0.692	0.698
3	0.783 (0.728–0.839)	0.318	0.722	0.720
4	0.816 (0.767–0.865)	0.277	0.731	0.712
5	0.825 (0.787–0.862)	0.429	0.763	0.761
6	0.857 (0.827–0.888)	0.463	0.796	0.799
7	0.860 (0.820–0.900)	0.382	0.772	0.796
8	0.865 (0.835–0.896)	0.770	0.790	0.798
9	0.896 (0.867–0.925)	0.319	0.821	0.826
10	0.958 (0.938–0.978)	0.134	0.899	0.895

in the emergency room to identify patients with CAP, but prospective data on performance, effects on patient care and outcome or cost-benefit analyses have not been reported to date.

The development and evaluation of our BDSS deserves some comments and one important limitation must be discussed. As all diagnostic studies for VAP, this study suffers from the absence of an available gold standard. An attractive alternative for the only true gold standard (i.e., histology) would be bronchoscopy with quantitative cultures. Yet, as this procedure was not common practice in our ICU, it could not be used for all patients. Therefore, we defined the decision tree (Figure 3.4) as used and prospectively categorised all patients with a clinical suspicion of VAP according to this scheme. Of note, this was performed before the BDSS model was used for calculation of likelihood predictions for VAP. Still, this does not exclude the possibility of misclassification. In the absence of a gold standard, a prospective randomised trial is needed to determine the clinical benefits, expressed in patient outcome, antibiotic use and costs of a new diagnostic test. In addition, we used the same dataset for validation and evaluation.

When used on a daily basis BDSS predictions had a PPV of 6.1%, which might seem low. Yet, with an overall incidence of 18% and average daily risk of 1.7%, a diagnostic test with sensitivity and specificity of 95% would still have a PPV of 25% only. Therefore, even good diagnostic tests can have poor PPV in low-prevalence settings [67]. The incidence rates as observed in our study are fully comparable to incidences reported in other studies [43, 68] and this problem, therefore, seems unavoidable when using a diagnostic test in daily practice. The problem of low PPV is avoided when restricting the diagnostic test to those days on which the physician has a clinical suspicion of VAP. In our setting, BDSS predictions had excellent test characteristics (AUC=0.846), but,

again, this is compared to our reference standard, which suffers from all the shortcomings mentioned.

The results show that the naive Bayesian network performs better than the expert model and that when the naive model is extended with expert knowledge performs even better in classifying patients who have VAP or do not have VAP. We showed that using machine learning techniques, i.e. structure and parameter learning from data, the diagnostic performance of a Bayesian network can be improved. The different Bayesian network models constructed by the TAN en FAN structure learning alogrithm (structures and results not shown) did not result in the best performing model. We conclude that expert knowledge can still result in a model for which the performance is superior to models with automatically-generated structures.

We, therefore, think that experts will not become useless in defining the structure and conditional probabilities for a given domain. The discussed algorithm could for example introduce arcs, i.e. dependencies between entities that would not be logical from a clinical point of view. Therefore, we combined expert knowledge with the best performing model according to the forest-augmented Bayesian network algorithm. The structural as well as the numerical part of the Bayesian network were updated, resulting in improved diagnostic performance of the Bayesian network for the management of VAP in Intensive Care Units.

The extended naive (Figure 3.7) and expert models (Figure 3.5) were investigated in order to study the effects of expert-based structure selection and probability estimation of a Bayesian network on classification performance. Normally, differences in structure influence the underlying probability distribution, due to the differences in modelled correlations; however, this does not necessarily imply that the model's classification performance will also be different (although one would wish that Bayesian networks designed by domain experts always perform better than networks extracted from data). Naturally, if both structure and probability distribution are based on expert knowledge, one would certainly expect some differences in classification performance, but again, it is hard to say beforehand how large those differences will be.

Validation of the model in other populations and prospective evaluation in a randomised trial are warranted. There are several explanations for false-positive test results. Because of the complex disease presentations in critically ill patients, there is always the possibility of other infections, also associated with systemic signs of infection. In our cohort, though, a comparison of the days with clinical suspicions of VAP to all other days on which no antibiotics were prescribed, reflecting absence of clinical suspicion of any infection, revealed similar sensitivity and specificity (data not shown). Another potential diagnostic problem is the sub-acute development of VAP, which may lead to false-positive predictions around the time of diagnosis. We, therefore, investigated whether patients developing VAP, already had higher prediction likelihoods in the days before diagnosis, in a matched-cohort analysis. And indeed, prediction likelihoods in patients developing VAP were already higher in the two days before diagnosis.

We have presented the first attempt to use a fully automated BDSS for diagnosing VAP. In two approaches (using it on a daily basis for all patients or only in case of clinical suspicion of VAP) the system performed very well. Next steps will be incorporation of pathogen prediction and guiding antimicrobial therapy. Yet the diagnostic properties of this model should be further investigated in other ICU-populations and, preferably, be compared to standard care in a randomised study design.

Part II

Dynamics in Diagnostic Parameters of VAP

[4]

Temporal Characteristics of ICU Patients

4.1 Introduction

Diagnosing disorders in patients is a temporal process that requires taking into account the appropriate signs, symptoms and laboratory test results. Very often, though, the dimension of time is kept implicit in the diagnostic process. From a clinical point of view, time is relevant in two different contexts: (1) the context of decision making, where a sequence of actions, such as ordering tests, possibly alternated by giving treatment is relevant; (2) the evolution of the disease process, in which, dependent of the disorder concerned, a different time scale comes into play. This is one of the reasons why a distinction is made between acute and chronic disease, and between early-onset and late-onset disease. Of course, these two mentioned aspects of time are closely linked to each other.

With respect to clinical decision making in time, different time scales can be distinguished:

- seconds: handling a simple or obvious disease by an experienced person, based on obvious signs and symptoms;
- hours: extra information obtained from laboratory tests is taken into account;
- days: after having ordered a radiological examination, a radiologist has to be consulted to learn whether abnormalities are present;
- weeks: a thorough examination is required in some disorders, such as cancer, and obtaining the final result may take a long time.

As was discussed in Chapter 1, there are various delays which may cause the times to be longer than needed.

Temporal evolution of disease is a process that, without treatment, has a timing determined by the nature of the disease, the health status of the patient, and sometimes, external conditions. This thesis is about infectious disease,

in particular ventilator-associated pneumonia (VAP), in which the natural history of the disease has a very strong temporal nature. A classical example is the natural history of lobular pneumonia, which students before the discovery of antibiotics had to learn by heart. However, also for modern disease management, the evolution of infectious disease is crucial, as knowledge of it has a major effect on the choice of the treatment.

In this chapter, we shed some light on the temporal evolution of ventilator-associated pneumonia, where time is again a major factor to be considered. Data of 157 patients diagnosed with VAP were analysed for this purpose, and a comparison was made with data from patients not diagnosed with VAP.

4.2 Bayesian-network model, VAP and time

The development of VAP is a process determined by time. This is an aspect which we have tried to capture in a Bayesian-network (BN) model shown in Figure 3.1, which incorporates the vertices *mechanical ventilation* and *hospitalisation* to model the influence of time. The first 48 hours a patient is mechanically ventilated, the probability of colonisation is low. After a period of two days, this probability is higher and as a result, the probability of developing VAP is also higher. Our database represents the situation for several patients from admission to discharge. We consider the period from admission to discharge of the patient as a *time series* $\langle X_t \mid t = 0, \ldots, n_p \rangle$, where $t = n_p$ is the time of discharge of patient p.

4.3 Methods

For exploring dynamic patterns in data of ICU patients, the following methods and materials were used.

4.3.1 Data of ICU patients

We used a temporal database of mechanically ventilated ICU patients for determining possible dynamics in the continuously distributed diagnostic parameters that indicate the development of VAP. For this purpose, three 'control' patients were matched to each of the 157 'case' patients diagnosed with VAP. Matching criteria were 'duration of mechanical ventilation', 'ICU ward' (medical and neurosurgical) and 'gender'. Table 4.1 shows mean $\pm$ standard deviation (SD) for the continues variables temperature, leukocytes, sputum production and pO_2/FiO_2 ratio from day t_{-4} until day of (matched) VAP, t_0.

Colonisation data were drawn from the hospital's medical microbiology files, as described in Chapter 2.

Table 4.1: Description of temporal data used: mean ± SD of continues diagnostic parameters. Dynamics from four days preceding the day of VAP or day of matched VAP were investigated. t_0 represents the day VAP was diagnosed.

Parameter	time point	**VAP** n = 157 'case' patients	**no VAP** n = 471 'matched control' patients
Temperature ($^\circ C$)	t_0	38.7 ± 0.8	38.1 ± 0.8
	t_{-1}	38.3 ± 0.8	38.0 ± 0.8
	t_{-2}	38.2 ± 0.8	38.0 ± 0.8
	t_{-3}	38.1 ± 0.8	38.0 ± 0.8
	t_{-4}	38.1 ± 0.8	38.0 ± 0.8
Leukocytes ($\cdot 10^9 l^{-1}$)	t_0	13.8 ± 6.7	12.4 ± 5.7
	t_{-1}	13.3 ± 6.5	12.2 ± 5.5
	t_{-2}	13.4 ± 5.8	12.4 ± 6.0
	t_{-3}	13.0 ± 6.2	12.7 ± 5.7
	t_{-4}	13.1 ± 6.0	12.7 ± 5.7
Sputum production (points [45])	t_0	22.7 ± 11.6	18.2 ± 11.0
	t_{-1}	18.5 ± 10.3	18.4 ± 10.5
	t_{-2}	16.9 ± 11.3	17.9 ± 10.8
	t_{-3}	14.4 ± 11.5	16.3 ± 12.0
	t_{-4}	13.8 ± 10.1	16.3 ± 12.0
pO_2/FiO_2 ratio (mmHg)	t_0	187 ± 78	210 ± 87
	t_{-1}	225 ± 114	214 ± 94
	t_{-2}	230 ± 100	219 ± 111
	t_{-3}	231 ± 95	220 ± 91
	t_{-4}	228 ± 98	216 ± 91

4.3.2 Colonisation patterns in the ICU

As described in Chapter 2, only a proportion of patients colonised with potentially pathogenic microorganisms will develop VAP, usually after a gradual change from asymptomatic colonisation to clinical infection. For patients for whom endotracheal culture results were known (positive/ negative) distinction was made between patients known to develop VAP and patients who do not. χ^2 tests were used to assess statistical significant differences in percentage of colonised (cultured) patients between the two groups.

4.3.3 Initial analysis of the Bayesian-network model

To test whether the model mentioned above behaved according to reality, a sensitivity analysis, known as one of the most powerful features in decision making, was performed. By definition, sensitivity analysis is a test of the stability of the conclusions of an analysis over a range of probability estimates, value judgments, and structural assumptions, which is useful especially for a

Bayesian network initially constructed upon expert estimates [56]. By performing such an analysis, the influence of the diagnostic variables on the likelihood of VAP could be investigated for the initially constructed, static Bayesian network.

In addition, it was investigated whether used the thresholds, as summarised in Table 3.1, for instantiating nodes in the network, such as for example

> IF number of leukocytes $> 11 \cdot 10^9 l^{-1}$
> THEN patient has *leukocytosis*

resulted in the best classification performance, compared to thresholds other than the original >11. Thus, for diagnostic variable x threshold t_x was validated by comparing the performance of the Bayesian network when using threshold t_x to the performance of the model when using different thresholds.

Furthermore, we investigated the relative contribution of each diagnostic variable to the posterior probability of VAP after parameter learning. Naturally, 'radiological signs' on a chest X-ray was presumed to be the most important variable for predicting VAP, as such an evident observation of the presence of infectious lung tissue is very specific for pneumonia. To objectify our hypothesis, increase or decrease in posterior probability of VAP when instantiating each diagnostic variable was reported. Here we assume that a patient is hospitalised for more than 5 days in the ICU and mechanically-ventilated for more than 144 hours.

4.4 Results

Having investigated dynamics and different pattern in dynamics involving VAP, the following results were obtained.

4.4.1 Colonisation patterns in the ICU

Figure 4.1 shows the distribution of the percentage of colonised patients over the first 30 days of mechanical ventilation in the ICU. During the first days of admission, the percentage of colonised patients increases, whereas, eventually, it stabilises at 60%. A third degree polynomial function was fit to the data points. Statistical differences were found for days 2–14. Thus, on days these the percentage of colonised patients differed between the two groups (VAP/ non-VAP). Note that bacterial persistence is assumed here, as described in 2.

Distinguishing between early-onset (Figure 4.2) and late-onset (Figure 4.3) pathogens, there is a clear difference apparent. Statistical differences in proportions were found at ICU days 2–10 and days 5–12 for early-onset and late-onset pathogens, respectively.

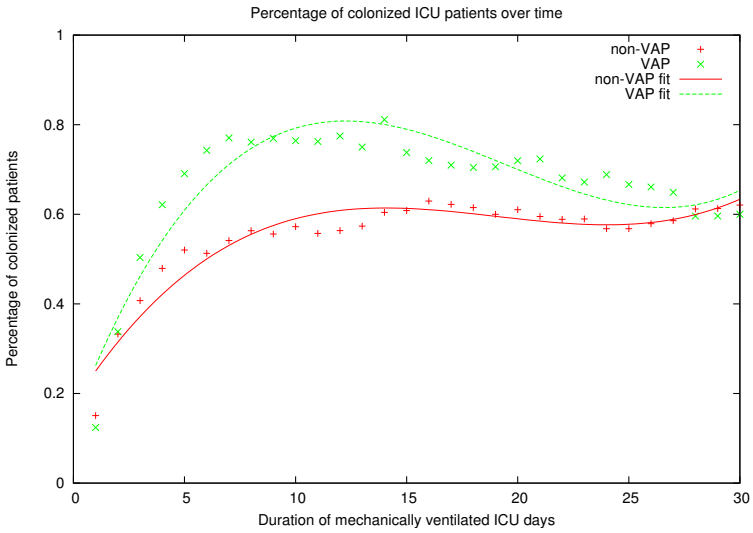


Figure 4.1: Colonisation in the ICU.

4.4.2 Initial analysis of the Bayesian-network model

By drawing graphs (with and without error bars) we illustrate the temporal behaviour of the average of some of the diagnostic variables. Characteristics of the data from 157 'case' and 471 'matched control' patients are shown in Table 4.1. The days before and after the day on which VAP was diagnosed as well as the days of matching are depicted in the following figures. In Figure 4.4, the amount of sputum production over time is depicted. It shows that the threshold of 14 points was most probably set too low, as the mean sputum production of the patients not diagnosed with VAP was already higher than 14 points. However, varying this threshold did not result in improved diagnostic performance of the Bayesian network. Figure 4.5. pO_2FiO_2 ratio decreased substantially on the day before VAP was diagnosed, which is depicted in Figure 4.6. Lowering the initial threshold of 240 mmHg to 205 mmHg and adding the condition that absolute decrease of $>$35 mmHg when comparing pO_2FiO_2 ratios between day t_{i+1} and preceding day t_i denotes lung function deterioration, improved the classification performance of the Bayesian network, resulting in the model described in Chapter 3. Blood leukocyte count, depicted in Figure 4.7, showed only small differences between VAP patients and matched controls.

From the assessement of the relative strength of each diagnostic parameter to the posterior probability of VAP, shown in Table 4.2, we can conclude that, indeed, positive radiological findings on the chest X-ray do have an increasing effect on the posterior probability of VAP. In addition, fever has a significant effect, whereas the number of leukocytes does not. The latter infection param-

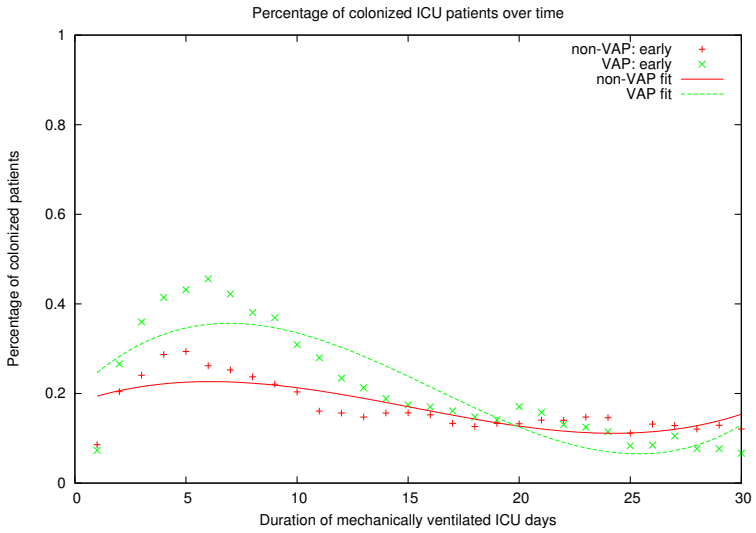


Figure 4.2: Early-onset colonisation in the ICU.

Table 4.2: Sensitivity analysis for the diagnostic part of the Bayesian network model. P(VAP | mechvent & hosp) = 34.1 6%.

Variable	Abnormal	Normal
Temperature	+ 22.57%	– 9.42%
Leukocytes	+ 2.64%	– 2.06%
Sputum amount	+ 15.47%	– 17.76%
Sputum colour	+ 18.66%	– 17.03%
pO_2/FiO_2 ratio	+ 11.15%	– 16.21%
Radiological signs (X-chest)	+ 26.73%	– 19.41%

eter proved to have low predictive performance on the development of VAP, as the majority of ICU patients has an increased leukocyte count, as already shown in Figure 4.7.

4.5 Conclusions

As physiological processes in the human body change over time, monitoring and interpreting them may be important. For example, continuously monitoring a patient's glucoses level in order to detect possible unexpected fluctuations. In intensive care units it is common practice to continuously measure clinical parameters like for example heart rate, blood pressure and ventilation frequency. It may be useful to monitor clinical symptoms and signs related to

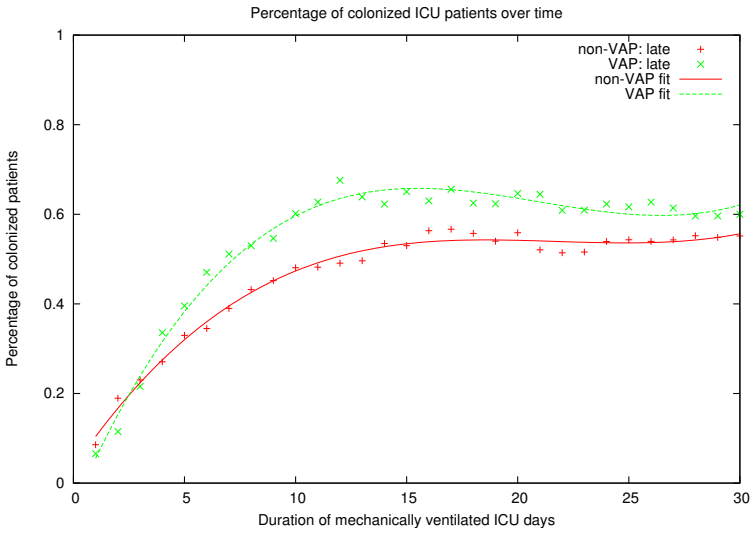


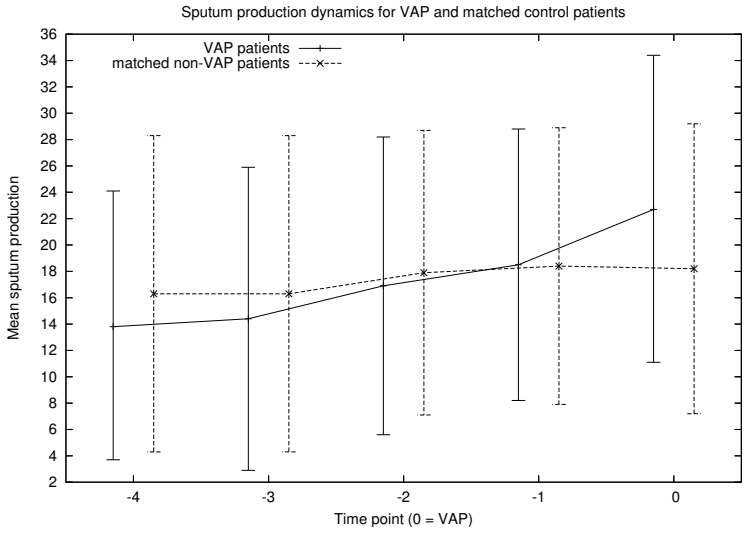
Figure 4.3: Late-onset colonisation in the ICU

the development of an infection like VAP, like, for example, an increasing body temperature. Therefore, we analysed the data of 157 patients diagnosed with VAP and did a comparison with data of patients not diagnosed with VAP.

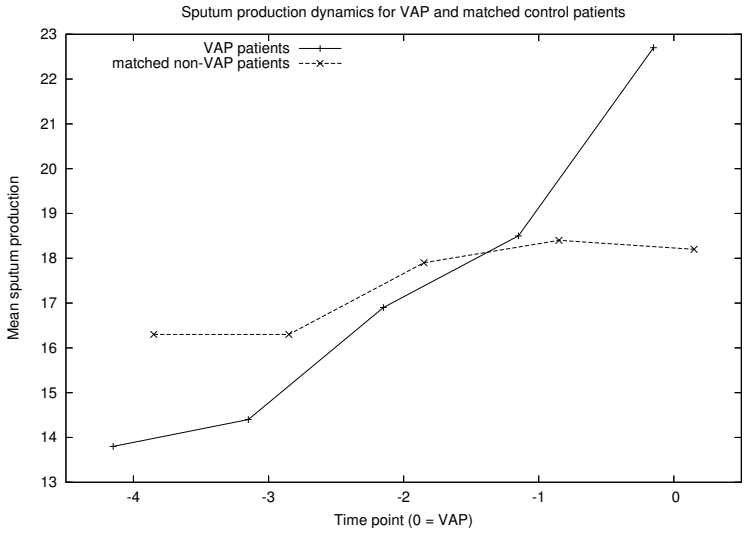
The development of VAP in patients is thought to be preceded by colonisation of the respiratory tract. Results showed, as expected, that the percentage of colonised patients known to develop VAP was higher than the percentage of colonised patients not developing VAP.

Differences in diagnostic variables between patients with and without VAP exist, though these are sometimes small. In addition, varying internationally recognised thresholds for abnormal values do not always result in a better classification performance of the Bayesian network model. It is known from ICU patients that, even when suffering from an infected tow, infection parameters such as number of blood leukocytes can be increased and, therefore, these parameters are regarded as not very specific for VAP. pO_2/FiO_2 ratio, representing the lung function, however, showed large temporal difference between both groups of patients. In addition, changing the threshold of this variable did improve the diagnostic performance of the BN model.

Though these results give more insight into the dynamic process of developing VAP, a fixed pattern, if any, for the predictive behaviour of the parameters involved is difficult to find.

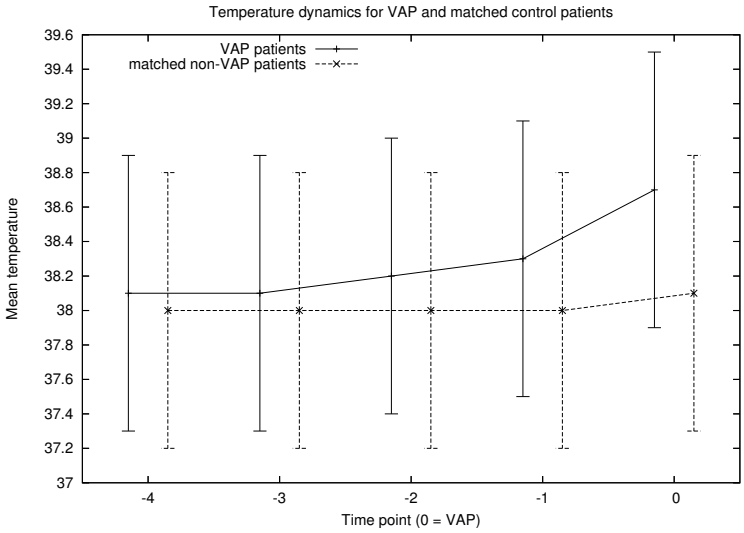


(a) Mean sputum production dynamics with error bars denoting the SD.

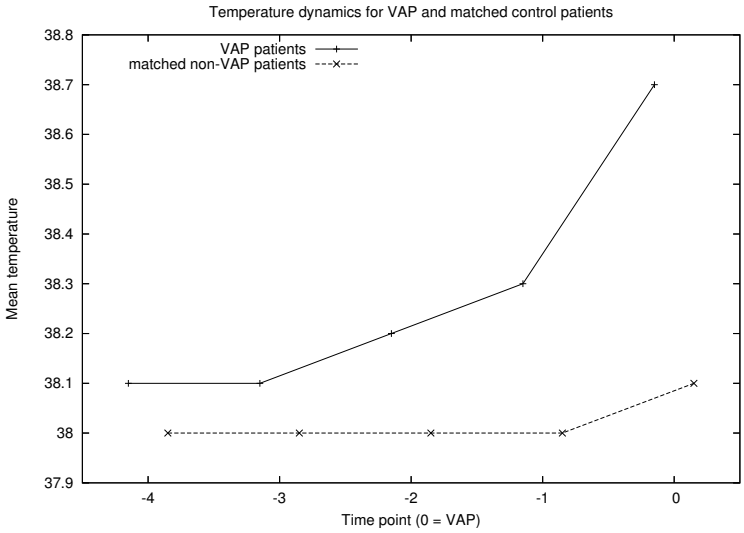


(b) Mean sputum production dynamics.

Figure 4.4: Sputum dynamics in ICU patients: VAP versus matched controls.

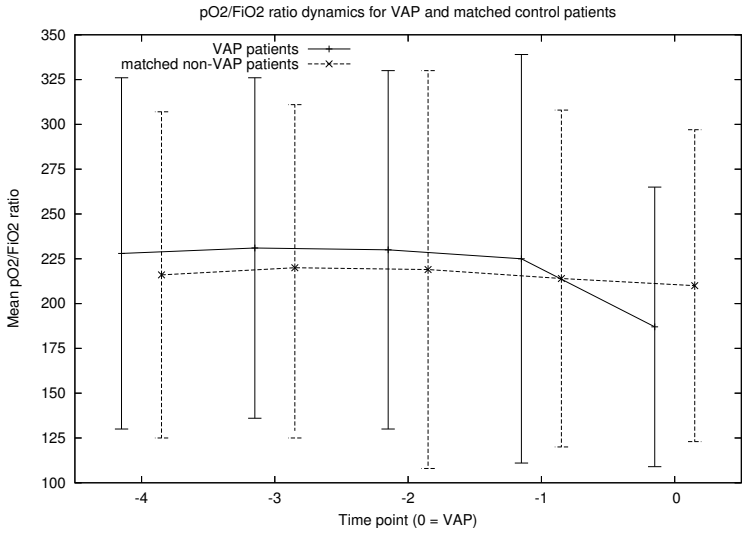


(a) Mean temperature dynamics with error bars denoting the SD.

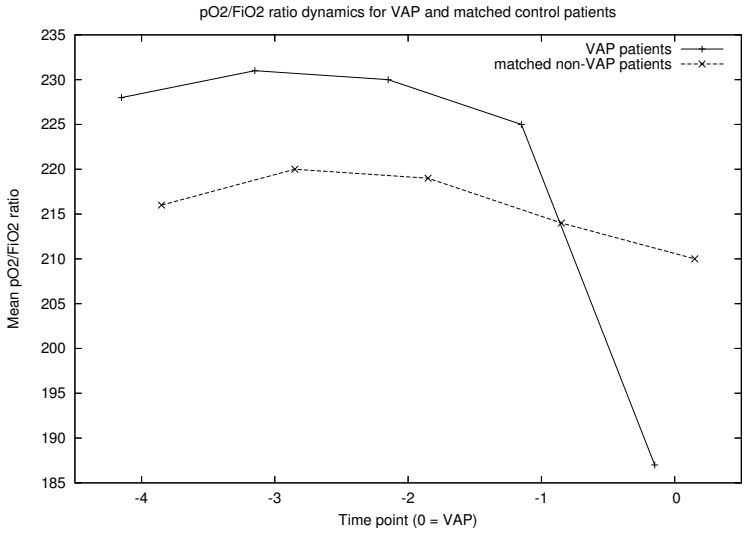


(b) Mean temperature dynamics.

Figure 4.5: Temperature dynamics in ICU patients: VAP versus matched controls.

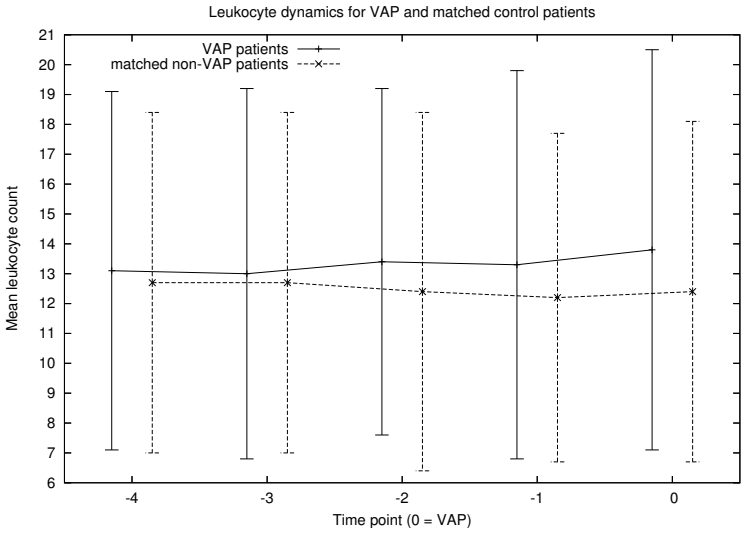


(a) Mean pO_2FiO_2 ratio dynamics with error bars denoting the SD.

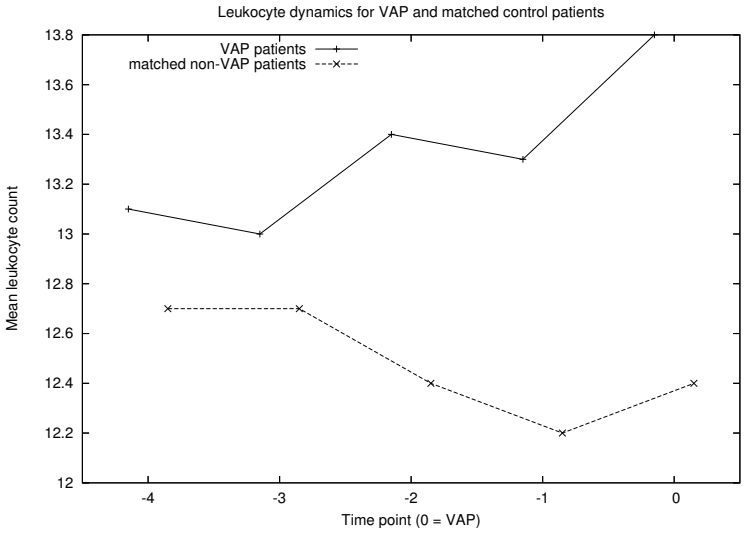


(b) Mean pO_2FiO_2 ratio dynamics.

Figure 4.6: pO_2FiO_2 ratio dynamics in ICU patients: VAP versus matched controls.



(a) Mean blood leukocyte count dynamics with error bars denoting the SD.



(b) Mean blood leukocyte count dynamics.

Figure 4.7: Blood leukocyte count dynamics in ICU patients: VAP patients versus matched controls.

[5]

The Exploratory Analysis of Disease Processes using Temporal Context-specific Independence

5.1 Introduction

Bayesian networks are statistical models that are well-suited for medical decision making, as for its output, can be understood in terms of the structure and probabilistic content of the Bayesian network [21]. Reasoning with a Bayesian network, i.e., filling in data and computing posterior probability distributions, may yield insight into the disease process, as well as directions in how to influence this by the selection of appropriate treatment. As the temporal nature of a disease process may also be relevant, *temporal* Bayesian networks are often selected for model construction. Such models can be used for temporal reasoning in clinical decision-support systems, the notion of time is taken into account [69]. Bayesian networks that ignore the notion of time are called *atemporal*.

So far, Bayesian networks have in particular been used for uncertainty reasoning in clinical decision-support systems; they have been used less frequently as tools for the analysis of clinical data, despite the availability of a wide range of Bayesian network structure and parameter learning techniques [22]. This is somewhat surprising as the statistical nature of Bayesian networks would them as useful as data-analytical tools as logistic regression, one of the main statistical tools of multivariate clinical data analysis.

There are several reasons why Bayesian networks, both atemporal and temporal, are used so rarely for data analysis in medicine: (1) in particular atemporal Bayesian networks are difficult to interpret, as the direction of the arcs is often counterintuitive; (2) whereas temporal Bayesian networks have the

The work described in this chapter has been submitted.

advantage that the direction of some of the arcs is in accordance to the order of time, their structure is usually restricted to being *repetitive* [70], which may not be compatible with the clinical problem at hand; (3) the conditional independences modelled in Bayesian networks only concern random variables and not their individual values; however, in medicine it is often the *context*, i.e., the specific values that random variables take, that determine how things relate to each other.

Temporal Bayesian networks that are repetitive are often called *dynamic* Bayesian networks [71, 72, 73, 74]. A dynamic Bayesian network consists of a specification of two timeslices: the first timeslice is an acyclic directed graph with associated prior probability distribution. The second timeslice consists of exactly the same acyclic directed graph as the first one, with the only difference that, rather than prior probability distributions, a collection of transition probability distributions is associated with the second timeslice, with variables conditioned on some of the variables in the first timeslice. Being a specification of a temporal Bayesian network, when reasoning with a dynamic Bayesian network, it is unrolled into a temporal Bayesian network with possibly infinite timeslices, where each timeslice contains exactly the same set of variables and graph structure, and the transition probabilities at different time instances are always the same. The latter property is known as *time invariance*; it is also said that the probability distribution is *stationary*. As they are attractive from a computational point of view, dynamic Bayesian networks, i.e. repetitive temporal Bayesian networks, have become the standard. However, dynamic Bayesian networks are unsuitable for elucidation of how a process evolves, and possibly adapts, in time. We, thus, propose to use non-repetitive temporal Bayesian networks in this paper, and investigate the usefulness of the formalism for data exploration for a clinical problem.

In addition, the use of context-specific independences and dependences is investigated, through *multinet* models [75].

In this paper, we use non-repetitive temporal Bayesian multi-networks in conjunction with context-specific independence information to analyse data. In doing so, we gained more insight in the evolution of a specific infection, by comparing such patients to others not developing this infection. We exploit a constraint-based learning algorithm, as these structure learning algorithms allow for the easy incorporation of medical background knowledge in the learning process. The ideas are illustrated by the analysis of temporal data of patients in the Intensive Care Unit (ICU), with or without ventilator-associated pneumonia (VAP). As only 10-15% of ICU patients will develop VAP, it was also necessary to exploit background knowledge in the learning process, as some clinically obvious relationships cannot be learnt from the data due to the sparsity of such data for a particular type of patient.

The chapter is organised as follows. In Section 5.2, Bayesian networks, temporal Bayesian networks and context-specific independence are briefly reviewed. Next, in Section 5.3, the basic theory underlying constraint-based structure learning is reviewed. In Section 5.4 exploration of dynamics in this cohort of patients is described. Finally, in Section 5.5 we discuss the results

achieved. The paper is rounded-off with some conclusions in Section 5.6.

5.2 Preliminaries

We briefly review the theory of temporal Bayesian networks, as discussed in more detail in [70]. Furthermore, the medical domain of ventilator-associated pneumonia, is described.

5.2.1 Temporal Bayesian networks

A *Bayesian network* $\mathcal{B} = (G, \Pr)$, BN for short, is a joint probability distribution $\Pr$ of a set of random variables X with an associated acyclic directed graph $G = (V, A)$, or ADG for short, where $\Pr$ is assumed to be decomposed into a set of conditional probability distributions in accordance to the structure of G. The random variables X and the vertices V have a 1–1 correspondence; thus we sometimes write X_W, $W \subseteq V$, for the random variables corresponding to the vertices W. Finally, $dom(X)$ denotes the domain of the set of random variables X (a Cartesian product).

Temporal Bayesian networks (TBNs) are an extension of ordinary Bayesian networks and allow for modelling uncertainty involved in processes regarding the dimension of time. Usually, a TBN is described in terms of a timeslice that has a fixed structure and is repeated several, possibly infinite, times, i.e., the TBN has a *repetitive* structure [76]. As mentioned above, often these repetitive temporal Bayesian networks are generated from a two-timeslice specification, and then they are called dynamic Bayesian networks. We, however, are convinced that disease processes are more complicated than that in the sense that independences may change over time and, therefore, a repetitive TBN would not suffice in every domain. This motivated some of us to develop a theory of modularisation of TBNs, with both repetitive and non-repetitive TBNs as special cases [70]. Evidence of the practical usefulness of non-repetitive TBNs has also come from work by Tucker et al. [77].

For the formal representation of the uncertain relations between variables over time, we need the following notions. Let T denote the (discrete and finite) time axis. Independence relationships between random variables with the *same* time point t are represented by means of an acyclic directed graph $G_t = (V_t, A_t^a)$, called a *timeslice*, with V_t denoting a set of vertices and $A_t^a \subseteq V_t \times V_t$ a set of *atemporal arcs*. Between timeslices, vertices corresponding to random variables may be linked to each other by means of so-called *temporal arcs*. Thus, a TBN consists of two parts: (1) an atemporal part (the timeslices), and (2) a temporal part. First, we consider the atemporal part.

Definition 1 ***(timeslice and atemporal arcs)*** *An ADG $G_t = (V_t, A_t^a)$, with the set of vertices V_t and the set of* atemporal arcs *$A_t^a \subseteq V_t \times V_t$, $t \in T$, is called a* timeslice *at time point t.*

The set of all timeslices G of a TBN is taken as:

$$G = \{G_t \mid t \in T\} = \{(V_t, A_t^a) \mid t \in T\} = (V_T, A_T^a). \tag{5.1}$$

Let G_t and $G_{t'}$, $t, t' \in T$, be two timeslices. Then, an arc $(u_t, v_{t'})$ with $t < t'$ is called a *temporal arc*. The set of temporal arcs of an ADG is denoted by A^t. Thus, temporal arcs connect timeslices with strict direction from the past to the future.

Definition 2 ***(temporal graph)*** *A* temporal graph N *is defined as a pair* $N = (V_T, A)$*, where* $G = (V_T, A_T^a)$ *and* $A = A_T^a \cup A^t$*, with* A_T^a *denoting the set of timeslices.*

Clearly, a temporal graph N is also an ADG. A *temporal Bayesian network* (TBN) is now defined as a pair $\mathcal{TBN} = (N, \Pr)$, where $\Pr$ is the joint probability distribution (JPD) on X_{V_T}.

5.2.2 Context-specific independences

Two sets of random variables X and Y are said to be *conditionally independent* given a third set of random variables Z, denoted by $X \perp\!\!\!\perp_{\Pr} Y \mid Z$, if it holds that

$$\Pr(X \mid Y, Z) = \Pr(X \mid Z)$$

if $\Pr(Y, Z) > 0$. If this does not hold, then X and Y are *conditionally dependent* given Z, which is denoted by $X \not\perp\!\!\!\perp_{\Pr} Y \mid Z$.

Conditional independence statements cannot only be represented in the form of probability distributions $\Pr$; they can also be read-off from the graphical structure of an associated ADG G using the notion of d-separation. Then, two disjoint sets of vertices A and B in G are said to be *d-separated* given a third disjoint set of vertices C, denoted by $A \perp\!\!\!\perp_G B \mid C$, if each (undirected) path from a vertex in A to a vertex in B is blocked by a vertex in C, taking into account paths with so-called v-structures (i.e., subgraphs of the form $u \rightarrow v \leftarrow w$), where the blockage is cancelled if either the vertex v or a descendant of v has been observed.

For Bayesian networks $\mathcal{B} = (G, \Pr)$, it holds that if $A \perp\!\!\!\perp_G B \mid C$ holds, then $X_A \perp\!\!\!\perp_{\Pr} X_B \mid X_C$ should also be satisfied. It is said that G is an *independence map* of $\Pr$. Similar, temporal and atemporal, notions of d-separation have been developed for temporal Bayesian networks, where the *atemporal d-separation* relationship $\perp\!\!\!\perp_G$ is defined for the part of the temporal Bayesian network where the temporal arcs are ignored, and *temporal d-separation*, denoted by $\perp\!\!\!\perp_{N_{|\Theta}}$, is defined by always taking into account at least one temporal arc when investigating blockage (for details, cf. [70]). Clearly, atemporal d-separation $\perp\!\!\!\perp_G$ can be defined in terms of atemporal d-separation for individual timeslices, i.e., in terms of $\perp\!\!\!\perp_{G_t}$, $t \in T$.

Despite the fact that temporal and atemporal notions of d-separation allow for the study of interesting independence patterns in temporal Bayesian

networks, we believe that many of these patterns are context specific, i.e., independence information may change for particular values of random variables. Formally, a set of variables Y is *conditionally context-specific independent* of a set of variables W given a third set Z in the context φ, written $Y \perp\!\!\!\perp_P W \mid Z; \varphi$, where φ is a nonempty set of random variables U with values u, i.e., $\varphi \equiv U = u$, if $\Pr(Y \mid W, Z, \varphi) = P(Y \mid Z, \varphi)$ and $\Pr(Y \mid W, Z, \varphi') \neq P(Y \mid Z, \varphi')$ for $\varphi' \equiv U = u'$, $u' \neq u$ [78]. For discrete random variables X with finite domain, it is possible to associate an ADG G^φ with every context φ. The result is called a *Bayesian multinetwork* $\mathcal{B} = (G, \Pr)$ with $G = \{G^\varphi \mid \varphi \equiv X = x, x \in dom(X)\}$. Temporal Bayesian multinetworks can be be defined along similar lines.

5.2.3 Ventilator-associated pneumonia

VAP is an infection of mechanically-ventilated ICU patients. Clinical symptoms, such as fever, are usually not very specific. Important symptoms and signs, providing evidence for the development of VAP, include *body temperature*, amount and colour of *sputum* production, radiological *signs* on the chest X-ray, duration of *mechanical ventilation*, number of *leukocytes* [27], and abnormal ratio between the arterial oxygen pressure and the fractional inspired oxygen level (pO_2/FiO_2-ratio).

5.3 Constraint-based structure learning

As only 10–15% of the long-term ventilated patients develop VAP, large cohorts are needed to collect sufficient data for analysis. If data are sparse, it will be difficult to learn independence relations from the data. Lack of data, though, can be compensated by augmenting the learning process through the exploitation of background knowledge. This is exactly what we have done. Learning algorithms that allow easy incorporation of background knowledge into the learning process are called *constraint based*. These algorithms derive a set of conditional independence statements from the data, taking supplied dependence and independence information as additional constraints, and build a structure with d-separation properties corresponding to the independence information available.

5.3.1 The NPC algorithm

One of the best constraint-based Bayesian network structure learning algorithms available is the Necessary Path Condition (NPC) It is a criterion that has been added to an earlier constraint-based algorithm, PC, by researchers at Siemens in Munich [79]. The algorithm is a variant of the CI algorithm by Verma and Pearl [80], and works as follows:

1. **Automatic phase:**
   (a) An undirected graph $H = (V, E)$, called *skeleton*, is derived through computation of the score $g_{\chi^2,\alpha}(X, Y, S)$, for pairs X, Y of random variables and the set of random variables S (with $X, Y \notin S$). The function $g_{\chi^2,\alpha}$ is based on the χ^2 test with *significance level* α [79]. Typically, one takes $\alpha \leq 0.05$. If $g_{\chi^2,\alpha}(X, Y, S) > 0$ then the conditional independence hypothesis $X \perp\!\!\!\perp_{\Pr} Y \mid S$ is rejected.
   (b) Modify subgraphs $X - Y - Z$ of H into $X \rightarrow Y \leftarrow Z$, if $X - Z \notin E$ and $X \not\perp\!\!\!\perp_P Z \mid S$, with $Y \in S$, using the same scoring function $g_{\chi^2,\alpha}$.
   (c) Orientate the remaining lines as to obtain arcs, where the creation of cycles in the resulting directed graph is avoided.
2. **User interaction phase:** To resolve inconsistencies in the conditional (in)dependence statements, the NPC algorithm, unlike the PC algorithm where in case of uncertain dependences directionality is chosen randomly, relies on user interaction where the user gets the opportunity to decide on the addition, removal and orientation of arcs. In addition, α can be arbitrarily chosen, so that lines with calculated p-value (called p below) larger than α are excluded.

In our domain, the direction of an arc has been determined by the use of background knowledge. By doing so, cause and effect can be distinguished. For example, when the NPC algorithm indicates a relation between 'VAP' and 'temperature', the most logical order is that 'VAP' should be parent of 'temperature', as when a patient suffers from VAP, normally, the temperature increases due to fever, and not the other way round. Also, it is possible to supply known relations at the start of the NPC learning process. By doing so, we were able to include constraints that have already been proved to exist and have been described in literature. We used the implementation of the NPC algorithm available in the Hugin tool set [81] for our research. This includes the EM algorithm for parameter learning [24].

5.3.2 Data

A dataset D with temporal data of ICU patients, containing 17710 records, was used. Each record represents data of one patient in the ICU during a period of 24 hours. The database contains 2424 admissions to the ICU. For 157 of these patient episodes, VAP was diagnosed by two infectious-disease specialists. From dataset D three subsets were created: D_{vap} containing data of all 157 VAP patients; $D_{\overline{\mathrm{vap}}}$ containing data of patients who were not diagnosed with VAP (so-called controls). Each patient with VAP was matched to 3 control patients, with a similar duration of mechanical ventilation on the days of matching. D_{VAP} contained data of both VAP and control patients, i.e., $D_{\mathrm{VAP}} = D_{\mathrm{vap}} \cup D_{\overline{\mathrm{vap}}}$. Each dataset contained data of 4 consecutive days, each representing a timeslice: t_0 was either the day on which VAP was diagnosed or the day of matching and t_{-3}, t_{-2}, t_{-1} were the three days preceding t_0.

5.3.3 Procedure of TBN construction

The construction of the context-specific and combined TBNs was performed as follows:

1. Using the NPC algorithm, under the author's supervision, the atemporal arcs between vertices in each separate timeslice were determined.
2. In the next run of the NPC algorithm, the temporal relationships of all variables were explored, taking into account the structure of the timeslices.
3. Medical background knowledge was sometimes employed to decide about the direction of arcs, or inclusion or deletion of arcs. However, this was only employed when the algorithm was unable to decide about the inclusion and direction of an arc. The direction of arcs was decided on by looking at the network in terms of cause–effect relationships. Expertise of a medical infectious disease specialist was used for that purpose.

5.4 Exploring dynamics in critically ill patients

The objective of this chapter was to explore data of ICU patients, in order to gain more insight into the dynamics of disease characteristics of patients developing and not developing VAP. Clinical parameters of patients can be very aspecific when using daily measurements. Therefore, we hypothesised that temporal (or dynamic) data mining should be prefered above static data mining approaches. But even when measuring clinical parameters dynamically, fluctuations will occur. For example, body temperature can fluctuate from day to day, depending on, for example, severity of illness and whether or not fever suppressing medication (antipyretics) is administered. Still, we seek for generalisable trends while exploring our data. A distinction is made between patients with and without VAP. By doing so, we can easily detect differences between these two groups of critically ill patients, denoted in the following by 'vap' and '$\overline{\text{vap}}$'.

The temporal and atemporal arcs of a non-repetitive temporal Bayesian network only indicate that variables influence each other to a sufficient extent to be included in the graph. However, no information is obtained about the strength of these dependences. It was, therefore, decided to explore addition of this information to the learnt temporal Bayesian networks. First, the level of significance α of accepting conditional dependence among variables was varied. Second, the dynamics of the change in difference between marginal probability distributions of variables in time slice $i+1$ and i, i.e., $\Pr(X_{i+1}) - \Pr(X_i)$. This measure provides insight into the development of the signs and symptoms as a function of time. Third, the temporal changes in the underlying conditional probability distributions, as determined by the EM-learning algorithm, were expected to yield insight into the probabilistic dynamics of the disease process.

Table 5.1: Qualitative interpretation of temporal probabilistic behaviour. **M**: marginal probability; **C**: conditional probability.

- - -	very strong decrease in **M** / **C** probability from i to $i+1$: $[\Pr(X_{i+1} = abnormal) - \Pr(X_i = abnormal)] < -0.20$
- -	strong decrease in **M** / **C** probability from i to $i+1$: $-0.20 \leq [\Pr(X_{i+1} = abnormal) - \Pr(X_i = abnormal)] < -0.10$
-	weak decrease in **M** / **C** probability from i to $i+1$: $-0.10 \leq [\Pr(X_{i+1} = abnormal) - \Pr(X_i = abnormal)] < -0.05$
	no change in **M** / **C** probability from i to $i+1$: $[\Pr(X_{i+1} = abnormal) - \Pr(X_i = abnormal)] \in [-0.05, 0.05]$
+	increase in **M** / **C** probability from i to $i+1$: $0.05 < [\Pr(X_{i+1} = abnormal) - \Pr(X_i = abnormal)] < 0.10$
++	increase in **M** / **C** probability from i to $i+1$: $0.10 \leq [\Pr(X_{i+1} = abnormal) - \Pr(X_i = abnormal)] \leq 0.20$
+++	very strong increase in **M** / **C** probability from i to $i+1$: $[\Pr(X_{i+1} = abnormal) - \Pr(X_i = abnormal)] > 0.20$

The significance of a dependence relation was indicated by making distinction between 'strong' and 'less strong' dependences, where the strength is indicated by the width of the arrow shafts in the graph. Strong dependence relations have a p-value less than significance level $\alpha_1 = 10^{-3}$, whereas all other relations have a p-value p with $\alpha_1 < p < \alpha_2$, with $\alpha_2 = 0.05$.

For the dynamics of the difference in the marginal probabilities as a function of time we are in particular interested in the changes in prior probabilities of the values denoting 'abnormal' situations. For example, for the variable 'temperature' we are interested in the value denoting the situation where a patient has fever, which is considered to be abnormal. Table 5.1 shows the meaning of the qualitative signs used to annotate each arc in the figures: the larger the number of plusses, the higher the increase in the marginal probability in the direction of the arc. This method was inspired by qualitative influences, which is part of the formalism of qualitative probabilistic networks [82]. The boundary values $-0.20, -0.10, -0.05, 0.05, 0.10, 0.20$ were experimentally determined.

We used similar qualitative signs as in Table 5.1 to indicate changes in the marginal as well as the conditional probability distributions, where the difference $\Pr(X_{i+1}) - \Pr(X_i)$ was replaced by $\Pr(X_{i+1} \mid \pi(X_{i+1}))$. Here, signs were only added for random variable with the same parents at each timeslice.

Finally, to gain more insight into the disease dynamics in critically ill patients, we selected two important diagnostic variables involved in the domain of lung function deterioration and compared the group of patients diagnosed with VAP with a group of patients without VAP. The chosen random variables were chest X-ray and pO_2/FiO_2. For these variables, the percentage of patients having 'normal', 'abnormal' or 'missing' values were visualised.

5.5 Results

Based on the three databases, D_{vap}, $D_{\overline{\text{vap}}}$ and D_{VAP}, three TBNs were constructed using the NPC learning algorithm. As described above, atemporal subgraphs were obtained separately, and then combined to a TBN by learning temporal arcs (taking into account the known timeslice structures). Sometimes uncertain arcs were removed (user interaction phase). For example, an arc between the variable *temperature* in timeslice t_0 and variable *leukocytosis* in timeslice t_{-2}, did not seem clinically relevant and was, therefore, excluded. In addition, we assumed that in all situations it holds that antipyretic drugs, such as paracetamol, always have a suppressive effect on a patient's body temperature. This background knowledge was accounted for when executing the NPC procedure, so the relation *antipyretics* $\rightarrow$ *temperature* was presented in every time slice automatically.

5.5.1 VAP Patients (D_{vap})

The timeslices (atemporal subgraphs) for the four different time points show different independences. For example, for t_{-3} an arc between *sputum* and *sputum-colour* ($p = 0.05$) was suggested by the NPC algorithm, whereas for t_{-1} and for t_{-2} that same relation was absent, but for t_0 it was again present. Also, an arc ($p = 0.02$) between *chest X-ray* and *pO$_2$/FiO$_2$* (as explained, a measurement of the lungs' functions) was often found, as well as between *temperature* and *sputum-colour* ($p = 0.05$). As the last arc was not considered clinically relevant, it was excluded from the models. All temporal arcs proved to have high significance ($p < 10^{-7}$). Combining the atemporal en temporal parts resulted in a TBN, called G_{vap}, shown in Figure 5.1, that includes all signs and symptoms describing the course of the development of VAP.

5.5.2 Patients not diagnosed with VAP ($D_{\overline{\text{vap}}}$)

The timeslices for the non-VAP patients were similar, but not identical, to those for VAP patients; in particular, arcs between *chest X-ray* and *pO$_2$/FiO$_2$*, and between *sputum* and *sputum-colour* were found. The only difference was that the strength of the arcs increased towards the time point matching the day of VAP, i.e., t_0, that is, $p(t_{-3}) \approx 10^{-2}$, $p(t_{-2}) \approx 10^{-3}$, $p(t_{-1}) \approx 10^{-4}$ and $p(t_0) \approx 10^{-5}$. Thus, $p(t_{-3}) > p(t_{-2}) > p(t_{-1}) > p(t_0)$. Temporal arcs were suggested between timeslices t_{-2} and t_{-1} for the variables *chest X-ray*, *sputum* and *pO$_2$/FiO$_2$* only. Moreover, these temporal relations proved to be less strong, i.e., $p \approx 0.01$, compared to the temporal relations in the context of VAP. Combining both temporal and atemporal structures resulted in TBN called $G_{\overline{\text{vap}}}$, shown in Figure 5.2 with again, $\overline{\text{vap}}$ representing the matched control patients.

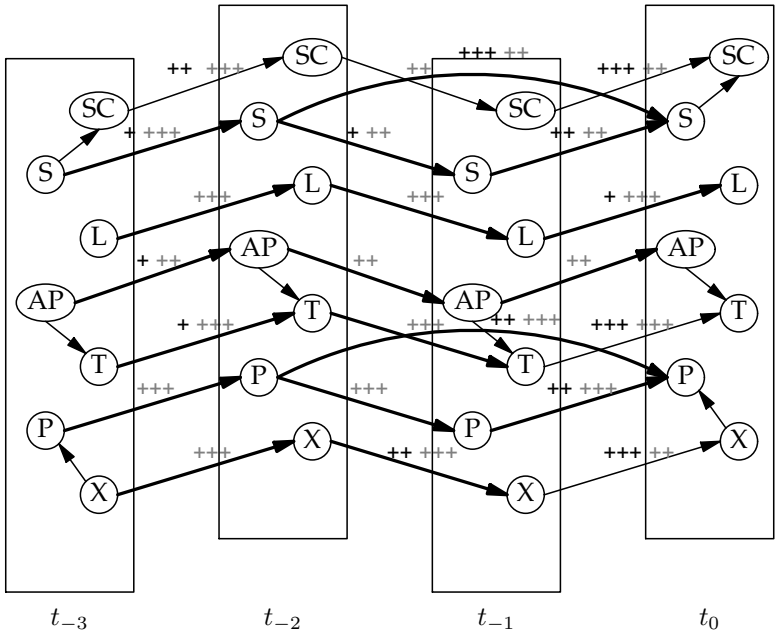


Figure 5.1: G_{vap} : Independences obtained for VAP patients. Abbreviations: SC: sputum colour; S: sputum; L: leukocytosis; AP: antipyretic drugs; T: temperature; P: pO_2/FiO_2; X: chest X-ray.

5.5.3 Patients with and without VAP (D_{VAP})

As a model only suitable for VAP patients would not be useful in practice, combining the two datasets mentioned above for building a TBN for VAP and non-VAP at the same time yields yet another view on structure learning. Structure learning based on the database including data of VAP as well as non-VAP patients resulted in a combination of the two TBNs G_{vap} and $G_{\overline{vap}}$, from here denoted by G_{VAP}. The temporal arcs were almost identical to those of G_{vap}, though less strong ($p \approx 10^{-4}$). The atemporal arcs had strong correlations and were, not surprisingly, found between the variables *chest X-ray* and pO_2/FiO_2 ($p \approx 10^{-3}$) and between *sputum* and *sputum colour* ($p \approx 10^{-3}$). In all, the temporal arcs again proved to be stronger than the atemporal arcs. The resulting model is shown in Figure 5.3. This TBN clearly shows that much of the clarity of the original context-specific TBNs was lost, and that it is no longer possible to gain insight into the development of VAP and non-VAP separately.

5.5.4 Probabilistic dynamics

We can learn from Figures 5.1 and 5.4 that, for example, in the days prior to the day at which VAP was diagnosed, the prior marginal probability of having fever increases (See also Table 5.1). From day t_{-1} to day t_0, there was even an

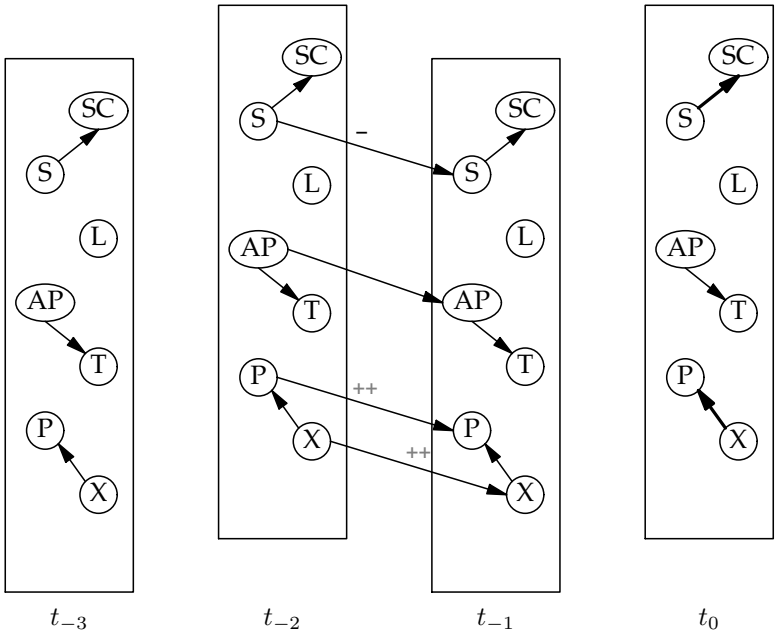


Figure 5.2: $G_{\overline{\text{vap}}}$: Independences obtained for patients *not* diagnosed with VAP. Abbreviations: SC: sputum colour; S: sputum; L: leukocytosis; AP: antipyretic drugs; T: temperature; P: pO_2/FiO_2; X: chest X-ray.

absolute increase of more than 20%, though this relation was less strong than the relations between prior days. In all, for all variables increasing prior probabilities towards the day of VAP as well as strong relations, except for *sputum colour* can be extracted from the model. In contrast, Figure 5.2 contains almost only relations that are considered less strong. In addition, very few temporal relations exist. Atemporal relations between *sputum* and *sputum colour* and between *chest X-ray* and pO_2/FiO_2 are obvious.

Table 5.2 shows the percentage of patients that have abnormal infection parameters, such as, for example, high body temperature. For patients not diagnosed with VAP, the percentage presenting abnormal clinical parameters remains relatively stable (leukocytosis, pO_2/FiO_2, X-chest). However, the variables concerning sputum amount and colour show a decrease of abnormal values on t_0.

As can be learnt from the temporal graphs, the random variables chest X-ray and pO_2/FiO_2 have significant temporal behaviour and are correlated. We, therefore, visualised the percentage of patients having 'normal', 'abnormal' or 'missing' values for these variables in Figures 5.5 to 5.8, which gives much more detailed insight into differences in the dynamics of signs and symptoms between patients with and without VAP.

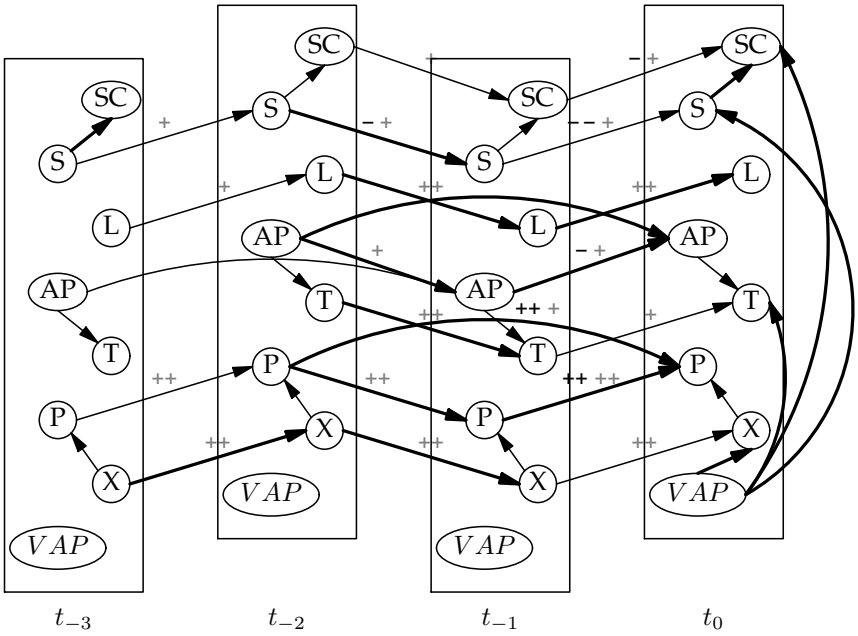


Figure 5.3: G_{VAP} : Independence model for variables (excluding context-specific independences). Abbreviations: SC: sputum colour; S: sputum; L: leukocytosis; AP: antipyretic drugs; T: temperature; P: pO_2/FiO_2; X: chest X-ray; VAP: ventilator-associated pneumonia.

5.5.5 Quality measures of models fitting the data

An important consideration in the construction of statistical models, Bayesian-network models included, is how well the constructed models fit the data. One the problems here is that we have three databases; so that the quality of the three resulting models cannot be compared. However, the different models per database can be compared mutually. For example, by comparing the models constructed on the basis of database D_{vap} (1) when using background knowledge for structure constraint and (2) when background knowledge was neglected, i.e., no structure constraints were imposed. Hugin offers measuring the quality of the probability density of a Bayesian network by calculating the Log-likelihood (LL) score, i.e., the likelihood of the data given the model is the product of the probabilities of the individual cases (assuming independence between the cases). The LL is then the sum of logarithms of these probabilities. In a single iteration of the EM algorithm, a propagation is performed for each case. This propagation computes (among other things) the probability of the case being propagated. The logarithm of this probability is added to the log-likelihood score being accumulated for the iteration. The closer the LL score is to zero, the better the model fits to the data. Table 5.3 includes the values of the LL scores for the three temporal Bayesian networks learnt

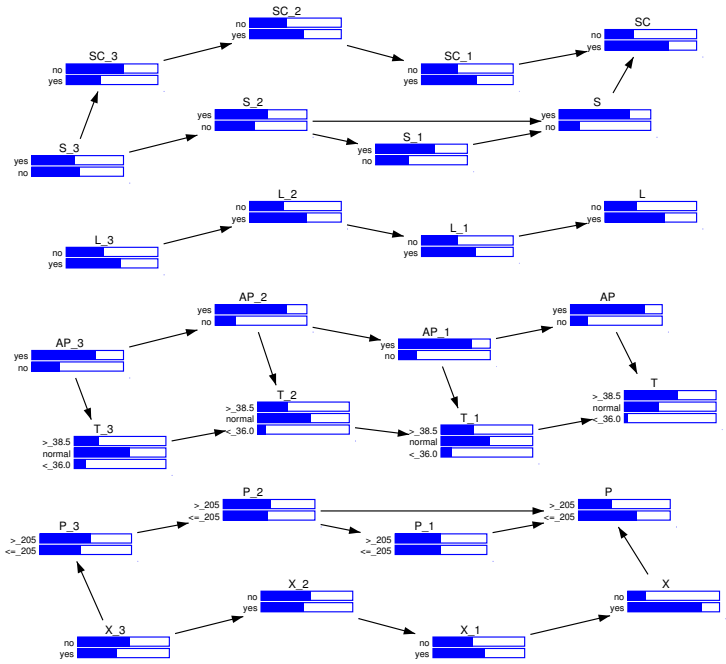


Figure 5.4: G_{vap} : Conditional probabilities for VAP patients. Abbreviations: SC: sputum colour; S: sputum; L: leukocytosis; AP: antipyretic drugs; T: temperature; P: pO_2/FiO_2; X: chest X-ray.

from the three databases. The models constructed by the NPC algorithm that included background knowledge are more tuned to real-life and have better quality outcomes compared to models constructed without using background knowledge. i.e., the PC algorithm, which is confirmed by the LL score for G_{vap} and G_{VAP}. The $G_{\overline{\text{vap}}}$ models only slightly differed when comparing LL for both NPC and PC constructed models.

5.6 Conclusions and discussion

The hypothesis underlying the research described in this paper was that in order to obtain insight in the evolution of disease processes, it is not necessary to explicitly model time, but also to consider context-specific independence information. The results we have obtained confirm this hypothesis, and to the best of our knowledge, this is the first paper combining context-specific independence and dynamic Bayesian networks.

When exploring variable dynamics, distinction was made between patients diagnosed and not diagnosed with VAP. For patients not diagnosed with VAP

Table 5.2: Dynamics of signs and symptoms in critically ill patients, showing 'abnormalities'. t_0 represents the day at which VAP was diagnosed for VAP patients; otherwise it denotes the day of matching for patients not diagnosed with VAP.

		Time point			
		t_{-3}	t_{-2}	t_{-1}	t_0
leukocytosis	vap	51%	60%	58%	65%
	$\overline{\text{vap}}$	46%	46%	47%	47%
body temperature	vap	34%	36%	41%	61%
	$\overline{\text{vap}}$	29%	24%	25%	28%
pO_2/FiO_2	vap	40%	47%	49%	61%
	$\overline{\text{vap}}$	24%	25%	22%	19%
X-chest	vap	32%	34%	46%	68%
	$\overline{\text{vap}}$	23%	22%	21%	17%
sputum amount	vap	47%	57%	64%	77%
	$\overline{\text{vap}}$	51%	54%	45%	27%
sputum colour	vap	38%	50%	57%	74%
	$\overline{\text{vap}}$	40%	41%	39%	28%

Table 5.3: Quality of the different probability tables of the investigated Bayesian network models, measured by the Log-likelihood score. Comparison was made between models constructed by the NPC and PC algorithm.

	LL G_{NPC} (LL G_{PC})
G_{vap}	-2121 (-2137)
$G_{\overline{\text{vap}}}$	-5978 (-5972)
G_{VAP}	-8722 (-8767)

the percentage presenting abnormal clinical parameters remains relatively stable (leukocytosis, pO_2/FiO_2, X-chest). However, the sputum variables show a decrease of abnormal values on t_0, whereas diagnostic variables for patients diagnosed with VAP show temporal behaviour.

The NPC learning algorithm proved to be useful, as it allowed for the incorporation of background knowledge, which appeared necessary to obtain clinically meaningful results. This algorithm combines the virtues of offering the capability of automatic learning of independence information from data, whereas uncertainty regarding both the presence of dependences and the directionality of arcs can be resolved by the user. Thus, the algorithm offers a natural role for the incorporation of expert background knowledge in the learning process.

The NPC algorithm is prefered above the PC algorithm, as the structure learning procedure of the latter lacks the opportunity to incorporate back-

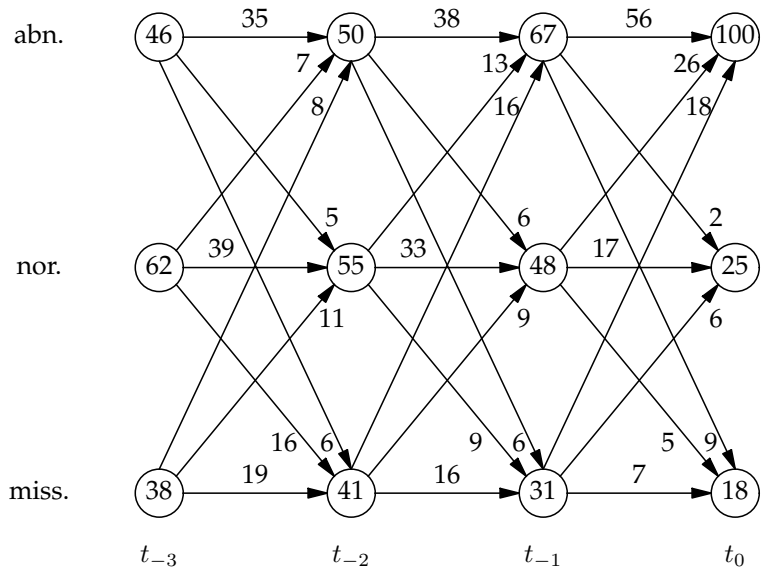


Figure 5.5: X-chest dynamics for patients diagnosed with VAP.

ground knowledge and, as a result, the algorithm allows for the inclusion of arcs directed from future to past (data not shown). In addition, LL scores showed that models constructed by means of the NPC algorithms have equal or better fit to the data, as compared to models constructed by the PC algorithm.

The results obtained for the ICU domain show that signs and symptoms of patients known to develop VAP have strong temporal relationships, whereas the temporal relationships between the signs and symptoms of patients not diagnosed with VAP were very weak. In addition, G_{vap} showed clear dynamics as denoted by the dynamics in marginal as well as conditional probabilities, as compared to $G_{\overline{\mathrm{vap}}}$. The combined model G_{VAP} included independences from both the VAP and non-VAP models. However, the model was not merely a union of the two independence sets.

We conclude from the constructed models that in inclusion of the association *X-ray* $\rightarrow$ *pO$_2$/FiO$_2$ ratio* should be considered when monitoring the lung function of ICU patients. It must be noted that this association was not incorporated in the structure of the static Bayesian network model described in Chapter 3, as its importance was not explicitly pointed out by clinical experts when constructing this model. In addition, neither was inclusion of this association accompanied by improved diagnostic performance of this model, according to the FAN learning algorithm described.

When comparing the atemporal parts of the networks, change in the independence information as a function of time was observed. This strengthens our belief that when constructing a temporal Bayesian model, it is not sufficient to resort to repetitive TBNs if one wishes to obtain models that capture the

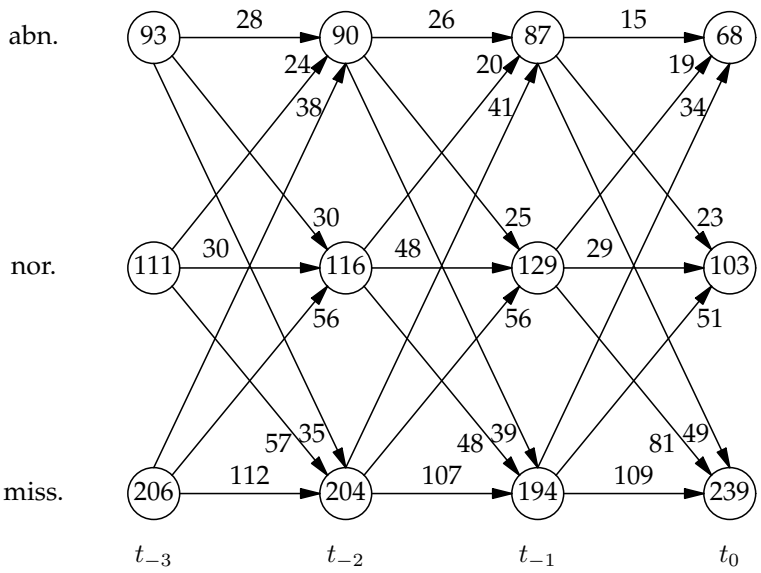


Figure 5.6: X-chest dynamics for patients for whom VAP was excluded.

characteristics of the domain; rather, non-repetitive TBNs should be investigated as well [70]. Further research is needed to evaluate our findings regarding the temporal behaviour of both models for VAP patients and non-VAP patients. It would, for example, be interesting to investigate the predictive value of an increasing body temperature of an ICU patient in relationship to the development of VAP. Moreover, a more detailed comparison of the ADGs of VAP and non-VAP for larger datasets may provide more insight in the course of the disease process of VAP. In conclusion, the combination of a general theory of TBNs, where repetitive and non-repetitive TBNs are both special cases, with the exploitation of context-specific independence information yielded a powerful data-analysis tool.

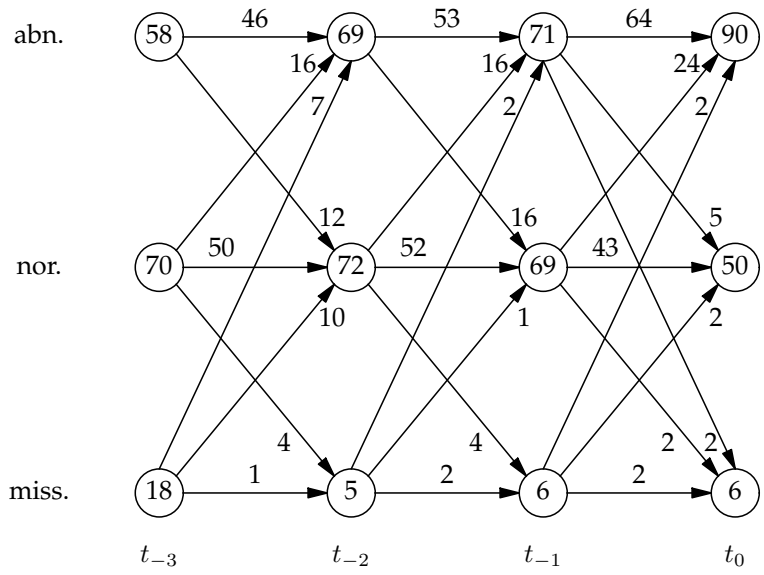


Figure 5.7: pO_2/FiO_2 dynamics for patients diagnosed with VAP.

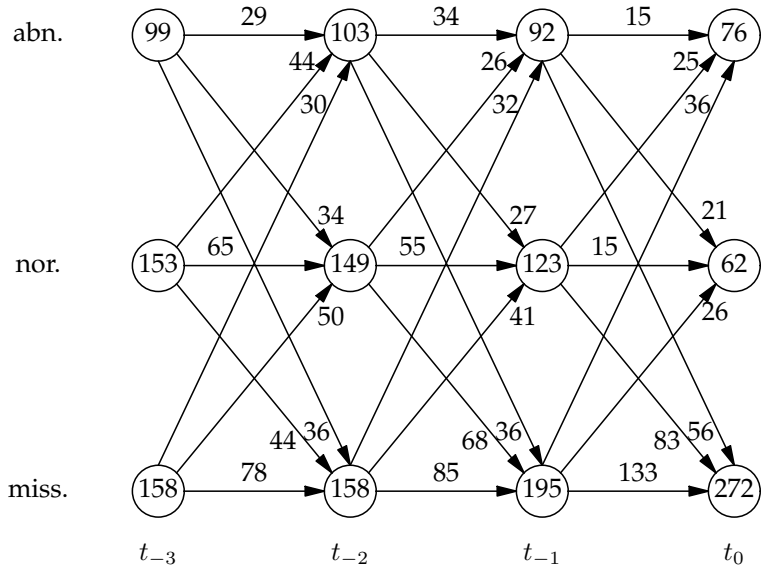


Figure 5.8: pO_2/FiO_2 dynamics for patients for whom VAP was excluded.

Part III

Antibiotic Treatment Selection for VAP

[6]

Predicting Causative Pathogens

6.1 Introduction

Ventilator-associated pneumonia (VAP) occurs in a considerable number of critically ill patients receiving mechanical ventilation. Reported incidence rates vary widely, depending on patient selection and diagnostic methods used [43]. VAP is thought to be associated with increased morbidity and higher healthcare related costs [43, 83]. In addition, delayed administration of appropriate antimicrobial treatment is associated with higher mortality and longer duration of mechanical ventilation [84]. Therefore, it is important to accurately identify infected patients and to prescribe appropriate antimicrobial treatment as soon as possible.

Diagnosing VAP remains a challenge as no gold standard exists. Usually a combination of different criteria is used, such as systemic signs of infection, abnormalities on chest roentgenogram and culture results of endotracheal secretions, i.e., sputum. However, each of these criteria has a low specificity for VAP. Only a proportion of patients colonised with potentially pathogenic microorganisms will develop VAP, usually after a gradual change from asymptomatic colonisation to clinical infection [85]. Invasive diagnostic techniques such as broncho-alveolar lavage (BAL) and protected specimen brush (PSB) have higher specificity for diagnosing VAP [43]. Nonetheless, these techniques are not commonly used in most ICUs. A consequence of the diagnostic uncertainty is the large amount of antibiotics prescribed for presumed VAP, which may be unnecessary and contributing to the emergence of resistant pathogens. Furthermore, current methods to determine bacteria and their susceptibility patterns bare a considerable diagnostic delay. For all these reasons, real-time decision-support systems may provide diagnostic benefits. Up till now only

This chapter is based on work published in the Journal of Antimicrobial Chemotherapy, 2008.

one computerised Bayesian network (BN) model for diagnosing VAP was constructed and evaluated [50, 34].

VAP is preceded by colonisation of the upper respiratory tract in almost all patients [60, 85]. Bacterial colonisation depends, amongst others, on the duration of mechanical ventilation and hospitalisation and on previous antibiotic use. And in daily clinical practice physicians base their judgment on the most likely cause of VAP on these variables and on the results of microbiological cultures. These two time-related variables (durations of hospitalisation and mechanical ventilation) and information of previous culture results and previous antibiotic use were, therefore, modelled in the previously described BN model as risk factors for ICU patients to acquire colonisation. The diagnostic performance of the BN model to predict the causative pathogen(s) of VAP was determined in a prospectively collected cohort of ICU patients.

6.2 Methods

6.2.1 Patients

Between January 1st 2000 and January 1st 2003 clinical, laboratory and microbiological data of mechanically-ventilated ICU patients were registered on a daily basis to study the development of VAP. All patients admitted to two ICUs (a ten-bed medical ICU and eight-bed neurosurgical ICU) of the University Medical Center Utrecht (UMCU) receiving mechanical ventilation > 48 hours were included in this cohort study. Patients with chronic mechanical ventilation at home as well as those with cystic fibrosis were excluded from analysis. As no intervention was evaluated and daily care was not influenced by the study, the institutional review board waived the necessity of informed consent.

6.2.2 Definitions

VAP was defined upon a diagnostic decision tree (Figure 3.4), described in detail previously [34]. For this, all episodes of clinical suspicion of VAP (n= 238; defined as days on which clinicians had prescribed antibiotics for presumed respiratory infections or infections without evident focus) that occurred in a consecutively studied cohort of 872 mechanically ventilated patients were retrospectively evaluated by two infectious-disease physicians (KS and MB) on the basis of clinical findings, the results of endotracheal aspirates and, when available, broncho-alveolar lavage (BAL) fluid. After adjudication, 157 episodes were considered as VAP, and the bacteria isolated from respiratory tract samples were considered the etiological cause of VAP (Table 1). These episodes and pathogens were considered the reference standard in the current study.

Pathogens were grouped as follows: *Pseudomonas aeruginosa* (PA), *Acinetobacter* spp. (AC), Enterobacteriaceae (Ent), *Staphylococcus aureus* (SA),

Haemophilus influenzae (HI) and *Streptococcus pneumoniae* (SP). The Enterobacteriaceae comprised multiple species; hence they were subdivided into two groups depending on the capacity of β-lactamase production: Ent1: *Escherichia coli*, *Klebsiella pneumoniae*, *Klebsiella* spp. and *Proteus Mirabilis*; Ent2: *Citrobacter* spp., *Enterobacter cloacae*, *Enterobacter* spp., *Moraxella* spp., *Morganella* spp. and *Serratia* spp..

Previous colonisation

Previous colonisation was defined as one or more positive cultures of endotracheal aspirate in the days prior to the day VAP was diagnosed. The pathogens were divided into: early-onset pathogens (SP, HI and SA), and late-onset pathogens (Ent1, Ent2, PA and AC). Hence, for early-onset
pathogens previous colonisation was defined as a positive culture within three days preceding the day on which VAP was diagnosed. For late-onset pathogens culture results during seven days prior to VAP were used. Obviously, if no cultures had been performed, previous colonisation was considered unknown and, consequently, no value for the node 'previous colonisation' in the BN model was filled in. Furthermore, if more than one culture was performed, only the results of the most recently-performed culture were selected.

Previous antibiotic use

Previously-administered antibiotics were considered effective when both of the following conditions were fulfilled: (1) the pathogen causing VAP was, based on in vitro susceptibility testing, susceptible to the antibiotics given; (2) the antibiotics were administered during at least two out of four days preceding VAP. In all other cases (including when no antibiotics were given) ineffective treatment was assumed [86]. See chapter 2 for more details.

6.2.3 Bayesian network model for diagnosing VAP

A Bayesian (or probabilistic) network model is a factorisation of a multivariate probability distribution, taking into account knowledge about conditional independences. It facilitates both the assessment of the probability distribution, as the amount of probabilistic information required may be greatly reduced, and computation of relevant probabilities that can be efficiently computed. The vertices in a Bayesian network represent variables, and arrows between the nodes indicate a causal relationship between the variables within a specific domain. More formally, the joint probability distribution of being colonised, as factorised by the structure of Figure 1, is denoted by
$P(\text{Col}_X,\text{MV},\text{Hosp},\text{Prev\_Col}_X,\text{Prev\_AB}_X)$
and can also be written using the structure of the network as
$P(\text{Col}_X \mid \text{MV},\text{Hosp},\text{Prev\_Col}_X,\text{Prev\_AB}_X) \cdot P(\text{VAP} \mid \text{Col}_X,\text{MV})$
$\cdot P(\text{Prev\_Col}_X) \cdot P(\text{Prev\_AB}_X) \cdot P(\text{Hosp})$
, where X denotes one of the seven modelled pathogens.

In many cases it is possible to give a causal interpretation of a Bayesian network in terms of common causes of an effect, which makes it easier to attach a clinical interpretation to the computed posterior probabilities given evidence about a patient's disease. In these cases the direction of the arc distinguishes the cause-node from the effect-node. The qualitative part of the Bayesian network consists of tables containing conditional probabilities for each variable in the network. The conditional probabilities of the original model are based on subjective estimates, i.e., expert opinions, and data from literature and were at a later stage updated by machine learning techniques [34].

In our BN model, 'duration of hospitalisation' was modelled as a node containing five values: $<$5 days in hospital, $\geq$5 days on the ward and/or in the ICU, $\geq$5 days in the ICU, $<$5 days in hospital for a patient with chronic obstructive pulmonary disease (COPD) who was recently hospitalised (within three weeks), and $\geq$5 days in hospital for a patient with COPD and recent hospitalisation. The node .duration of mechanical ventilation. in the Bayesian network incorporates three – arbitrarily chosen – categories: 48–96h, 96–144h and $\geq$144h.

Each of the seven groups of pathogens was modelled as a single node in the BN, as the presence of a certain pathogen does not imply the absence of other pathogens. As acquisition of pathogens depends on the duration of hospital stay and duration of mechanical ventilation these two time-related variables were modelled as parents of the pathogen nodes. In addition, for each pathogen group a parent-node representing whether effective or non-effective treatment was previously administered and a parent-node indicating whether previous colonisation had been demonstrated was added.

The appropriateness of antibiotic therapy was pragmatically analysed assuming standard antibiotic prescription for each pathogen group predicted by the model, using the following choices: amoxicillin for *S. pneumoniae*, amoxicillin-clavulanic acid for *H. influenzae* and Ent1, flucloxacillin for *S. aureus*, ciprofloxacin for Ent2 and ceftazidime for *P. aeruginosa* and *Acinetobacter* species. Appropriateness was determined upon in vitro susceptibilities of the reference pathogens.

6.2.4 Analyses

The BN model predicts the likelihood (from 0% to 100%) for a certain pathogen to be a causative agent of VAP. This likelihood was dichotomised to denote either the presence or absence of a pathogen as a cause of VAP. The cut-off was defined as the point on the ROC curve that resulted in the optimal trade-off between sensitivity and specificity. BN model likelihood predictions above this threshold were interpreted as positive for that pathogen, i.e. the model predicts that that specific pathogen should be considered as causative for VAP. Naturally, these thresholds may differ for each model.

The diagnostic accuracy of the BN model to correctly predict pathogens causing VAP was assessed by successively adding information. In the first analysis, only information on duration of mechanical ventilation and duration

of hospital stay was used (Analysis 1). Subsequently, information on endotracheal culture results (Analysis 2) and previously-administered antimicrobial treatment (Analysis 3) were added. Finally, both information (previous antibiotics and culture results) was added simultaneously (Analysis 4). The performance of the BN model, using the different amounts of information, was analysed with Receiver Operating Curve (ROC) characteristics. Performance was classified according to the traditional academic point system [55], in which an Area Under the Curve (AUC) between 0.90 and 1.00 is considered excellent, between 0.80–0.90 good, 0.70–0.80 fair, 0.60–0.70 poor and 0.50–0.60 fail. Diagnostic test accuracy was further assessed by calculating the sensitivity, specificity, and positive and negative predictive value (PPV and NPV) for all episodes of VAP. The output of the best performing model was then used to analyse how well the model predicted polymicrobial VAP episodes.

IDEAL was used to calculate posterior probabilities, SPSS version 14 was used to calculate descriptive statistics and for ROC analyses. Paired samples T-tests or Non-parametric Tests were used, when appropriate, to assess statistical differences. Probability levels less than 0.05 were considered statistically significant.

6.2.5 Quality of Bayesian-network model's fit to the data

Several scores exist that measure the quality of the probability density of a Bayesian network, such as the Bayesian Information Criterion (BIC) [87], Akaike Information Criterion (AIC)[88] and log-likelihood score. BIC and AIC scores take complexity of different models into account. As the models to be compared have the same complexity, i.e., have the same structure, the log-likelihood score was used for comparison. The log-likelihood expresses how well the model, in terms of underlying structure and parameters, fits the data. The closer the score is to zero, the better the model fits to the data. The sum of the log-likelihood scores of the model was assessed for each pathogen.

6.3 Results

6.3.1 Data

The cohort contained 157 episodes of VAP in 140 patients; 105 episodes were monobacterial and 52 episodes were polymicrobial. However, in 4 episodes of monobacterial and 6 episodes of polymicrobial VAP
Stenotrophomonas maltophilia was considered causative. As these were not incorporated as a pathogen or pathogen group in the BN model, these pathogens were excluded from analysis. In two episodes both pathogens belonged to the same group (Ent1). Thus, in total 199 pathogens were considered causative (153 episodes in 140 patients; 107 monobacterial and 46 polymicrobial (in all cases caused by two pathogens)) and were included in our analyses. The

Table 6.1: Reference standard: frequency of VAP-causing pathogens. Pathogens part of the Enterobacteriaceae groups are shown in gray.

	Episodes of VAP		
Causative pathogens	Monobacterial N=107 episodes (70%)	Polymicrobial N=46 episodes (30%)	Total N=153 episodes
Pseudomonas aeruginosa	19	11	30
Acinetobacter spp.	6	8	14
Enterobacteriaceae 1	29	17	46
Escherichia coli	17	7	23
Klebsiella pneumoniae	11	4	15
Klebsiella spp.	1	2	3
Proteus mirabilis	0	4	4
Enterobacteriaceae 2	17	15	32
Serratia spp.	5	4	9
Morganella spp.	0	2	2
Citrobacter spp.	3	2	5
Enterobacter spp.	2	0	2
Enterobacter cloacae	7	7	14
Staphylococcus aureus	25	16	41
Haemophilus influenzae	8	14	22
Streptococcus pneumoniae	3	11	14
Total number of pathogens	107	92	199

largest group of pathogens (23%) was Ent1 and the smallest were AC and SP (both 7%) (See Table 6.1).

The proportion of patients with previous colonisation with the same pathogen ranged from 46% for *S. aureus* to 70% for *P. aeruginosa* (Table 2). Proportions of patients that had received antibiotics effective for the causative pathogen of VAP ranged from 9% for *H. influenzae* to 47% for Ent2 (See Table 6.2). Details on antimicrobial agents effective for the causative pathogen are specified in Table 6.3.

6.3.2 Pathogen-specific performance of the BN model

In analysis 1, pathogens were predicted using information on duration of hospitalisation and mechanical ventilation only (See Table 6.4). For *P. aeruginosa*, the threshold for positivity in this model was 27.8% and with this cut-off the AUC for predicting *P. aeruginosa* as a cause of VAP was 0.718 (0.626–0.809 95% Confidence Interval (CI)). The highest AUC was obtained for *S. pneumoniae* (0.772 (0.64–0.905 95% CI) with a cut-off of 4.3% and the lowest for both Enterobacteriaceae groups, both with an AUC of 0.511.

In analysis 2, information of previous culture results was added and the performance improved for all pathogens (See Table 6.4).

Table 6.2: Previous colonisation and previous antibiotic use for all 153 episodes of VAP.
Previous colonisation status was considered unknown, when culture data were not available.
PA: *P. aeruginosa*; AC: *Acinetobacter* spp.; Ent1: Enterobacteriaceae 1; Ent2: Enterobacteriaceae 2; SA: *S. aureus*; HI: *H. influenzae*; SP: *S. pneumoniae*.

		Previous colonisation			Previous antibiotic use	
Pathogen	N	yes	no	unknown	effective	ineffective or none
PA	30	70%	27%	3%	13%	87%
AC	14	64%	36%	0%	36%	64%
Ent1	46	67%	20%	13%	44%	56%
Ent2	32	59%	25%	16%	47%	53%
SA	41	46%	27%	27%	15%	85%
HI	22	55%	18%	27%	9%	91%
SP	14	50%	21%	29%	21%	79%

Table 6.3: Previous antibiotic use in more detail. For all VAP episodes, only effective antibiotics against the VAP-causing pathogen administered during 4 days (when data were available) preceding the day VAP was diagnosed, are reported.
PA: *P. aeruginosa*; AC: *Acinetobacter* spp.; Ent1: Enterobacteriaceae 1; Ent2: Enterobacteriaceae 2; SA: *S. aureus*; HI: *H. influenzae*; SP: *S. pneumoniae*.

	Causative pathogen						
Antimicrobial agent	PA N=4	AC N=5	Ent1 N=20	Ent2 N=15	SA N=6	HI N=2	SP N=3
Cephalotin			1	3	1	1	
Ceftazidime	2		1				
Ceftriaxone			2	2			
Ciprofloxacin	1	2	1	1	1		1
Amoxicillin/ Clavulanic acid			12	7	3	1	
Cotrimoxazole		2	3	1	1		1
Meropenem	1	1		1			
Vancomycin							1

AUCs were now 0.916 (0.846–0.987 95% CI) for *P. aeruginosa* (cut-off now 13.7%) and 0.916 (0.85–0.982 95% CI) for *S. pneumoniae* (cut-off now 3.4%). The lowest AUC (0.831 (0.681–0.981 95% CI)) was obtained for *Acinetobacter* species. The confidence areas of the AUCs of the second analysis did not overlap with those of the first analysis for *P. aeruginosa*, Ent1, Ent2 and *S. aureus*, indicating that model predictions improved statistically significant for these pathogen groups.

Adding information on previous antibiotic use to the model of the first analysis (as in Table 6.4) hardly changed model performances (analysis 3: data not shown). Similarly, adding information on previous antibiotic exposure to the model already using information on previous cultures (analysis 2, Table 6.4) increased the model performance for all pathogens slightly, but not significantly (analysis 4: See Table 6.4). As shown in Tables 6.4, 6.4 and 6.4, the sum of log-likelihood scores increased with adding information (from analysis 1 to 3; $p<0.05$), indicating improved fit of the BN model to the data.

6.3.3 Prediction of monobacterial and polymicrobial episodes of VAP

In this analysis we evaluated to what extent the best performing model (as in Table 6.4) identified one (and the correct) pathogen causing monobacterial VAP and the two pathogens causing polymicrobial VAP. According to model predictions VAP was monobacterial in 67 cases (107 episodes according to reference); this was correct for 60 episodes (90%) and in 52 episodes (78%) the correct pathogen was predicted (See Table 6.5). In 86 episodes VAP was, according to the model, caused by multiple pathogens (46 according to reference); 43 times by two pathogens, 23 times by three pathogens, 3 times by four pathogens, 12 times by five pathogens and five times by six pathogens (See Table 6.5). In all, 91 of the 107 pathogens (85%) causing monobacterial VAP were correctly predicted. Sensitivities to predict the correct
pathogen in case of monobacterial VAP were highest for *H. influenzae* (8/8, 100%) and lowest for *S. aureus* (18/25, 72%) and *S. pneumoniae* (2/3, 67%).

In 46 episodes the model predicted a polymicrobial etiology whereas the reference standard contained only one pathogen. The two pathogens causing VAP (according to the reference) were correctly identified as the only two pathogens in 17 of 46 (37%) episodes of polymicrobial VAP. However, in all, the two pathogens were correctly identified (although frequently together with other 'incorrect' pathogens) in 29 of 46 (63%) episodes. When combining accuracies of predicting monobacterial as well as polymicrobial episodes of VAP, the BN model correctly identified pathogens of 78% (91 + 29 / 153) of all VAP episodes. When a fixed antibiotic choice was linked to the pathogen(s) predicted as causative for VAP by the BN model, 92% (140 of 153) of all episodes of VAP would have received appropriate therapy. Causative pathogens in 13 VAP episodes that were not covered by the model, were *Acinetobacter* spp (n=4), *Enterobacter cloacae* (n=3), *P. aeruginosa* (n=2), *Klebsiella* spp. (n=2), *S. aureus* (n=2), *Serratia* spp. (n=1), and *H. influenzae* (n=1).

Tables 6.4, 6.4 and 6.4: Predictive performance of the BN model using information on duration of hospitalisation and mechanical ventilation.
PA: *P. aeruginosa*; AC: *Acinetobacter* spp.; Ent1: Enterobacteriaceae 1; Ent2: Enterobacteriaceae 2; SA: *S. aureus*; HI: *H. influenzae*; SP: *S. pneumoniae*. AUC: area under the ROC curve; CI: confidence interval; T: threshold; Sens: sensitivity; Spec: specificity; PPV: positive predictive value; NPV: negative predictive value; Σ LL: sum of log-likelihood scores.

Table 6.4: Analysis 1. Predictive performance per pathogen: without information of previous colonisation and without information on previous antibiotic use.

	AUC	95% CI	T	Sens	Spec	PPV	NPV	Σ LL
PA	0.718	0.626–0.809	27.8	0.900	0.472	0.293	0.951	-71.6
AC	0.603	0.469–0.737	12.3	0.857	0.424	0.130	0.967	-46.7
Ent1	0.511	0.416–0.606	33.3	0.848	0.252	0.328	0.794	-92.7
Ent2	0.511	0.391–0.631	18.7	0.906	0.058	0.203	0.700	-78.4
SA	0.633	0.532–0.733	17.1	0.976	0.080	0.280	0.900	-85.9
HI	0.768	0.666–0.870	6.6	0.818	0.627	0.295	0.957	-60.3
SP	0.772	0.640–0.905	4.3	0.786	0.662	0.190	0.968	-45.4

Table 6.4: Analysis 2. Predictive performance per pathogen: with information of previous colonisation, but without information on previous antibiotic use.

	AUC	95% CI	T	Sens	Spec	PPV	NPV	Σ LL
PA	0.916	0.846–0.987	13.7	0.833	0.837	0.591	0.963	-34.5
AC	0.831	0.681–0.981	1.4	0.786	0.655	0.385	0.969	-30.8
Ent1	0.877	0.816–0.938	29.7	0.783	0.776	0.607	0.902	-63.7
Ent2	0.887	0.830–0.944	17.9	0.781	0.777	0.524	0.910	-52.4
SA	0.861	0.794–0.927	37.2	0.805	0.812	0.447	0.909	-67.7
HI	0.893	0.823–0.963	8.9	0.818	0.824	0.378	0.954	-43.7
SP	0.916	0.850–0.982	3.4	0.857	0.799	0.476	0.970	-30.5

Table 6.4: Analysis 4. Predictive performance per pathogen: with information of previous colonisation and with information on previous antibiotic use.

	AUC	95% CI	T	Sens	Spec	PPV	NPV	Σ LL
PA	0.921	0.854–0.989	28.9	0.833	0.959	0.641	0.956	-33.9
AC	0.859	0.739–0.980	1.3	0.786	0.698	0.526	0.970	-29.9
Ent1	0.879	0.818–0.939	27.1	0.783	0.766	0.569	0.863	-63.5
Ent2	0.890	0.835–0.946	21.5	0.781	0.793	0.328	0.876	-54.5
SA	0.873	0.809–0.937	27.2	0.805	0.812	0.493	0.907	-64.9
HI	0.899	0.833–0.965	12.0	0.818	0.832	0.375	0.962	-39.3
SP	0.929	0.875–0.983	5.8	0.857	0.827	0.224	0.971	-28.4

Table 6.5: Performance of pathogen prediction. Numbers in boldface denote correctly predicted episodes of VAP, otherwise the numbers denote partially correct or incorrect pathogen detection.
NA: not applicable; None: no pathogen; Mono: monobacterial; poly: polymicrobial.

		Total number of predicted pathogens	Number of correctly/ incorrectly predicted pathogens		Reference standard 153 VAP episodes	
			correct	incorrect	Mono N=107	Poly N=46
Model Predictions	None	0	NA	NA	0	0
	Mono	1	1	0	**52**	5
			0	1	8	2
	Poly	2	0	2	6	2
			1	1	**15**	3
			2	0	NA	**17**
		3	0	3	1	3
			1	2	**12**	1
			2	1	NA	**6**
		4	1	3	**2**	0
			2	2	NA	**1**
		5	0	5	1	0
			1	4	**6**	1
			2	3	NA	**4**
		6	1	5	**4**	0
			2	4	NA	**1**
Total					**91/107 (85%)**	**29/46 (63%)**

6.4 Discussion

We have described a Bayesian network model that is able to accurately predict the most likely cause(s) of VAP. Combining information of the time of intubation and of previous culture results from respiratory tract samples appeared essential, supporting the usefulness of regular surveillance as a means to assist physicians in choosing appropriate antibiotics. If confirmed in prospective studies in other settings, this Bayesian decision model might offer a reliable and valuable tool in the management of critically ill patients.

We have previously published on the performance of this model to predict presence or absence of VAP [34]. The ROC curve for the BN model to correctly predict presence or absence of VAP had an AUC of 0.857 (0.827–0.888 95% CI) with sensitivity and specificity of 80% and positive and negative predictive value of 6.1% and 99.6%, respectively.

Despite the positive results of the previous and the present study, some aspects preclude widespread use of this model in daily clinical practice, at this stage. Importantly, the model has been tested only retrospectively in a single cohort. Moreover, some of the model parameters need to be discussed. The prior probabilities (e.g., the likelihood of colonisation with *P. aeruginosa* on successive days after intubation) used by the model are to a large extent based upon best guess estimates of the investigators, and where possible, upon published evidence. Naturally, many of the estimates on etiology and antibiotic susceptibility may be site-specific and might even change in time. For instance, outbreaks of pathogens or changes in antibiotic or infection control policies may well influence a priori probabilities. Therefore, a diagnostic model as presented will always need optimisation when used in a new setting.

There are also several possibilities to improve the model. The distribution of Enterobacteriaceae in two different pathogen groups can be justified upon the differences in antibiotic susceptibility, as witnessed in our setting. However, grouping was also needed to avoid the presence of bacteria with extreme low prevalence rates. Using the model in larger patient populations, though, will allow for analyses of more individual bacterial species. In addition, newly emerging pathogens, such as S. maltophilia in our setting, could easily be included. In addition, the likelihood of colonisation was a time-dependent variable and was also influenced by previous antibiotic exposure. The likelihood of persistence of colonisation, despite antibiotic therapy for which that pathogen is susceptible in vitro, differs considerably between pathogens, and this information was used [86]. However, a certain antibiotic may increase the likelihood of colonisation (and subsequent infection) with pathogens that are not susceptible to that antibiotic, and such associations were not included.

Another relevant aspect in the microbial ecology within ICUs is colonisation pressure [89], which is defined as the proportion of all other patients colonised with a certain pathogen, and which is positively associated with the risk of cross-transmission. Adding a parameter representing colonisation pressure of each pathogen might further enhance pathogen prediction.

In this study we have determined thresholds for pathogens (causing or not causing VAP) upon the points of the ROC curves with optimal sensitivity and specificity. However, from a clinicians. point of view it might be preferable to use higher sensitivities for certain pathogens (e.g., *P. aeruginosa*), in order to minimise the risk of not treating those episodes of VAP appropriately. Naturally, this would improve the negative predictive value, at the cost of a lower specificity and the possibility of unnecessary treatment of patients.

Finally, model accuracy might improve if more categories of duration of hospitalisation and mechanical ventilation are used. We have now used an arbitrary cut-off between early- and late-onset pathogens. Early-onset VAP, occurring during the first days of mechanical ventilation, is most frequently caused by pathogens such as *S. aureus*, *H. influenzae* and *S. pneumoniae*. In contrast, late-onset VAP is usually caused by organisms such as *P. aeruginosa*, *Acinetobacter baumannii*, Enterobacteriaceae and multiresistant pathogens [43, 39, 90, 91]. The cut-off time points between early-onset and late-onset VAP episodes have not been clearly defined [92]. Four, six and seven days after initiation of mechanical ventilation have been suggested [39, 90, 91, 93].

In addition to these points for model improvement, our findings suggest that decision-support systems could enhance patient management. Appropriateness of antimicrobial therapy in 92% of the episodes would be much higher than reported rates from international studies that have been as low as 32% [94], 49.5% [95] and 46% [96].

Of note, antibiotic selection in our analysis was based upon a fixed choice linked to each pathogen and could be improved by linking antibiotic choices to information on recent trends of antibiotic susceptibilities of bacteria isolated in the unit. However, because of extremely low levels of antibiotic resistance in our setting, it is unlikely that appropriateness of antibiotics could be further improved, and this will be much more relevant in settings with high resistance levels.

The ultimate proof of the clinical usefulness of our model should be determined in a randomised study. To the best of our knowledge, no other decision-support model has been developed for VAP. Recently, though, a causal probabilistic network for predicting sites of infection and specific pathogens in patients with common bacterial infections (TREAT) was evaluated in observational and experimental study designs including more than 3500 patients in several European countries. The performance of this model, also expressed in the AUC of ROC curves, ranged from 0.63 for infection caused by *S. aureus* to 0.80 for those caused by *E. coli*. The investigators concluded that their model could be used to triage patients by the risk for specific pathogens and enable antimicrobial therapy prior to pathogen detection [97]. In a subsequent cluster randomised trial comparing wards with and without using TREAT, the rate of appropriate empirical antibiotic treatment improved in the intervention wards (85%; odds ratio 3.40 (1.96–5.90 95% CI) per protocol, adjusted for location and clustering) while reducing antibiotic costs [98].

[7]

Application of the Close-World Assumption to a Bayesian Network for Optimal Treatment Selection

7.1 Introduction

Patients who are admitted to the Intensive Care Unit (ICU) are often severely ill and are prone to colonisation by hospital-acquired (nosocomial) bacteria. Some of these bacteria are more pathogenic[1] than others. Our medical domain is restricted to mechanically ventilated patients who are at risk of developing an infection of the lower respiratory tract. This infection is called ventilator-associated pneumonia, or VAP for short. To treat pathogens that have caused an infection, antimicrobial treatment is needed.

Prescribing optimal treatment is a difficult task for physicians. They have the tendency to prescribe broad spectrum antibiotics to cover as much pathogens as possible. However, this stimulates the development of resistance of pathogens to specific antibiotics, which will eventually reduce the effectiveness of these antibiotics in time [99]. Therefore, more and more research is being performed in the area of standardisation of antibiotic prescription; guidelines and protocols are well-known practical outcomes of such efforts. In our research, we focus on supporting physicians in prescribing antimicrobial treatment for ICU patients by means of a Bayesian network (BN) augmented by a decision-theoretic model [50]. Bayesian networks have been introduced in the 1980s as a formalism to compactly represent and reason efficiently with joint probability distributions. Bayesian networks are in particular well suited for the representation of uncertain causal relations within a specific domain of expertise.

This chapter is published in the Proceedings of the 3rd European Medical and Biological Engineering Conference, 2005.

[1] Pathogenicity is the ability of an organism of causing an individual to get ill.

This paper describes our attempt to improve the therapeutic performance of our Bayesian network. We used real patient data to test the accuracy of the network in selecting optimal antimicrobial treatment. It is shown that performance improvement can be achieved by using the concept of the Closed-World Assumption (CWA) from database theory and logic [100].

This paper describes our attempt to improve the therapeutic performance of our Bayesian network. We used real patient data to test the accuracy of the network in selecting optimal antimicrobial treatment. It is shown that performance improvement can be achieved by using the concept of the Closed-World Assumption (CWA) from database theory and logic [100].

The paper is organised as follows: Bayesian networks and the Bayesian network for the treatment of VAP we have developed previously are first described, followed by a brief description of the basic ideas underlying the CWA. The exploitation of the CWA in the context of probability theory is studied next. The practical usefulness of these ideas are subsequently investigated for the Bayesian network concerning VAP. The paper is rounded off by some conclusions.

7.2 Materials and methods

7.2.1 A Bayesian network for VAP

Formally, a *Bayesian network*, BN for short, is defined as a pair $\mathcal{B} = (G, \Pr)$, where $G = (\mathbf{V}(G), \mathbf{A}(G))$ is a directed acyclic graph with a set of vertices $\mathbf{V}(G) = \{V_1, \ldots, V_n\}$, corresponding one to one to stochastic variables, here denoted by the same indexed letters, and a set of arcs $\mathbf{A}(G) \subseteq \mathbf{V}(G) \times \mathbf{V}(G)$, and $\Pr$ is a joint probability distribution $\Pr(V_1, \ldots, V_n)$ representing statistical dependences and independences among the variables, respecting the independences represented in the graph, as follows:

$$\Pr(V_1, \ldots, V_n) = \prod_{i=1}^{n} \Pr(V_i \mid \pi(V_i)),$$

where $\pi(V_i)$ stands for the variables corresponding to the parents of vertex V_i. In the following, upper-case letters , such as X, stand for (free) variables, whereas lower-case letters, such as x (short for $X = \textit{yes}$) or $\neg x$ (short for $X = \textit{no}$), stand for (collections) of values of variables.

The formalism of BNs supports the kind of reasoning under uncertainty that is typical for medicine when dealing with diagnosis, treatment selection and planning, and prediction of prognosis. A Bayesian-network model concerning VAP was previously constructed with the help of two infectious-disease domain experts, who helped in establishing the network's structure and estimated all conditional probabilities required [50]. This
Bayesian network can be seen as consisting of a diagnostic part, modelling signs and symptoms of VAP, and a therapeutic part, that models the temporal

evolution of the disease based on duration of stay in the ICU and duration of mechanical ventilation. These parts are linked together through the 'PNEUMONIA' vertex (the 'PNEUMONIA' vertex has one incoming arc from the therapeutic part).

Entities that play an important role in the development of VAP and that belong to the *diagnostic part* of the Bayesian network for VAP include: the duration of *mechanical ventilation*, the amount of *sputum*, *radiological signs*, i.e., whether the chest radiograph shows signs of an infection, *body temperature* of the patient and the number of *leukocytes* (white blood cells) [27]. Each of these entities is modelled as one vertex in the diagnostic part of the Bayesian network for VAP, as shown in Figure 3.5.

The *therapeutic part* of the network models the situation of a patient from the colonisation and development of pneumonia as temporal processes, to the selection of optimal antimicrobial treatment. We have modelled seven groups of microorganisms, each as one vertex in the Bayesian network. These microorganisms are: *Pseudomonas aeruginosa*;
Haemophilus influenzae; *Streptococcus pneumoniae*; two groups of Enterobacteriaceae, depending on which antibiotics these are susceptible to; *Staphylococcus aureus* and *Acinetobacter* spp. Also, for each modelled microorganism the pathogenicity was included in the model; the pathogenicities were assumed to be equal for each microorganism.

The presence of certain bacteria is influenced by antimicrobial therapy. Each microorganism is susceptible[2] to some particular antibiotics, and these susceptibilities were taken into account while constructing the model. The infectious-disease experts assigned utilities[3] to each combination of microorganism(s) and antimicrobial drug(s) using a decision-theoretic model [101].

7.2.2 Modelling joint interactions

To model the probabilistic interaction of the various pathogens on the likelihood of development of pneumonia and overall susceptibility, the notion of *causal independence* was used [102, 103]. For example, the interaction among susceptibility or coverage variables, as shown in Figure 7.1, was expressed by a logical-AND gate. The probability distribution of the vertex that represents the overall susceptibility or coverage is expressed as the *conjunctive effect* of the seven different pathogens, which can be defined formally as follows:

$$\Pr(\text{coverage} \mid \text{ab}) = \prod_{i=1}^{n} \Pr(\text{susceptibility-pathogen}_i \mid \text{ab}) \tag{7.1}$$

[2]Susceptibility, here, is stated as the sensitivity to or degree to which a microorganism is affected by treatment with a specific antibiotic.

[3]Utility: by definition a quantitative measure of the strength of the preference for an outcome.

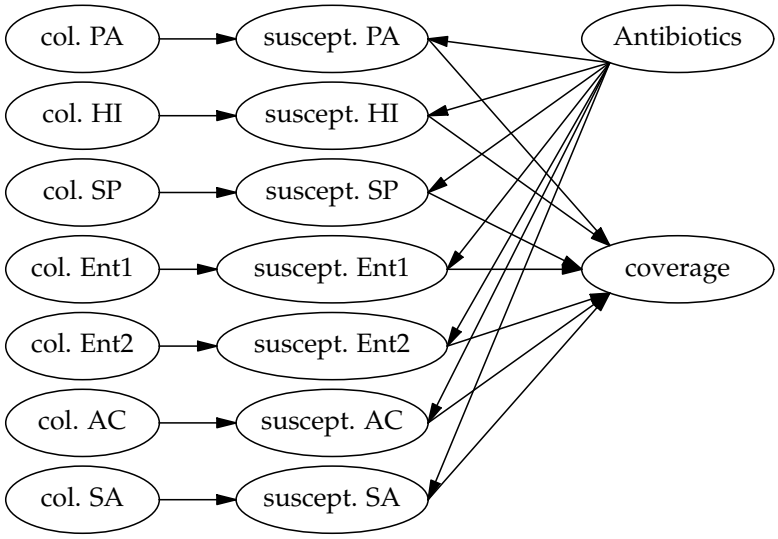


Figure 7.1: Most important fragment of the therapeutic part of the Bayesian network. PA: Pseudomonas aeruginosa; HI: Haemophilus influenzae; SP: Streptococcus pneumoniae; Ent{1,2}: Enterobacteriaceae{1,2}; AC: *Acinetobacter* spp.; SA: Staphylococcus aureus. Each pathogen is susceptible (suscept.) to particular antibiotics and an optimal coverage of the pathogens is what the model tries to achieve.

Using the logical-AND gate, the model tries to cover all pathogens, i.e., the probability $\Pr(\text{coverage} \mid \text{ab})$ for all possible antimicrobial treatment values 'ab' of the variable 'ANTIBIOTICS' is computed. This is balanced against the broadness of the antimicrobial spectrum. Thus, there is a trade-off between coverage and broadness of antimicrobial spectrum of the prescribed treatment.

7.2.3 Preliminary evaluation

Preliminary evaluation of the performance of the current network model learnt us that the model, generally, advised broad spectrum antibiotics, even when the patient was only colonised by one or two pathogens. When prescribing antibiotics, the spectrum should be chosen based on information concerning the susceptibilities of causative pathogens. In general, when increasing the number of different possible causative pathogens, there will be a need for broader coverage and, thus, more different antibiotics. If the patient is only infected by one or two pathogens, prescription of one narrow to intermediate spectrum antibiotic is often sufficient. Our conclusion was that there was room for improvement of the model. More detailed evaluation results that support this claim are discussed below.

7.2.4 The closed-world assumption

In this paper, we study the idea from Artificial Intelligence (AI) concerning the assumption that unknown, absent data can be taken as being negative. We investigate whether this idea can have a positive impact on the performance of a Bayesian network. We first describe the basic ideas from the perspective of AI, followed by how these ideas can be exploited in the context of probability theory.

A typical logical AI system stores its knowledge about its domain as a finite set Γ of first-order logic formulas. To answer queries, the system will have to decide whether or not a formula, say φ, can be obtained by performing logical deductions on Γ; formally: $\Gamma \models \varphi$, meaning that φ is a logical consequence of our knowledge represented in the *knowledge base* Γ. A typical use of this logical approach is in rule-based expert systems, where Γ is a collection of logical rules, and φ could be something such as the diagnosis of a disease based on entered *patient findings* or *evidence* E:

$$\Gamma \cup E \models \varphi$$

Thus, the knowledge base Γ is static, and augmented by non-static, patient-specific findings E to conclude about a disease φ, which is then interpreted as a diagnosis, or prognosis, etc., depending on the medical purpose of the knowledge in Γ.

This model has been proven to be quite useful for various tasks requiring knowledge about a domain, however, it has its limitations. Let us assume that what we know about our world is stored in Γ. When we want to know something about, for example, a certain φ, we search in Γ for information concerning this φ. It might be the case that it is not possible to conclude φ from Γ, indicating that somehow our knowledge base Γ is incomplete. Clearly, this is often the case, as it may be impossible to ensure that a knowledge base offers a complete coverage of all the knowledge in a domain of concern. The assumption that a knowledge base is incomplete is known as the *open world assumption* [100, 104].

However, when the knowledge base Γ cannot tell us anything about φ, we may also assume that the negation of φ, i.e., $\neg\varphi$, holds and can be added to Γ. Formally, we have that if $\Gamma \not\models \varphi$, we assume that $\neg\varphi$ is a member of a set of assumptions or beliefs B, i.e., B is the smallest set of negative assumptions or beliefs, such that for each $\psi \in B : \Gamma \not\models \psi$ and $\Gamma \not\models \neg\psi$, where $\psi \equiv \neg\varphi$. Now, if it is the case that

$$\Gamma \cup B \models \neg\varphi$$

then we say that $\neg\varphi$ is in the CWA-augmented knowledge base Γ, formally $\neg\varphi \in \mathrm{CWA}(\Gamma)$. This is called the *closed world assumption* [100, 104]. It is used to provide a default, negative solution in the absence of a positive solution.

There are at least two situations where the CWA is used. The first is where it is assumed that a knowledge base contains all relevant facts. This is common in corporate databases. That is, the information it contains is assumed to be

complete. The second situation is where it is known that the knowledge base is incomplete (does not have enough information to produce an answer to a question) and a decision must be made without complete information — a situation familiar to most people. The closed world assumption is designed to solve a reasoning problems in both of these situations.

In medicine, it is unlikely that all information concerning a patient is known, as only the results of a selected set of tests are known. Thus, it is common to assume that unless explicitly stated, everything not known about a patient is normal. For example, if the blood pressure of a patient is unknown, and there are no indications that absence of information concerning the blood pressure may be a medical mistake, it is normally assumed that the blood pressure is *not* high. Hence, the CWA in medicine is commonly used to interpret data concerning a patient; without it, it is hard to draw clear conclusions. Here we, therefore, assume that the CWA is only used to handle patient data. This means that in reasoning with logical rules in a knowledge base Γ and patient findings E, such that

$$\Gamma \cup E \models \varphi$$

we compute the CWA, not of the knowledge base Γ, but of the set of patient findings E in the context of Γ. Hence, logical reasoning in medicine can be formalised as reasoning according to the following definition

$$\Gamma \cup \mathrm{CWA}_{\Gamma}(E) \models \psi$$

where if $\neg\varphi \in \mathrm{CWA}_{\Gamma}(E)$: $\Gamma \cup E \not\models \varphi$, $\Gamma \cup E \not\models \neg\varphi$ and $\psi \equiv \neg\varphi$.

7.2.5 The CWA and probability theory

Despite the fact that the CWA, as described above, is useful in a medical context, a limitation is that, being fully based on logic, it is unable to deal with uncertainty. Thus, there is a need to extend the ideas described above towards probability theory.

In the context of Bayesian networks, the use of the CWA is comparable, but somewhat different. We use as much information as available from our clinical database to fill in the nodes of the network. The information used for inference in the network is called *evidence*. Let us assumed that we have particular evidence e concerning a patient, excluding a variable X, such that $0 < \Pr(X = \textit{yes} \mid e) < 1$. This is interpreted as saying that X is not fully known. However, if variable X is potentially part of the patient evidence, then we may, adopting the CWA, assume that $X = \textit{no}$ holds, thus as a consequence:

$$\Pr(X = \textit{yes} \mid e \cup \{X = \textit{no}\}) = 0$$

We have that $\mathrm{CWA}_{\mathrm{Pr}}(e) = e \cup \{X = \textit{no}\}$. This means that we always compute

$$\Pr(X \mid \mathrm{CWA}_{\mathrm{Pr}}(E))$$

for any set of evidence E.

With regard to our Bayesian network about VAP, we have applied the CWA to data of sputum cultures, but in a restrictive fashion. When no pathogens are found, but a sputum culture has been taken, it is assumed that all pathogens are absent. This is a negative test result and not really an instance of the CWA. When culture data is not missing for a specific patient day, and one or more of the seven possible groups of pathogens are found positive, we deduce that the other groups of pathogens will be absent. These negative culture data will then augment the patient evidence e, and this is an instance of the CWA.

7.2.6 Evaluation of therapeutic performance

The performance of the Bayesian network, before and after taking into account the CWA for probability theory, was subsequently evaluated. The following approach was taken:

1. We first assume that when a patient is colonised by one or more microorganisms on a given day t_c, that on the three successive days, i.e. $t_c + 1,\ t_c + 2,\ t_c + 3$, this patient is still colonised. This assumption seems valid, as (1) clinical cultures usually are not performed daily and (2) when treated with antibiotics, microorganisms are not eradicated immediately.

2. Secondly, we interpret the clinical culture data as follows:

   (a) When no microorganisms are found, this is interpreted as a *negative* result for the seven nodes in the network we fill in the evidence: 'colonisation_by_pathogen$_i$ = no', $1 \leq i \leq 7$.

   (b) When one or more pathogens are found positive in the culture, these pathogens are processed in the network as being present; for the other pathogens, however, we assume that they are absent. This is in instance of the CWA in probability theory.

   (c) When on a specific day no cultures were performed, i.e. culture data is missing, no evidence is filled in in the network.

To test whether making these assumptions improves the therapeutic performance of the network, we used a temporal database of 17710 records for this purpose. This database contains data for more than 2000 patients, admitted to the ICU between 1999 and 2002 in the University Medical Center Utrecht, The Netherlands. For 157 of these 17710 episodes, a VAP was diagnosed according to the judgment of two infectious-disease specialists (IDS), who based their judgment upon a decision tree (Figure 3.4). During the period of seven days from time-point of diagnosis, the patient is treated with antibiotics. When the number of days after the day of VAP is less than 7, we assume that this patient recovered, or died. See Figure 7.2 for an example of a time line which shows the evidence and actions for a patient from the time point of admission to the ICU until the time point of discharge from the ICU. Using the CWA and the clinical evidence as shown on the time line, on day $t_{VAP} + 3$, i.e. on the third

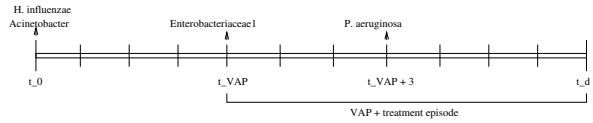


Figure 7.2: Example time line. On t_0, clinical cultures showed that this patient was colonised with *H. influenzae* and *Acinetobacter* spp. Four days later, here denoted by t_{VAP}, the patient was diagnosed with VAP and antimicrobial treatment was started to cover the Enterobacteriaceae1 pathogen and on day 3 of the treatment pathogen *P. aeruginosa*. At time point t_d, the patient was discharged from the ICU.

day of treating this VAP patient, the network would look as shown in Figure 7.3.

7.3 Results

Two infectious-disease specialists selected optimal antimicrobial treatment for the same patient data as we used in our analysis. By doing so, we were able to compare their therapy advice, here considered the *gold standard*, to the treatment selected by the Bayesian network. The information shown in Table 7.1 can be interpreted as follows: each combination of causative pathogens was put in the first column. The second column denotes the number of occurrences of each combination, adding up to the total number of VAP patients, i.e., 157. The third column is the antibiotic treatment, prescribed by the infectious-disease experts. Broadness of spectrum, matching to the the treatment, can be found in the last column. The results of the evaluation of the Bayesian network, with and without using the CWA, are shown in Table 7.2.

To be able to draw conclusions from Tables 7.1 and 7.2, we have compared them. We have done this in such a way that we are able to say for each group of causative pathogens whether the advised antibiotics by the models are right or wrong. Wrong here means that the antibiotic spectrum, as advised by one of the two interpretations of the Bayesian network, is either too narrow or too broad. Table 7.3 summarises the results of this comparison.

7.4 Conclusions and discussion

In this paper, we have described a way to improve the performance of a Bayesian network for treatment selection in patients with VAP. The way in which data about sputum cultures was interpreted previously gave rise to the selection of broad-spectrum antibiotics in most of the cases. This is also due to the fact that overall susceptibility is modelled by means of a logical-AND gate. However, prescription of unnecessary broad-spectrum antibiotics is undesirable from a

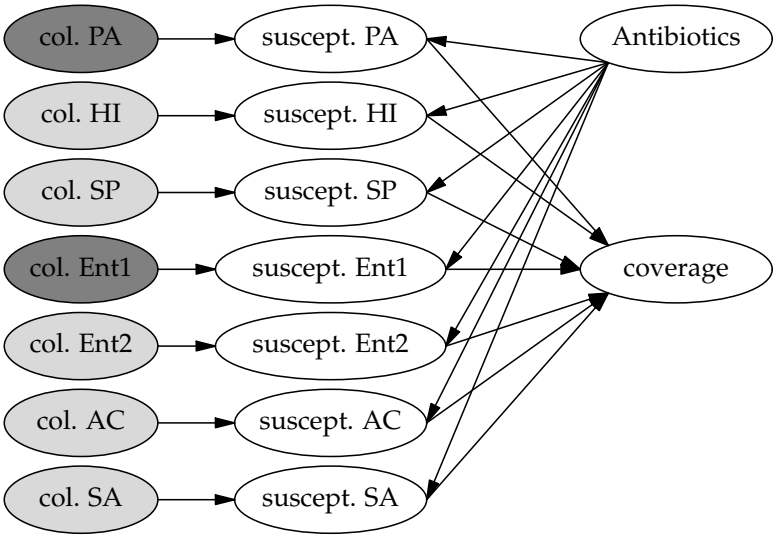


Figure 7.3: The therapeutic part of the network including clinical evidence as shown on the time line on the third day of treatment ($t_{VAP} + 3$). Positive evidence is denoted by the darkest shaded ellipses, whereas the lightest shaded ellipses denote the pathogens which are assumed negative, using the CWA.

medical point of view, as it stimulates the development of antibiotic resistance in pathogens. We subsequently adopted the closed-world assumption in interpreting sputum-culture data when filling in evidence in the Bayesian network, which resulted in an improved therapeutic performance of the model. When using the CWA, the percentage of patients for whom a correct antibiotic therapy was advised by the Bayesian network, improved by 50%. We conclude that the CWA, which was originally developed in the context of relational database theory, may also be useful when interpreting clinical data using a Bayesian network.

In the near future, we plan to investigate ways to model interaction among variables in a more natural way using causal independence models [102].

Table 7.1: Gold standard. P1 and P2 represent the causative pathogens. Abbreviations used for the pathogens: SA: *S. aureus*; SP: *S. pneumoniae*; HI: *H. influenzae*; Ent1: Enterobacteriaceae1; Ent2: Enterobacteriaceae2; AC: *Acinetobacter*; PA: P.aeruginosa. Freq. stands for 'frequency'. Abbreviations for antibiotic spectrum: v = very narrow; n = narrow; i = intermediate; b = broad.

P1	**P2**	**Freq.**	**Antibiotics**	**Spectrum**
SA		25	Floxapen	vn
SP		4	Penicylin	vn
HI		8	Augmentin	n
Ent1	Ent2	4	Ceftriaxone	n
Ent2	SA	4	Ceftriaxone	n
SA	SP	3	Floxapen	vn
HI	SP	6	Augmentin	n
Ent1		27	Ceftriaxone	n
Ent2		18	Ceftriaxone	n
SA	HI	5	Ceftriaxone	n
AC	Ent1	3	Ceftriaxone	n
PA		19	Ceftazidime	i
Ent2	AC	3	Ceftriaxone	n
HI	PA	1	Ceftazidime	i
Ent2	SP	1	Cotrimoxazol	n
PA	Ent2	5	Ciproxin	b
AC	SA	1	Cotrimoxazol	n
PA	Ent1	3	Ceftazidime	i
Ent1	HI	2	Cotrimoxazol	n
PA	AC	1	Ceftazidime	i
SA	PA	1	Tazocin	b
SA	Ent1	2	Ceftriaxone	n
AC		6	Cotrimoxazol	n
Ent2	HI	1	Ceftriaxone	n
		4	Cotrimoxazol	n

Table 7.2: Results. The column 'Old' denotes the results for the original Bayesian network, whereas 'New' gives information about the performance of the Bayesian network, when using the CWA. Abbreviations for antibiotic spectrum: v = very narrow; n = narrow; i = intermediate; b = broad.

	Antibiotics & Spectrum			
Freq.	**Old**	**SP**	**New (CWA)**	**SP**
25	Clinda+aztr	i	Floxapen	vn
4	Clinda+aztr	i	Penicillin	vn
8	Clinda+aztr	i	Erythromycin	vn
4	Clinda+aztr	i	Clinda+aztr	i
4	Clinda+aztr	i	Clinda+aztr	i
3	Clinda+aztr	i	Floxapen	vn
6	Clinda+aztr	i	Erythromycin	vn
27	Meropenem	b	Ceftriaxone	n
18	Clinda+aztr	i	Clinda+aztr	i
5	Clinda+aztr	i	Ceftriaxone	n
3	Meropenem	b	Meropenem	b
19	Ceftazidime	i	Ceftazidime	i
3	Meropenem	b	Cotrimoxazol	n
1	Ceftazidime	i	Ceftazidime	i
1	Clinda+aztr	i	Clinda+aztr	i
5	Clinda+aztr	i	Clinda+aztr	i
1	Meropenem	b	Cotrimoxazol	n
3	Ceftazidime	i	Ceftazidime	i
2	Meropenem	b	Ceftriaxone	n
1	Meropenem	b	Ceftazidime	i
1	Clinda+aztr	i	Clinda+aztr	i
2	Meropenem	b	Ceftriaxone	n
6	Clinda+aztr	i	Cotrimoxazol	n
1	Meropenem	b	Clinda+aztr	i
4	Clinda+aztr	i	Metronidazole	vn

Table 7.3: Absolute and relative numbers of incorrectly and correctly advised antibiotics for both the original and new Bayesian-network model interpretation.

Model	Incorrect		Correct
	too narrow	too broad	
Original	6 (4%)	129 (82%)	22 (14%)
CWA	23 (14%)	31 (20%)	103 (66%)

[8]

Treatment Effects in a Clinical Bayesian Network using Boolean Threshold Functions

8.1 Introduction

Establishing an accurate diagnosis and choosing appropriate treatment for infections are desirable, especially when it concerns critically ill patients. In the intensive care unit (ICU) patients who depend on respiratory support are prone to develop *ventilator-associated pneumonia*, or VAP for short. Although it is important to start antimicrobial treatment for VAP as soon as possible, when indicated, unnecessary antimicrobial treatment will enhance selection of antibiotic-resistant pathogens, which may subsequently hamper the treatment future infections. Since the accurate diagnostic test for diagnosing VAP (i.e., bronchoscopy with quantitative microbiological cultures) is invasive, expensive and labour-intensive, some form of computer-based decision support could be helpful in the process of early diagnosis and treatment of VAP.

Previously, we have developed a computer-based decision-support system (DSS) that is aimed at assisting physicians with the diagnosis and treatment of VAP. The model underlying the DSS consists of a Bayesian network with an associated decision-theoretic part. The structure as well as the conditional probabilities and utilities were elucidated with the help of two infectious disease specialists. The resulting decision-theoretic model, or influence diagram, was translated into a Bayesian network, and this is the model currently used (See Ref. [50] for details concerning the model and the construction process of the model). The probability of VAP is computed using the diagnostic part of

The work described in this chapter has been accepted for publication in Artificial Intelligence in Medicine.

the Bayesian network. In addition, the therapeutic part of the network can be used to determine the best possible combination of antibiotics.

When prescribing antimicrobial treatment a physician intends to cover all microorganisms causing the infection, using an antibiotic with the narrowest possible spectrum. This policy aims at preventing antibiotic resistance and at saving costs [99]. This was already taken into account when constructing the DSS, described in more detail in Ref. [50]. To cover as many pathogens as possible by the antibiotic treatment advised by the DSS, a noisy-AND probabilistic model was used in the Bayesian network for the modelling of the probabilistic interactions of the effects of the prescribed antibiotics on the pathogens, taking into account colonisation by those pathogens. However, this approach yielded antibiotic choices that were considered much too broad.

The start of the current research reported in this paper is an analysis of the reason of this behaviour of the model, and a number of alternatives to define this particular conditional probability distribution are studied. As the noisy AND is a special case of the more general class of probabilistic models based on the Boolean threshold functions, Boolean functions that are defined in terms of the number of truths among Boolean values, it is also studied theoretically what happens when the AND is replaced by a threshold function. The resulting probabilistic models are called *noisy-threshold models* [102, 105]. Arguments are presented why we expect that a noisy-threshold model might work better than the noisy AND. Finally, behaviour of the various noisy-threshold probabilistic models is investigated by replacing the noisy AND used, using data of patients with VAP from the ICU of the University Medical Center, Utrecht. It is also investigated whether the therapeutic performance of the Bayesian network for VAP improves in this way.

Although we focus on an actual clinical problem – the prescription of antibiotics for patients with VAP – this problem can be seen as an instance of a common and important problem in medicine: the modelling of the effectiveness of treatment. In this sense, the results achieved here have a bearing on a clinical area wider than infectious disease.

The paper is organised as follows. In the next section, our earlier work on the development of a Bayesian network that is able to assist physicians in the diagnosis and treatment of VAP is briefly reviewed. In Section 8.3, the mathematical principles of causal independence models are discussed and noisy threshold models are introduced. Furthermore, three different models of antimicrobial coverage are constructed and analysed. In Section 8.4, the data and methods used in evaluating the Bayesian networks incorporating the threshold functions are described. The results achieved are commented on in Section 8.5. The paper is rounded off by some conclusions in Section 8.6.

8.2 A Bayesian network for the management of VAP

Bayesian networks, or BNs for short, have been introduced in the 1980s as a formalism to compactly represent and reason efficiently with joint probability

distributions. Bayesian networks are in particular well suited for the representation of causal relations within a specific domain of expertise.

Formally, a Bayesian network $\mathcal{B} = (G, \Pr)$ is an acyclic directed graph $G = (\mathbf{V}(G), \mathbf{A}(G))$ with set of vertices $\mathbf{V}(G) = \{V_1, \ldots, V_n\}$, corresponding to random variables, here denoted by the same letters or strings of characters, and a set of arcs $\mathbf{A}(G) \subseteq \mathbf{V}(G) \times \mathbf{V}(G)$, representing statistical dependences and independences among the variables. On the set of random variables, a joint probability distribution $\Pr(\mathbf{V}(G))$ is defined that is factorised according to the structure of the graph:

$$\Pr(\mathbf{V}(G))) = \prod_{V \in \mathbf{V}(G)} \Pr(V \mid \pi(V)),$$

where $\pi(V)$ stands for the variables corresponding to the parents of vertex V.

In the following, we will often make use of binary random variables. If the variable X assumes the value 'true' or 'yes', this will be sometimes indicates by x, whereas if X assumes the value 'false' or 'no', this will be indicated by $\neg x$.

The formalism of BNs supports the kind of reasoning under uncertainty that is typical for medicine when dealing with diagnosis, treatment selection, planning, and prediction of prognosis. Our clinical domain is restricted to patients who are mechanically ventilated and are at risk of developing VAP. Entities that play an important role in the development of VAP and that belong to the diagnostic part of the Bayesian network for VAP include: the duration of *mechanical ventilation*, the amount of *sputum*, *radiological signs*, i.e., whether the chest radiograph shows signs of an infection, body *temperature* of the patient and the number of *leukocytes* (white blood cells) [27]. The structure of the Bayesian network for VAP is shown in Figure 8.1. Mechanically ventilated ICU patients become colonised by bacteria. When colonisation of the lower respiratory tract occurs within 2–4 days after intubation, this is usually caused by antibiotic-sensitive bacteria, whereas after one week of intubation often antibiotic-resistant bacteria are involved in colonisation and infection. Such infections are more difficult to treat and immediate start of appropriate treatment in case of infection is, therefore, important. Duration of hospital stay and severity of illness are associated with an increased risk of colonisation and infection with Gram-negative bacteria. We modelled seven groups of microorganisms, each by one vertex in the Bayesian network, and the pathogenicity, i.e., the influence of that particular microorganism on the development of VAP, was included in the model. The presence of certain bacteria is influenced by antimicrobial therapy; however, a microorganism is susceptible only for particular antibiotics. Susceptibility, in this case, is stated as the sensitivity to or degree to which a microorganism is affected by treatment with a specific antibiotic. The susceptibility of each microorganism was taken into account while constructing the model. The infectious-disease experts assigned utilities, by definition quantitative measures of the strength of the preference for an outcome [101], to each combination of microorganism(s) and antimicrobial

drug(s) using a decision-theoretic model. These variables are included in the therapeutic part of the Bayesian network for VAP.

8.3 Causal independence modelling

Causal independence is a popular means to facilitate the specification of conditional probability distributions $\Pr(V_i \mid \pi(V_i))$ involving many parent variables $\pi(V_i)$. Its basic principles and some special forms are briefly discussed below.

8.3.1 Basic principles

Consider the conditional probability distribution $\Pr(E \mid C_1, \ldots, C_n)$, where the variable E stands for an *effect*, e.g., coverage, and the variables C_j, $j = 1, \ldots, n$, denote *causes*, e.g., colonisation by pathogens in combination with treatment by means of antibiotics. By taking a number of assumptions into account, which are summarised in Figure 8.2, it is possible to simplify the specification of $\Pr(E \mid C_1, \ldots, C_n)$. These assumptions are: (1) the causes C_j are assumed to be mutually independent, and (2) the variable E is conditionally independent of any cause variable C_j given the intermediate variables $I_1, \ldots, I_n$. In our domain the intermediate variable I_j stands for susceptibility of pathogen$_j$ to a specific antibiotic. Using basic probability theory, it follows that:

$$\Pr(e \mid C_1, \ldots, C_n) = \sum_{I_1, \ldots, I_n} \Pr(e \mid I_1, \ldots, I_n) \prod_{j=1}^{n} \Pr(I_j \mid C_j). \tag{8.1}$$

Now, if we assume that the probability distribution $\Pr(E \mid I_1, \ldots, I_n)$ that is specified for variable E expresses some deterministic function $f : I_1 \times \cdots \times I_n \rightarrow E$, with $I_j, E \in \{\bot, \top\}$, called an *interaction function*, an alternative formalisation is possible. Using the interaction function f and the causal parameters $\Pr(I_j \mid C_j)$, it follows that [103, 102, 106]:

$$\Pr(e \mid C_1, \ldots, C_n) = \sum_{f(I_1, \ldots, I_n) = e} \prod_{j=1}^{n} \Pr(I_j \mid C_j). \tag{8.2}$$

The result is called a *causal independence model* [103, 107, 102]. In this paper we assume that the function f in Equation (8.2) is a Boolean function. The consequences are that instead of a specification of a conditional probability distribution that is exponential in size, one only needs to specify a conditional probability distribution in terms of a linear number of parameters $\Pr(I_j \mid C_j)$ and a Boolean function f.

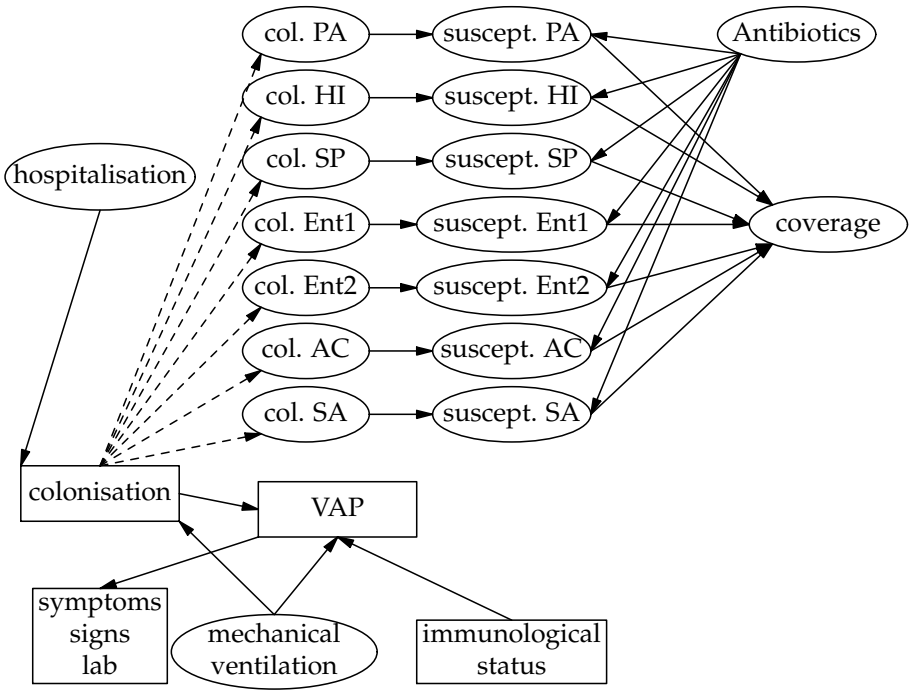


Figure 8.1: Abstract model of the Bayesian network for the management of VAP. Colonisation and VAP play a central role in this model. The duration of hospitalisation and mechanical ventilation influence colonisation (col.) of the patient. PA: *Pseudomonas aeruginosa*; HI: *Haemophilus influenzae*; SP: *Streptococcus pneumoniae*; Ent{1,2}: Enterobacteriaceae{1,2}; SA: *Staphylococcus aureus*; AC: *Acinetobacter* spp.. Each pathogen is susceptible (suscept.) to particular antibiotics and an optimal coverage of pathogens is what the model aims to achieve. The duration of mechanical ventilation, immunological status and colonisation influence the development of VAP. When a patient is diagnosed with VAP, the patient often has symptoms such as an increased body temperature. Boxes denote entities or processes which are observed; processes that change or can be changed are denoted by ellipses.

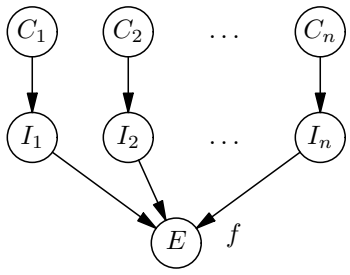


Figure 8.2: Causal independence model.

Systematic analyses of the global probabilistic patterns in causal independence models based on restricted Boolean functions were presented in Ref. [102] and Ref. [108]. There are 2^{2^n} different n-ary Boolean functions [109, 110]; thus, the potential number of causal interaction models is huge. If we assume that the order of the cause variables does not matter, the Boolean functions become symmetric; formally, an interaction function f is called *symmetric* if [110]

$$f(I_1, \ldots, I_n) = f(I_{j_1}, \ldots, I_{j_n})$$

for any index function $j : \{1, \ldots, n\} \to \{1, \ldots, n\}$ [110]. The number of different Boolean function reduces then to 2^{n+1}. Examples of symmetric binary Boolean functions include the logical OR, AND, exclusive OR and bi-implication. An example of a general, possibly non-binary, symmetric Boolean function is the *exact* Boolean function e_k, which is defined as follows [110, 102]:

$$e_k(I_1, \ldots, I_n) = \begin{cases} \top & \text{if } \sum_{j=1}^{n} \nu(I_j) = k \\ \bot & \text{otherwise} \end{cases} \tag{8.3}$$

with $k \in \mathbb{N}$, and

$$\nu(I) = \begin{cases} 1 & \text{if } I = \top \\ 0 & \text{otherwise} \end{cases}$$

where $\top$ stands for 'true', and $\bot$ for 'false'. Symmetric Boolean functions can be decomposed in terms of the exact functions e_k as follows [110]:

$$f(I_1, \ldots, I_n) = \bigvee_{k=0}^{n} e_k(I_1, \ldots, I_n) \wedge \gamma_k \tag{8.4}$$

where γ_k are Boolean constants only dependent of the function f. Using this result, the conditional probability of the occurrence of the effect E given the causes $C_1, \ldots, C_n$ can be decomposed in terms of probabilities that exactly l amongst the intermediate variables $I_1, \ldots, I_n$ are true, as follows:

$$\Pr(e \mid C_1, \ldots, C_n) = \sum_{\substack{0 \leq l \leq n \\ \gamma_l}} \sum_{e_l(I_1, \ldots, I_n)} \prod_{j=1}^{n} \Pr(I_j \mid C_j). \tag{8.5}$$

Thus, Equation (8.5) yields a general formula to compute the probability of the effect in terms of exact functions in any causal independence model where an interaction function f is a symmetric Boolean function.

The interaction among variables modelled by the susceptibility, or coverage variables, as shown in Figure 8.1, was modelled by assuming f to be a logical AND. The resulting probabilistic model $\Pr(E \mid C_1, \ldots, C_n)$ is usually called the *noisy-AND* model, or *noisy AND* for short. The probability distribution of the variable that represents the overall susceptibility (coverage in Figure 8.1), models the conjunctive effect of the seven different pathogens. This principle

is modelled by a probability distribution $\Pr(E \mid C_1, \ldots, C_n)$ that is defined as in Equation (8.1) by the noisy AND, yielding the following equation:

$$\Pr(\text{coverage} \mid \text{Colonisation}_1, \ldots, \text{Colonisation}_n, \text{Antibiotics}) = \prod_{j=1}^{n} \Pr(\text{susceptibility-pathogen}_j \mid \text{Colonisation}_j, \text{Antibiotics}).$$

By adopting this modelling approach, the network attempts to cover *all* pathogens in choosing appropriate antimicrobial treatment. As revealed in our dataset, patients were colonised by at most 3 pathogens. Therefore, covering all 7 possible groups of pathogens is simply too much and results, most of the time, in antimicrobial treatments that are too broad. Furthermore, in modelling the effect of antibiotics on the susceptibility, one also needs to express what effect the absence of colonisation has in the absence or presence of antimicrobial treatment. In the next section, both issues will be explored in detail. Both issues have raised doubts on the appropriateness of the noisy AND for the modelling of interactions concerning coverage of bacteria by antibiotics.

As the methods which are developed subsequently are generic, a slightly more general terminology than the one above will be adopted. Thus, in the following, 'coverage' will be abbreviated to O (outcome), 'susceptibility-pathogen$_j$' by S_j (susceptibility), 'Colonisation_j' by C_j (causal factor) and 'Antibiotics' by M (medication), i.e., the conditional probability distribution that is studied is of the form

$$\Pr(O \mid C_1, \ldots, C_n, M). \tag{8.6}$$

8.3.2 Bayesian network coverage models

As argued in Section 8.1, it is generally felt that clinicians could be more careful in the prescription of antibiotics as they tend to prescribe antibiotics that are either not needed or have a too broad spectrum [111]. A symmetric Boolean function that is useful in designing a generalised version of the noisy AND is the *threshold function* τ_k, which simply checks whether there are at least k trues among its arguments; it is defined as follows:

$$\tau_k(I_1, \ldots, I_n) = \begin{cases} \top & \text{if } \sum_{j=1}^{n} \nu(I_j) \geq k \\ \bot & \text{otherwise} \end{cases}$$

where again $\nu(I_j)$ equals 1 if I_j equals $\top$ (true) and 0 otherwise [110]. Let us denote a conditional probability of the effect E given causes $C_1, \ldots, C_n$ in a noisy-threshold model with interaction function τ_k as $\Pr_{\tau_k}(e \mid C_1, \ldots, C_n)$. Then, from Equation (8.5) it follows that:

$$\Pr_{\tau_k}(e \mid C_1, \ldots, C_n) = \sum_{k \leq l \leq n} \sum_{e_l(I_1, \ldots, I_n)} \prod_{j=1}^{n} \Pr(I_j \mid C_j). \tag{8.7}$$

Note that the Boolean AND corresponds to the threshold function τ_k with $k = n$, whereas the Boolean OR is a threshold function with $k = 1$. Hence, the AND

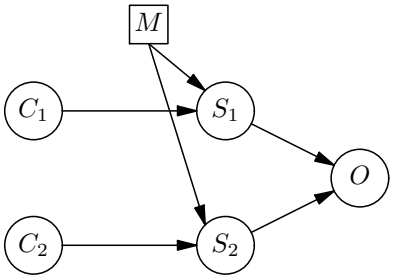


Figure 8.3: Prototypical structure for investigating outcome of medication M, based on the susceptibility of the causal factors C_j to that medication. The meaning of the used abbreviations is as follows: C_j: causal factor j; S_j susceptibility to medication; M: treatment by antimicrobial medication; O: overall outcome. The variable O is taken as a measure of success of the treatment.

and OR can be seen as the extremes of a spectrum of Boolean functions based on the threshold function.

The prototypical structure that can be used to model the outcome O of medication M on the causal factors C_j is shown in Figure 8.3; it offers a possible way to model the conditional probability distribution of Formula (8.6).

The resulting decomposition then has the following form:

$$\Pr_{\tau_k}(o \mid C_1, \ldots, C_n, M) = \sum_{k \leq l \leq n} \sum_{e_l(S_1,\ldots,S_n)} \prod_{j=1}^{n} \Pr(S_j \mid C_j, M). \tag{8.8}$$

There are various choices possible for the conditional probability distributions

$$\Pr(S_j \mid C_j, M) \tag{8.9}$$

that are associated with the vertex S_j (Susceptibility) indicated in Figure 8.3. Let us for simplicity's sake assume that M takes only two values: 'no' (sometimes indicated by $\neg m$), which corresponds to the situation where no treatment is given, and 'yes' (sometimes indicated by m), which corresponds to the situation that drugs are given that influence some susceptibilities S_j. For the example, we also assume this probability distribution to be deterministic, which normally will not hold in real life. One way to define the probability distribution (8.9) is as follows:

$$\Pr(s_j \mid C_j, M) = \begin{cases} 0 & \text{if } C_j = \text{yes}, M = \text{no} \\ 1 & \text{otherwise} \end{cases}$$

We call this definition the '**susceptibility I model**'. The implication of this definition is that treatment is always successful in the absence of the causal factors, such as the absence of colonisation in the case of VAP. Although this may seem natural at first sight, a disadvantage is that when optimising the medication, it is likely that causal factors that have *no* effect, will have a major

influence on the choice of medication. In case of VAP this means that absence of colonisation is consistent with medication.

Another way to model susceptibility might be to change the probability distribution above by stating that $\Pr(s_i \mid \neg c_i, \neg m) = 1$, whereas $\Pr(s_i \mid \neg c_i, m) = 0$. We call this definition the '**susceptibility II model**'; it has the advantage that when optimising medication, it is unlikely that a drug will be selected in the absence of the causal factor. However, a disadvantage is that the model may select *no* medication when only a few causal factors are active, and covering the inactive causal factors would already be optimal. In the case of VAP this corresponds to covering absent colonisation by no medication.

A third way to model likelihood of susceptibility is to take almost the reverse of the definition above:

$$\Pr(s_j \mid C_j, M) = \begin{cases} 1 & \text{if } C_j = \text{yes}, M = \text{yes} \\ 0 & \text{otherwise} \end{cases}$$

We call this definition the '**susceptibility III model**'. This implies that as long as causal factors are active, the optimal policy is to cover those by medication. For VAP this means coverage of the microorganisms colonising a patient by means of appropriate treatment.

Another issue that should be considered concerns the choice of the Boolean interaction function f corresponding to the deterministic probability distribution $\Pr(O \mid S_1, \ldots, S_n)$. In the original model, we took the logical AND as an implementation for this probability distribution. Using the threshold function τ_k with $k \neq 1, n$, may result in a more intuitive model. Using a particular way of exploiting the noisy threshold functions, the network might be redesigned such that it only covers a fixed number of the causal factors. For the VAP model, it might, for example, cover 1 ($k = 1$), i.e. the noisy-OR, 2 ($k = 2$), 3 ($k = 3$), 4 ($k = 4$), 5 ($k = 5$) or 6 ($k = 6$) of the causative pathogens compared to the noisy-AND gate, where all pathogens, i.e. $k = 7$, are taken into account. However, as the behaviour of the entire Bayesian network model shown in Figure 8.3 is not only determined by the probability distribution $\Pr(O \mid S_1, \ldots, S_n)$ but, in addition, also by $\Pr(S_i \mid C_j, M_j)$ we first need to study the various behaviours obtained by combining various definitions of these two conditional probability distributions. Next, we explore the consequences of choosing a particular combination of these two conditional probability distributions for the case of VAP.

The behaviour of the susceptibility I model, as shown in Figures 8.4(a) and (b), indicates that no distinction is made between presence and absence of colonisation by a particular microorganism. As a consequence, all three threshold functions indicate coverage even when the patient is only colonised by one microorganism. Figures 8.4(a) and (b) also indicate that as soon as effective treatment against the single microorganism by which the patient is colonised is selected, the noisy AND concludes that it is able to cover all, i.e., both microorganism by which the patient is and is not colonised.

The results for the susceptibility II model with no medication, as shown in Figure 8.4(c), are identical to those of Figure 8.4(a). However, as medication

is no longer assumed to cover microorganisms by which the patient has not been colonised, the noisy AND and noisy τ_2 models indicate failure of coverage. This behaviour is identical to the probabilistic behaviour shown in Figure 8.4(f).

As shown in Figure 8.5(b) and (d), the susceptibility II and III models also give identical results if the patient is colonised by two microorganisms and being appropriately treated. However, as Figure 8.5(a) and (c) indicate, the susceptibility II model indicates coverage of the single microorganism by which the patient has not been colonised when no medication is given, whereas the susceptibility III model indicates failure. Clearly, the susceptibility II model encodes a sort of symmetry between no treatment in the absence of colonisation and treatment in the presence of colonisation, where the susceptibility III model is asymmetric and incorporates the implicit assumption that it is unlikely that patients are completely uncolonised and that taking this situation into account is therefore unnecessary. The susceptibility model III also clearly indicates probable coverage of microorganisms, as shown in Figure 8.5(f) in the vertex concerning the noisy AND, which appears to be another advantage.

Probability distributions $\Pr(E \mid C_1, \ldots, C_n)$ defined in terms of Boolean threshold functions using the same probabilistic parameters $\Pr(I_k \mid C_k)$ have the following important property:

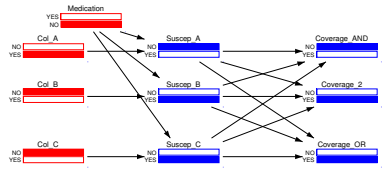
$$\Pr_{\tau_k}(e \mid C_1, \ldots, C_n) \geq \Pr_{\tau_{k+1}}(e \mid C_1, \ldots, C_n) \tag{8.10}$$

where again $\Pr_{\tau_k}$ is a probability distribution defined in terms of the threshold function τ_k. The proof follows directly from Equation (8.7), as according to this equation

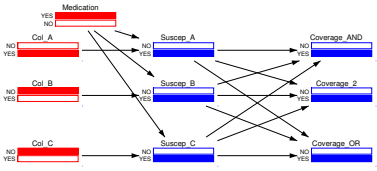
$$\Pr_{\tau_k}(e \mid C_1, \ldots, C_n) + \sum_{e_{k+1}(I_1,\ldots,I_n)} \prod_{j=1}^{n} \Pr(I_j \mid C_j) = \Pr_{\tau_{k+1}}(e \mid C_1, \ldots, C_n),$$

and $\sum_{e_{k+1}(I_1,\ldots,I_n)} \prod_{j=1}^{n} \Pr(I_j \mid C_j) \geq 0$. This property is consistent with the probabilities of coverage depicted in the bar charts in Figure 8.4 and 8.5, as the probability for the noisy AND is never above that for τ_2, which is never above that for the noisy OR.

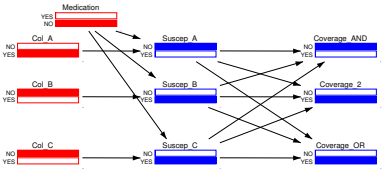
In the following we therefore investigate properties of the threshold function, and subsequently study its use in improving the Bayesian network model shown in Figure 8.1.



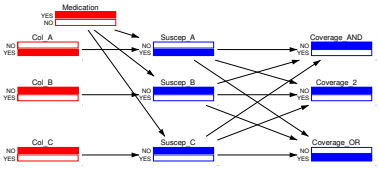
(a) Susceptibility I model; colonisation by 1 microorganism, no medication



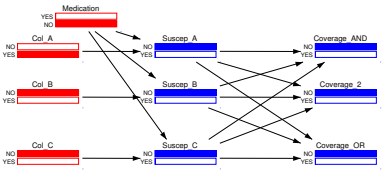
(b) Susceptibility I model; colonisation by 1 microorganism, effective medication



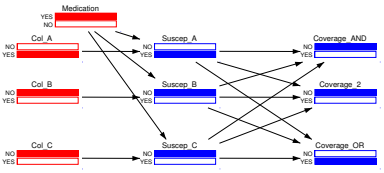
(c) Susceptibility II model; colonisation by 1 microorganism, no medication



(d) Susceptibility II model; colonisation by 1 microorganism, effective medication

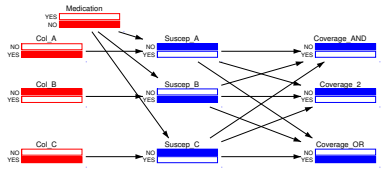


(e) Susceptibility III; colonisation by 1 microorganism, no medication

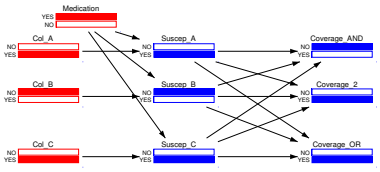


(f) Susceptibility III model; colonisation by 1 microorganism, effective medication

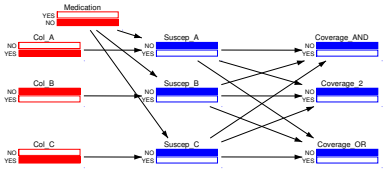
Figure 8.4: Comparison of the two different susceptibility models, assuming the patient is colonised by one microorganism, with three threshold interactions functions: the AND (τ_3), τ_2 and OR (τ_1).



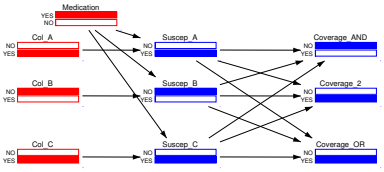
(a) Susceptibility II model; colonisation by 2 microorganisms, no medication



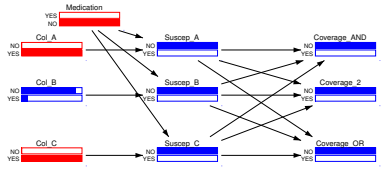
(b) Susceptibility II model; colonisation by 2 microorganisms, effective medication



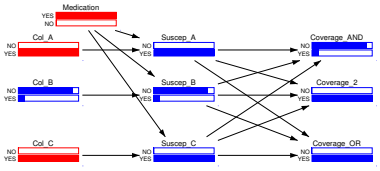
(c) Susceptibility III model; colonisation by 2 microorganisms, no medication



(d) Susceptibility III model; colonisation by 2 microorganisms, effective medication



(e) Susceptibility III model; colonisation by 2 microorganisms, no medication



(f) Susceptibility III model; colonisation by 2 microorganisms, effective medication

Figure 8.5: Comparison of susceptibility models II and II, assuming the patient is colonised by two microorganisms, with three threshold interactions functions: the AND (τ_3), τ_2 and OR (τ_1).

8.3.3 Counting functions

So far we have assumed that the probability distributions

$$\Pr(O \mid S_1, \ldots, S_n)$$

are defined as single big tables. However, it is possible to decompose these probability distributions using a basic property of symmetric Boolean functions [110]. The values of a symmetric Boolean function can be represented as a vector $(v_0, \ldots, v_n)$ such that $f(I_1, \ldots, I_n) = v_i$ if $I_1 + \cdots + I_n = i$. This means that it suffices to count the number of trues in the arguments of f, interpreting 'true' arithmetically as 1 and 'false' as 0, and this can be done incrementally, as addition is commutative and associative: $I_1 + \cdots + I_n = (\cdots((I_1 + I_2) + I_3) + \cdots + I_{n-1}), I_n) = i$. The probability distribution that corresponds to this counting is defined in terms of the following conditional probability distributions:

$$\begin{aligned}
\Pr(O_1 = \nu(I_1, I_2) \mid I_1, I_2) &= 1 \\
\Pr(O_2 = \nu(O_1, I_3) \mid O_1, I_3) &= 1 \\
&\vdots \\
\Pr(O_{n-2} = \nu(O_{n-3}, I_{n-1}) \mid O_{n-3}, I_{n-1}) &= 1 \\
\Pr(O_{n-1} = v_i \mid O_{n-2}, I_n) &= \begin{cases} 1 & \text{if } \nu(O_{n-2}, I_n) = i \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

Note that the last variable, $O_{n-1} \equiv O$, is binary, whereas the other O_k variables, $1 \leq k \leq n-2$, take values from the set $\{0, \ldots, \nu(o_{k-1}, i_{k+1})\}$, where o_{k-1} indicates the maximum value of the random variable O_{k-1}. The resulting Bayesian network structure, when the susceptibility variables S_k are taken as the intermediate variables I_k, is shown in Figure 8.6.

For a threshold function τ_k it is only necessary that random variables take values out of the set $\{0, 1, \ldots, k\}$, as when the maximum k is reached, $P(O \mid S_1, \ldots, S_n) = 1$. An example is shown in Figure 8.7.

8.4 Validation

The usefulness of methods described above has been investigated for the Bayesian network concerning VAP, using data of ICU patients. The characteristics of the data are described in the next section, after which we return to the problem of the prescription of antibiotics to patients with VAP.

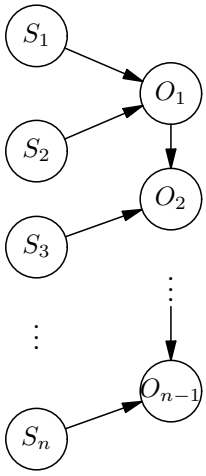
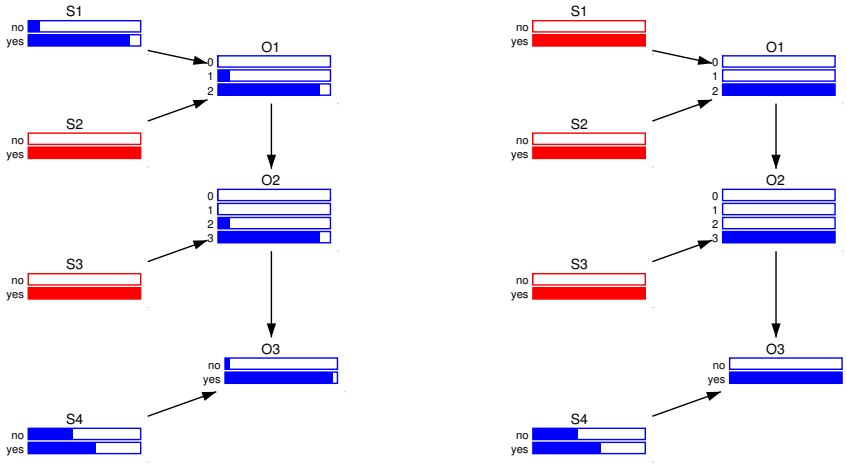


Figure 8.6: Decomposition of the conditional probability distribution $\Pr(O \mid S_1, \ldots, S_n)$.



(a) Posterior probabilities for 2 observations

(b) Posterior probabilities for 3 observations

Figure 8.7: Example Bayesian network implementation of the decomposition of threshold function τ_3.

Table 8.1: Reference standard: frequency of VAP-causing pathogens.

	Patients with VAP		
Causative pathogens	Monomicrobial $N = 107$ episodes (70%)	Polymicrobial $N = 46$ episodes (30%)	Total $N = 153$ episodes (100%)
Pseudomonas aeruginosa	19	11	30
Acinetobacter spp.	6	8	14
Enterobacteriaceae 1	29	17	46
Enterobacteriaceae 2	17	15	32
Staphylococcus aureus	25	16	41
Haemophilus influenzae	8	14	22
Streptococcus pneumoniae	3	11	14
Total number of pathogens	107	92	199

8.4.1 ICU data

We used a temporal database with 17710 records, each record representing a period of 24 hours of a mechanically ventilated patient in the ICU. The database contains information of 2233 distinct patients, admitted to the ICU of the University Medical Center Utrecht between 1999 and 2002. For 157 of these 2233 patients, a VAP was diagnosed according to the judgement of two infectious-disease specialists (IDS), which was subsequently considered as the *reference standard*. Four of 157 patients with VAP were excluded from analyses, as these infections were caused by a type of pathogen that was not modelled in the Bayesian network model. The distribution of all monomicrobial (caused by one pathogen) and polymicrobial (caused by 2 or more pathogens) among patients with VAP is shown in Table 8.1.

8.4.2 Methods

In order to improve the therapeutic performance of the Bayesian network, the network was inspected in detail. Points for possible improvement that were identified included the way pathogen coverage was modelled for the assessment of the conditional probability distribution

$$\Pr(\text{Coverage} \mid \text{Colonisation}_1, \ldots, \text{Colonisation}_7, \text{Antibiotics}),$$

as described in detail in Section 8.3.

Prescribing antibiotics that cover all likely pathogens is not an easy task for non-specialists. Normally, a fixed list of antimicrobials to which pathogens are susceptible, so-called *susceptibility patterns*, is available. As antibiotic resistance patterns differ between countries and even hospitals, this list may be different for each hospital. When susceptibility tests indicate resistance of the pathogen against antimicrobial a, another antibiotic, or combination of antibiotics, should be prescribed. In the model, susceptibility was incorporated using

Table 8.2: Antibiotics and their effectiveness (+). PA: *P. aeruginosa*; AC: *Acinetobacter* spp.; Ent{1,2}: Enterobacteriaceae; SA: *S. aureus*; HI: *H. influenzae*; SP: *S. pneumoniae*.

	Antibiotic	**Spectrum**	**Effectiveness.**						
			PA	AC	Ent1	Ent2	SA	HI	SP
A	amoxicillin	vn						+	+
B	amoxicillin / clavulanic acid	n						+	+
C	benzyl penicillin	vn							+
D	ceftazidime	i	+		+			+	+
E	ceftriaxone	n		+	+	+	+		
F	ciprofloxacin	b	+	+	+	+			
G	cotrimoxazole	n				+			
H	flucloxacillin	vn					+		
I	meropenem	b	+	+	+	+	+	+	+
J	vancomycin	vn					+		
K	no antibiotics	–							

data from the Department of Medical Microbiology, indicated that in a particular percentage of cases pathogen p was susceptible to antibiotic a.

As described earlier, there are several types of antibiotics; some antibiotics have a narrow spectrum and are effective against specific pathogens, whereas other antibiotics have a broad spectrum, that usually cover
difficult-to-treat pathogens. In addition, two groups of pathogens are clinically distinguished: *early-onset* and *late-onset pathogens*. The former are pathogens that colonise patients predominantly during the first 5 days of ICU admission, whereas the latter pathogens mainly occur after day 6 of ICU admission. Subdividing the pathogens modelled in our Bayesian network yields:

Early	*S. aureus*
	H. influenzae
	S. pneumoniae
Late	Enterobacteriaceae
	Acinetobacter spp.
	P. aeruginosa

For each pathogen, we selected a number of antibiotics that are highly effective for treatment, based on laboratory data. For some pathogens, these effective antibiotics overlap (Table 8.2). Some of these antibiotics overlap (to say >90%), for example, amoxicillin is a very narrow spectrum antibiotic that is effective against (in other words, covers) H. influenzae (HI) and S. pneumoniae (SP). Furthermore, in general, one would expect broad-spectrum antibiotics to have better coverage numbers than narrow-spectrum antibiotics.

For the assessment of the coverage behaviour of the Bayesian network, the overall coverage for the pathogens of all 153 episodes of VAP was calculated, using different susceptibility models and threshold functions. In particular, we

explored the question how well the model was able to cover the pathogens. To answer this question, the following assumptions were made:

1. presence of VAP was assumed based on reference standard and, therefore, the VAP vertex in the Bayesian-network model was instantiated;

2. based on previous endotracheal culture data, colonisation vertices were instantiated to denote either the presence or absence of pathogens colonising a patient.

8.5 Results

The results of using different threshold functions and susceptibility models were computed. The probability of coverage, which was between 0% and 100%, denotes how well the model was able to cover the present VAP-causing pathogens. However, these probabilities may not be correct. For example, it might be the case that according to the model there is 100% coverage of a pathogen by an antibiotic, where in reality this probability should be 0%.

The tables have been split up in terms of early- and late-onset VAP, as well as by the number of VAP-causing pathogens; Nr. 1 denotes that the patient was infected by one pathogen (monomicrobial episodes), where Nr. 2 denotes that the patient has been infected by 2 pathogens (polymicrobial episodes). Table 8.5, for example, shows the results for the susceptibility I model for threshold functions τ_1 (equivalently, the noisy-OR) and τ_2. In the 'Patho' column, the name of the pathogen is listed. The mean coverage for the in total 13 early-onset polymicrobial episodes of VAP when prescribing antibiotic B, i.e., amoxicillin-clavulanic acid, is 97. Prescribing no antibiotics in this case would not be advisable, as column K indicates that coverage would then be zero.

The various definitions of Boolean threshold functions, from τ_1 (noisy-OR), $\tau_2, \ldots, \tau_7$ (noisy AND) were combined with the three susceptibility models defined above. The following tables summarise the results obtained:

Susceptibility I model: Tables 8.3 to 8.8 show the results for susceptibility model I for threshold functions τ_k, with $k = 1, \ldots, 7$.

Susceptibility II model: For $k = 1, \ldots, 5$ the coverage results obtained for this model are identical to those obtained for susceptibility model III, with the exception of the probabilities in column K, i.e., the case when no antibiotics are prescribed, which are always 0% for model III and 100% for model II. For $k = 6, 7$ the pathogen coverages for susceptibility model II are fully equal to the coverages for susceptibility model III. To save space, the outcome tables for model II have therefore been omitted.

Susceptibility III model: Tables 8.9 to 8.14 show the results for susceptibility model III for the various threshold functions τ_k, for $k = 1, \ldots, 7$.

Table 8.3: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SI,** $k = 1, 2$**)** model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		100					100		100		0
1	SA	25					100			100	100	100	0
	HI	8	100	100		100					100		0
	SP	3	100	100	100	100					100		0
2		33						100			100		0
1	PA	19				100		100			100		0
	AC	6					100	100			100		0
	Ent1	29				100	100	100			100		0
	Ent2	17					100	100	100		100		0

Table 8.4: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SI,** $k = 3$**)** model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		99					100		100		0
1	SA	25					100			100	100	100	0
	HI	8	98	99		99					100		0
	SP	3	100	100	100	100					100		0
2		33						100			100		0
1	PA	19				100		100			100		0
	AC	6					100	100			100		0
	Ent1	29				100	100	100			100		0
	Ent2	17					100	100	100		100		0

Note that the changes in probabilities when going from τ_1 to τ_7 is according to Property (8.10), and this is thus as expected. What is important is to look for cases where the coverage, computed using the Bayesian network, becomes very low, even though the antibiotics are known to be effective, or very high, even though the antibiotics are known to be ineffective. In the next section, the clinical implications of these results are discussed in detail.

8.6 Conclusions and discussion

In this chapter, we have shown that by reconsidering the modelling of interactions between the random variables in a Bayesian network, it is possible to refine its performance. We used a Bayesian network for the diagnosis and treatment of ventilator-associated pneumonia as an example. Intensive use was made of the theory of causal independence, which not only facilitates the as-

Table 8.5: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SI,** $k = 4$**)** model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		97					97		100		0
1	SA	25					100			95	100	96	0
	HI	8	86	98		98					100		0
	SP	3	93	97	67	97					100		0
2		33						100			100		0
1	PA	19				100		100			100		0
	AC	6					100	100			100		0
	Ent1	29				99	99	100			100		0
	Ent2	17					98	98	99		100		0

Table 8.6: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SI,** $k = 5$**)** model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		90					95		100		0
1	SA	25					98			88	100	88	0
	HI	8	74	90		93					100		0
	SP	3	37	90	33	87					100		0
2		33						98			100		0
1	PA	19				98		100			100		0
	AC	6					100	100			100		0
	Ent1	29				97	96	100			100		0
	Ent2	17					93	96	98		100		0

sessment of probability tables by allowing the specification of a table in terms of a linear number of parameters of the form $\Pr(I_j \mid C_j)$, but also allows taking into account domain characteristics [102]. This was clearly shown for our Bayesian network concerning VAP, where motivation was derived from the domain of infectious diseases, indicating that only specific noisy-threshold model might be appropriate for the modelling of the interaction between pathogens and antimicrobial treatment with respect to susceptibility.

Table 8.7: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SI,** $k = 6$**)** model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		62					88		100		0
1	SA	25					85			63	100	64	0
	HI	8	71	73		78					100		0
	SP	3	33	50	33	53					100		0
2		33						91			99		0
1	PA	19				97		97			99		0
	AC	6					98	98			100		0
	Ent1	29				91	82	100			100		0
	Ent2	17					67	94	89		99		0

Table 8.8: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SI,** $k = 7$**)** (noisy-AND gate) model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		41					66		98		0
1	SA	25					49			24	97	24	0
	HI	8	46	60		62					98		0
	SP	3	33	33	36	33					95		0
2		33						68			92		0
1	PA	19				70		85			81		0
	AC	6					90	75			100		0
	Ent1	29				69	64	88			100		0
	Ent2	17					12	78	66		91		0

As the Boolean threshold functions are examples of *symmetric* Boolean functions, the exploitation of symmetry may suggest that the effect of the presence of pathogens is judged equally. However, this is not the case, as both the probability of presence of a pathogen and the susceptibility of a pathogen to specific antibiotics determines to what extent a pathogen-treatment combination contributes to the overall effect. For example, the presence of *P. aeruginosa* has more effect on the type of antibiotics to be prescribed than the presence of *S. pneumoniae*.

Table 8.9: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SIII,** $k = 1$**)** (noisy-OR gate) model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		96					97		100		0
1	SA	25					94			72	100	72	0
	HI	8	94	99		99					100		0
	SP	3	71	97	65	97					100		0
2		33						96			100		0
1	PA	19				85		91			88		0
	AC	6					92	81			100		0
	Ent1	29				88	89	96			100		0
	Ent2	17					48	98	90		96		0

Table 8.10: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SIII,** $k = 2$**)** model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		67					72		85		0
1	SA	25					42			25	67	28	0
	HI	8	31	34		41					50		0
	SP	3	27	47	0	43					67		0
2		33						67			79		0
1	PA	19				22		26			28		0
	AC	6					13	8			17		0
	Ent1	29				22	10	28			31		0
	Ent2	17					23	40	37		46		0

Table 8.11: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SIII,** $k = 3$**)** model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		32					42		54		0
1	SA	25					19			0	31	0	0
	HI	8	1	31		23					38		0
	SP	3	0	27	0	23					63		0
2		33						17			29		0
1	PA	19				6		11			13		0
	AC	6					0	0			0		0
	Ent1	29				10	6	14			16		0
	Ent2	17					16	24	22		28		0

Table 8.12: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SIII, $k = 4$)** model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		3					25		31		0
1	SA	25					4			0	12	0	0
	HI	8	0	5		11					25		0
	SP	3	0	0	0	7					33		0
2		33						7			8		0
1	PA	19				0		0			0		0
	AC	6					0	0			0		0
	Ent1	29				3	3	8			9		0
	Ent2	17					6	19	14		22		0

Table 8.13: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SIII, $k = 5$)** model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		0					2		7		0
1	SA	25					0			0	4	0	0
	HI	8	0	0		0					11		0
	SP	3	0	0	0	0					0		0
2		33						0			0		0
1	PA	19				0		0			0		0
	AC	6					0	0			0		0
	Ent1	29				0	0	1			3		0
	Ent2	17					0	2	2		5		0

When using *susceptibility I model* (i.e., prescribing antimicrobial therapy results in coverage of pathogens colonising as well as pathogens not colonising a patient), the model always gives high coverage. It is counter-intuitive that even threshold functions with high k give high coverage, as patients are usually colonised with at most two pathogens. In addition, this probabilistic model does not support obtaining insight into the actual effects of the antibiotics on the pathogens that cause the infection. This is the information a clinician would like to obtain from a probabilistic model.

Table 8.14: Predicting optimal treatment for 153 patients diagnosed with VAP using the **(SIII,** $6 \leq k \leq 7$**)** model, distinguished by different number of colonising pathogens (Nr.). Mean coverage is specified for a selection of antibiotics.

			Antibiotic coverage (See Table 8.2)										
			A	B	C	D	E	F	G	H	I	J	K
Nr.	Patho	N											
2		13		0					0		0		0
1	SA	25					0			0	0	0	0
	HI	8	0	0		0					0		0
	SP	3	0	0	0	0					0		0
2		33						0			0		0
1	PA	19				0		0			0		0
	AC	6					0	0			0		0
	Ent1	29				0	0	0			0		0
	Ent2	17					0	0	0		0		0

The results for *susceptibility II model* (i.e., when there is no colonisation no medication should be prescribed), however, indicate that when a patient is colonised with, for example, 2 pathogens, the model advises to prescribe no antimicrobial therapy. This is due to dominance of the remaining five pathogens by which the patient has not been colonised. This is clearly undesirable for a life-threatening disease like VAP. This model might, therefore, be used for patients with a low likelihood of having VAP.

Incorporation of *susceptibility III model* (i.e., when there is colonisation, cover it with antibiotics) yields a Bayesian network that performs best at prescribing antimicrobial therapy for monomicrobial as well polymicrobial VAP. It appeared that a threshold function τ_k with $k = 1$ and $k = 2$ yielded the best results, according to the reference standard. Using a model that is able to combine and compare covering results of both $k = 1$ and $k = 2$ would be worth considering. As can be learnt from these tables, coverage probabilities for $k = 1$ are high for monomicrobial as well as for polymicrobial VAP, whereas for $k = 2$ coverage probabilities for monomicrobial infections were relatively low, compared to those for polymicrobial VAP episodes. Therefore, a combination model should be used as follows:

```
IF   results for k = 2 indicate relatively high coverage
     probabilities for polymicrobial VAP
THEN use k = 2 to calculate treatment for
     polymicrobial VAP
ELSE use k = 1 to calculate treatment for
     monomicrobial VAP
FI
```

Clearly, susceptibility model III is able to distinguish between monomicrobial and polymicrobial VAP, which is important for the selection of appropriate therapy. In addition, early-onset VAP requires other, often narrower-spectrum antibiotics compared to late-onset VAP. These two findings that are important to limit creation of antibiotic resistance have certain implications for the construction of a clinical Bayesian network model for assisting in prescribing antimicrobial therapy.

Naturally, susceptibility model II has other implications to the selection of antimicrobial therapy, as compared to susceptibility model I, as it accounts for the pathogens that are absent. Therefore, the specificity of the model predictions for selecting antimicrobial treatment will be high for model II, whereas for model I the sensitivity will be high.

In general, these results also provide evidence that the noisy-OR and noisy-AND, which are very popular in Bayesian network modelling, might not always be the best functions to model interactions among random variables.

Although in this paper, we have studied the use of susceptibility models in combination with the use of Boolean threshold functions for treatment selection in VAP, it is likely that the techniques introduced in this paper are also relevant to other clinical fields. In particular in clinical areas where it is relevant to consider the *number* of causes of disease in selecting treatment of the disease, the methods can be of use.

To conclude, it was shown that the noisy-threshold model is useful from a practical point of view by using it as a basis for the refinement of an existing real-world Bayesian network for the management of critically ill patients.

[9]

Summary and General Discussion

In this chapter, the achievements of the work described in this thesis are summarised. In addition, possible alternative approaches to the development of probabilistic models for the management of VAP as well as suggestions for further research are presented.

9.1 Diagnosis of VAP

In the first part of the thesis, we aimed to improve our knowledge about the process of respiratory tract colonisation in critically ill and mechanically ventilated patients, irrespective whether they develop VAP.

The nature of the critical illness of ICU patients implies several specific aspects of patient care. For instance, chest X-rays are – for multiple reasons – performed almost daily; samples obtained from different body sites (such as sputum, wounds, urine, blood) are frequently submitted for microbiological analysis (i.e. culture) to monitor or detect sources of infection; biochemical analyses of blood samples are obtained on a daily basis to closely monitor infection parameters. Patients depending on ventilatory support are most vulnerable to infection of the lungs, as the endotracheal tube eliminates an important natural barrier function (swallowing) providing bacteria colonising the oropharyngeal cavity almost free access to the distal lung tissues. The bacterial etiology of endotracheal colonisation depends on the duration of risk exposure, and this can be expressed as the duration of stay in ICU or the duration of intubation. During the first 6 days, patients are primarily at risk to become colonised with so-called early-onset pathogens, such as *Staphylococcus aureus*, *Haemophilus influenzae* and *Streptococcus pneumoniae*, whereas late-onset pathogens, such as *Pseudomonas aeruginosa*, *Acinetobacter* species and Enterobacteriaceae most frequently colonise patients after 6 days in the ICU, as reported in Chapter 2. Of

Part of this work is published in Studies in Health Care Technology and Informatics, 2004

note, the early-onset pathogens represent the normal flora of the upper respiratory tract and these bacteria will also be detected frequently in healthy persons. The late-onset pathogens, though, are rarely detected in the respiratory tract of healthy subjects, and these pathogens, therefore, represent the typical "hospital flora".

Using a prospectively collected dataset of more than 2000 mechanically ventilated ICU patients, with microbiological culture results of respiratory tract samples available for 70% of them, we investigated the persistence of respiratory tract colonisation. This persistence (defined as the duration that a patient remained colonised) appeared to be highly pathogen-dependent. Median durations of colonisation ranged from 4–5 days for early-onset pathogens and from 5–8 days for late-onset pathogens. Since administration of antibiotics is common practice in ICU patients, we further investigated the effects of antibiotics on acquisition and persistence of bacterial respiratory tract colonisation. Antibiotics reduced the relative risk of acquiring endotracheal colonisation with early-onset pathogens and Enterobacteriaceae, under the condition that these bacteria were – *in-vitro* – susceptible to these antibiotics. In addition, when patients who were colonised with these bacteria (early-onset pathogens and Enterobacteriaceae) received antibiotics to which these pathogens were susceptible, the relative risk for disappearance of colonisation was significantly higher (as compared to late-onset pathogens). Clearly, late-onset pathogens such as *P. aeruginosa* are difficult to eradicate and persist long, even when antibiotics with – *in-vitro* – activity are administered intravenously. Optimally, we would have analysed pathogen persistence in the absence of antibiotic treatment. Such an analysis would have allowed elucidation of the role of time in colonisation and persistence. Unfortunately, the number of patients being colonised without having received antibiotics was too low for such an analysis.

Only a proportion of colonised patients will develop VAP, usually after a gradual change from asymptomatic colonisation to clinical infection. Unfortunately, there is no accurate and easily available gold standard for VAP, which seriously hampers diagnostic research. As an alternative, we have used a reference standard that was based on a combination of diagnostic tests, summarised in a diagnostic tree, as described in Chapter 3. In the study cohort, ICU physicians prescribed antimicrobial treatment for 238 episodes of presumed VAP. By thorough adjudication of all parameters at the time of presumed VAP and the clinical response in the days thereafter by two infectious disease experts, 157 episodes were eventually considered VAP.

In case of VAP, diagnosing involves the use of parameters that lack absolute certainty. We argue in this thesis that the paradigm of Bayesian networks is suited to deal with these uncertain diagnostic processes. For a Bayesian network, addressing a disease process as complex as VAP, the structure of the model may be given by the domain expert. The structure denotes the various causalities and independences within the domain of concern. At the same time it holds that the more vertices are introduced in a model, the more complex it becomes. Hence, when it comes to estimating all the conditional

probabilities by the expert (on the basis of experience, or gained from literature), employing a parsimonious structure still holding all the independences that are important in the domain, is warranted. In Chapter 3 we demonstrated that the Bayesian network model with the best performance to distinguish between VAP and non-VAP, was a model with a simple structure, and that an extended naive Bayesian classifier outperformed all the other structures studied. The conclusion that a simple Bayesian-network structure often performs best has also been reached in many other studies, also outside the field of medicine.

As we did not want to rely on expert opinions only, we also used the longitudinal database of mechanically ventilated ICU patients to update the – so-called – subjective expert estimates. In the process of constructing a clinical decision-support system, physicians should never be ruled out, as their experience and domain knowledge are as important (or even more important) as prospectively collected data. To optimise the model, the data set was split: the conditional probabilities were compared and updated where needed in one half of the dataset and tested on the remainder to prevent overfitting. By using this procedure, the performance of the Bayesian network improved and had good test characteristics. AUC was 0.857 [0.827–0.888 95% CI], sensitivity and specificity were 80%, and positive and negative predictive values were 6.1% and 99.6%, respectively. The low positive predictive value partly results from the extremely low daily prevalence of VAP. When the model predicts that the posterior probability of VAP is above 0.46, a physician should consider the possibility that the patient is suffering from VAP. With posterior probabilities below 0.46, the likelihood of VAP is extremely low.

9.2 Dynamics of Diagnostic Parameters of VAP

In the second part of the thesis we aimed to elucidate the temporal evolution of VAP. For this purpose, standard descriptive statistical analysis was complemented by learning temporal Bayesian networks from data of patients with and without VAP. The concept of context-specific dependence was the main tool used.

The Bayesian network model, described in Chapter 3, provides daily predictions of the likelihood of VAP, without accounting for any possible temporal effects. We, therefore, in Chapter 5 investigated whether possible temporal trends in diagnostic values such as body temperature, sputum production, leukocytes and pO_2/FiO_2 ratio could be detected. For this, non-VAP patients were matched to VAP patients on three criteria: duration of mechanical ventilation, ICU ward (medical or neurosurgical) and gender. In a collaborative project with Charitos and Van der Gaag [112] we developed a repetitive temporal Bayesian network for diagnosing VAP. Here, a fixed transition was assumed between two time slices, representing two consecutive days, with fixed structure, for calculating the likelihood of VAP on a certain day while taking that likelihood on the previous day into account. This approach yielded good test characteristics, measured in AUC and Brier score [112]. In this thesis, however,

we used an alternative approach. We investigated whether non-repetitive temporal Bayesian networks can be constructed as a tool for exploring the temporal evolution of the disease processes in mechanically ventilated ICU patients. We found that associations between variables were alternately present and absent at different time points, which supported our hypothesis of non-repetitiveness. The constructed temporal models using the Necessary Path Condition (NPC) structure learning algorithm revealed different temporal as well as atemporal patterns for VAP and non-VAP patients. Thereby, the independences found for VAP patients were much stronger, i.e., had higher significance levels than independences for non-VAP patients. From these observations, we concluded that the NPC algorithm is useful, as it allowed for the incorporation of background knowledge. The algorithm combines the virtues of offering the capability of automatic learning of independence information from data, whereas uncertainty regarding both the presence of dependences and the directionality of arcs can be resolved by the user. In addition, these results show that 'normal' repetitive dynamic Bayesian networks may not suffice in capturing the characteristics of a domain.

9.3 Antibiotic Treatment Selection for VAP

In the third part of the thesis, we investigated the models' performance to predict the microbial cause of VAP. When VAP is diagnosed, targeted antimicrobial treatment should be started as soon as possible. Culture results from endotracheal aspirates, though, will not be immediately available. Actually, definite results may only be available after 2 days. Therefore, most frequently, empirical treatment must be initiated. In most cases the uncertainty about microbial cause and its antibiotic susceptibility necessitates a rather broad antimicrobial coverage.

The Bayesian network model, however, uses a different approach to select antibiotic treatment for VAP. As described in Chapter 6, durations of hospital stay as well as duration of mechanical ventilation are used as predictors for the presence of certain pathogens. In addition, results of previously performed endotracheal cultures and prior antibiotics (linked to the *in-vitro* susceptibility of pathogens) were used as parameters. Based on these parameters, posterior colonisation probabilities were calculated and model performance was analysed, for each group of pathogens, by ROC analysis. With this analysis, the optimal cut-off point, i.e., the point on the curve that yielded the highest combination of sensitivity and specificity, was determined for each pathogen. For *P. aeruginosa* (PA), for example, the optimal cut-off value was 0.289, which was interpreted as follows: when the posterior probability of being colonised by PA is above 0.289, it is likely that this pathogen is colonising the patient's respiratory tract; with lower estimates, absence of this pathogen can be assumed. This procedure was followed for all seven pathogen groups and the Bayesian network model had good test characteristics for predicting VAP-causing pathogens. However, the disadvantage of dichotomising posterior colonisation

probabilities is that – according to model predictions – 20 of the 153 episodes of VAP were caused by more than 3 pathogens, which is unlikely and which resulted in unnecessary prescription of broad-spectrum antibiotics.

In Chapters 7 and 8, a variety of modelling methods that affect the selection of antimicrobial treatment for VAP are described. In addition, different ways of modelling the process from endotracheal colonisation to covering these colonising pathogens by antibiotics were investigated. As mentioned, antibiotic treatment should be started as soon as possible when VAP is diagnosed. The starting point then is the colonisation status of the patient. Usually, the physician will check results of previously performed cultures and, if available, will initiate antimicrobial treatment based on these data. Naturally, one would expect that when a patient was previously colonised by *P. aeruginosa*, it is likely that the patient is still colonised with this pathogen on the day VAP is diagnosed, as has been shown in Chapter 2. However, the probability that a pathogen is still colonising a patient, given that colonisation was demonstrated previously, is much higher for *P. aeruginosa* than for *S. pneumoniae* and, therefore, the colonisation status of a patient may be uncertain. Thus, data on previous colonisation might be useful, but do not guarantee a patients' colonisation status. In addition, cultures may not detect every pathogen, for example when growth is below a detection limit. In all these situations, the Bayesian network model may provide more reliable information on colonisation status, based on durations of hospital stay and mechanical ventilation.

In Chapter 7 it was shown that applying the closed-world assumption (CWA), i.e., the assumption that unknown test results are negative, proved to be useful in interpreting previously performed endotracheal cultures. A disadvantage is that the CWA does not allow for the hypothetical situation that a pathogen – other than the one(s) already found present or absent according to previous culture results – additionally colonises a patient. Subsequently, the possibility of losing colonisation is not accounted for either.

In Chapter 8, three different ways of modelling antibiotic coverage were described, called 'susceptibility models':

I antibiotic coverage of both identified (by means of microbiological cultures) and non-identified pathogens is considered appropriate;

II antibiotic coverage of identified pathogens is considered appropriate, whereas antibiotic coverage of non-identified pathogens is not;

III antibiotic coverage of identified pathogens is considered appropriate only.

The probability distribution of the vertex that represents the overall susceptibility or coverage is expressed as the conjunctive effect of the seven different pathogen groups, i.e., the noisy-AND causal independence model was used for this purpose. The disadvantage of using the noisy-AND is that the model tries to cover all seven pathogens, whereas a patient is hardly ever colonised by more than three pathogens. This resulted in prescribing unnecessarily broad-spectrum antibiotics. Therefore, Boolean threshold functions were studied as

generalisations of the Boolean AND, i.e., 100% coverage for all pathogens results in 100% overall coverage, and Boolean OR, i.e., 100% coverage for one or more pathogens results in 100% overall coverage; in principle it is possible by the threshold functions to restrict coverage to less than seven pathogens. We conclude that susceptibility model III in combination with τ_k with $k = 1, 2$ performs best in covering present causative pathogens in monobacterial or polymicrobial episodes of VAP.

9.4 Future research

Diagnosis and treatment of VAP is just one example of complex patient management. Some of its facets were investigated in this thesis, but many questions have not, or only insufficiently, been dealt with. Bayesian networks were selected as the formalism of choice, as they are one of few statistical formalisms that allow for the easy integration of knowledge from experts and knowledge derived from data. By using this formalism for the actual construction of a model that could be used for diagnosis and treatment selection, we came across fundamental problems that needed to be solved, some of which have already been discussed in this thesis. However, many more research questions remain and these are discussed in the following.

9.4.1 Modelling colonisation dynamics

Having described and quantified dynamics of endotracheal colonisation in the ICU, a logical next step would be to construct a dynamic Bayesian network (DBN) modelling these colonisation patterns. Such a DBN could predict presence and absence of the most frequently occurring pathogens colonising ICU patients, based on hospital stay and duration of mechanical ventilation.

9.4.2 Temporal Bayesian networks

Temporal Bayesian networks (TBNs) are increasingly used, as there are many problems that involve dimensions of time. TBNs are suitable tools for the construction of compact probabilistic models that include time. Almost without exception, these TBNs are repetitive Bayesian networks with transition probability distributions that are time-invariant. These TBNs are also called dynamic Bayesian networks, as they can be generated from a specification consisting of two time slices, which offers an even more compact specification. Although dynamic Bayesian networks offer very powerful representations for various forms of decision making, dynamic Bayesian networks are not suitable for determining subtle changes in dependence and inadvertence information in time. For this, non-repetitive TBNs networks are more suitable and inclusion of context-specific dependence would make the formalism even more powerful. At the moment, there are no tools available that allow exploring of the context-specific independence in time, and we thus had to resort to the use of

standard Bayesian networks tools. Extending existing Bayesian-network software packages by including facilities for non-repetitive TBNs seems an appropriate follow-up to the research described in this thesis.

9.4.3 Modelling interactions

When modelling interactions between variables that affect a particular outcome, probabilistic models rapidly grow unwieldy, as the number of probabilities that must be assessed in a discrete conditional probability distribution is exponential in the number of variables associated to the parents of a variable. Linear approximation methods, such as logistic regression, are, therefore, frequently used. This problem also occurred in our studies, and it was tackled by deploying the concept of causal independence, one of the many types of canonical probabilistic models, where probability distributions are generated based on a fixed number of parameters. We used Boolean functions, in particular Boolean threshold functions. However, as there are 2^{n^n} different Boolean functions, and 2^{n+1} symmetric Boolean functions, with n the number of Boolean variables involved, there are many other functions that could have been investigated. These could offer a more generalised way of modelling interactions – if required – than the noisy-AND does.

9.4.4 Utility modelling and treatment selection

A lot of attention has been paid (in this thesis) on structured probability distributions, with less attention for modelling of utility information needed to select appropriate antimicrobial treatment. Utility assessment in antimicrobial treatment selection, therefore, is an obvious topic for further research.

9.4.5 Role of decision support in the management of VAP

As VAP – until now – is considered to be associated with attributable mortality, some form of decision support in its management might be beneficial. Despite the apparent problems in diagnosing VAP we succeeded to construct a diagnostic part of the Bayesian network that is able to distinguish between VAP and non-VAP patients. Still, the positive predictive value is low, but the extremely high negative predictive value could assist physicians in reducing unnecessary antibiotic prescriptions. Yet, the impact and practicability of the constructed Bayesian decision-support system (BDSS) in real life is to be determined and evaluated in future.

9.4.6 Clinical assessment of decision-support systems

The Bayesian-network model described in this thesis has already been incorporated in a computational infrastructure, involving a Linux server, which allows to use the model on an Internet browser via Ethernet. This set-up would have allowed a clinical assessment of the system. However, much more time was

needed for the refinement of the model than originally planned, as management of VAP in patients appeared more difficult than expected. Unfortunately, this course of events in the development of clinical decision-support systems is rather standard, and few clinical decision support systems have undergone clinical evaluation. Clearly, making a decision-support system available on the Internet ensures that people can use the system, and could be a first step in its clinical assessment. More importantly, though, is that the system has been appropriately evaluated in a clinical setting.

Before incorporating a BDSS into clinical practice, the effects of such a system on patient management and outcome and the willingness of physicians to use such a form of decision-support should be determined.

To evaluate the performance of the BDSS in diagnosing and treatment selection we propose a randomised study. The domain of the study would be all mechanically ventilated patients with a clinical suspicion of VAP, as judged by the responsible physician. The determinant of the study will be the advice provided by the BDSS. Accepting the absence of a suitable gold standard for VAP, several clinical endpoints could be chosen, such as patient outcome (e.g., survival, length of stay in ICU or hospital) or antibiotic use (e.g., total antibiotic use or proportion of patients receiving appropriate empirical treatment).

For every patient meeting the eligibility criteria (i.e., being mechanically ventilated and having some clinical criteria compatible with VAP) the physician is asked to judge the presence of VAP (either a yes/ no or a proportion of likelihood ranging from 0% to 100%) as well as to select the most appropriate antimicrobial therapy. Subsequently, the computer presents the patients' clinical details (See Figure 9.1), directly derived from the clinical database. If needed some of these criteria can be manually adapted (or added) by the physician. After the completing this form, the BDSS provides a management advice (likelihood of VAP and antibiotic proposal) that is randomly presented to the physician. For those 50% of the consultations in which the advice is offered, the physician is asked – with the information provided by the BDSS – whether he or she remains with the initial opinion and antibiotic choice. This design would allow us to quantify the effects of BDSS use in patient management.

Evaluation biases

When evaluating the BDSS in the ICU setting, results may not faithfully reflect the reality due to unknown systematic effects. These confounding effects, or biases, have to be taken into account when designing such a study. Known types of biases include [113, 114]:

- The volunteer effect: physicians that volunteer to use the system may perform better than others. This could be prevented by including all physicians.
- The assessment bias: knowledge of the gold standard (or reference standard) result may influence the users' decisions. Since there is no real gold standard for VAP, this does not apply here.

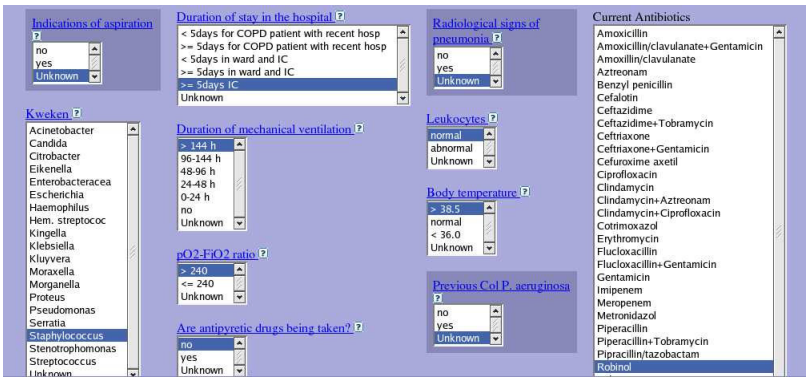


Figure 9.1: Screen-shot of the BDSS for VAP.

- The Hawthorne effect: clinical performance improves if clinicians know they are being observed or studied. This is an almost non-preventable effect. By having all physicians using the first part of the intervention (entering the data in the model), both study groups (those that do and do not receive feedback from the BDSS) will experience some form of 'being watched'.
- The checklist effect: performance may improve if clinicians use a checklist, as the physician is reminded of the factors that are considered relevant in, for example, diagnosing VAP. This may cause improvement in clinical performance.

Resulting BDSS for VAP

In this paragraph we describe the technical aspects of performing the proposed evaluation study in the ICUs of the UMC Utrecht. First, the system should be integrated in the clinical information system used in the ICUs. This system is a full-fletched system that functions as an electronic patient records system (EPR), containing a wealth of information that can be exploited for decision-support purposes.

The front-end of the BDSS consists of a graphical user interface (GUI) with pop-up menus, tables and some graphics, allowing clinicians to enter patient data and to inspect patient data, mostly in textual form, but sometimes – for certain laboratory data – in graphical form, as a time plot. The back-end is a relational database management system, which is linked to the EPR system's relational database system [115]. This not only offers modern facility for secure data storage, updating and retrieval, but also a Structured Query Language (SQL) interface, allowing external systems to have access to the data.

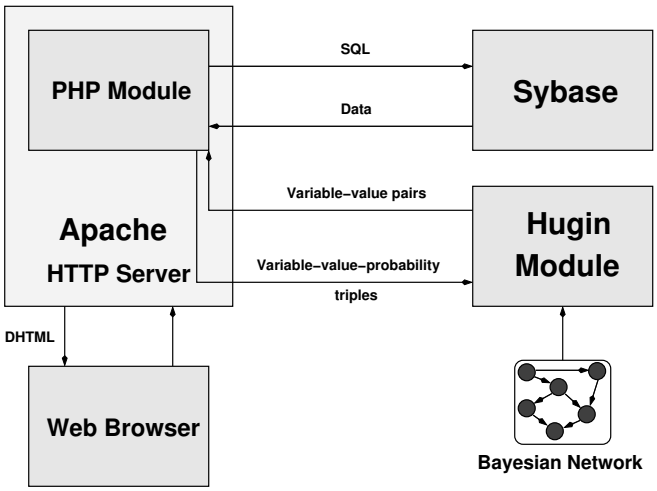


Figure 9.2: Architecture of the BDSS for VAP.

It was decided to develop a separate user-interface for the decision-support system in particular because this makes separate development of both underlying Bayesian network model and user-interface possible. We used modern distributed information system technology, in our case Hypertext Preprocessor (PHP). PHP is a server-side HTML-embedded cross-platform scripting language, which allows one to generate content of HTML pages dynamically [116].

PHP is actually used as the language to link various parts of the system together. The majority of the functionality of the system is offered by a commercially available Bayesian-network and decision-network package, which is used in our project to implement the inference engine for the decision-support system. This part of the system processes patient data, and offers various types of advice based on the results computed after instantiating the Bayesian network model of VAP with the patient data. This advice is presented to the user.

The final components for the system are an HTTP server, and a Web browser. The HTTP server that is used in the project is Apache. The Web browser acts as the user interface to the decision-support system. The architecture of the system is visualised in Figure 9.2.

References

[1] Grimes DA and Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet*, 365:1500–1505, 2005.

[2] Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, and Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard: a review of methods. *Health Technology Assessment*, 11:iv–51, 2007.

[3] Rijnders BJ, Peetermans WE, Verwaest C, Wilmer A, and Van Wijngaerden E. Watchful waiting versus immediate catheter removal in icu patients with suspected catheter-related infection: a randomized trial. *Intensive Care Medicine*, 30:1073–1080, 2004.

[4] Burbank F. A computer diagnostic system for the diagnosis of prolonged undifferentiated liver disease. *American Journal of Medicine*, 46:401–415, 1969.

[5] de Dombal FT, Leaper FJ, Staniland JR, McAnn AP, and Horrocks JC. Computer-aided diagnosis of acute abdominal pain. *British Medical Journal*, ii:9–13, 1972.

[6] Gorry GA and Barnett GO. Experience with a model of sequential diagnosis. *Computers and Biomedical Research*, 1:490–507, 1968.

[7] Knill-Jones RP, Stern RB, Girmes DH, Maxwell JD, Thompson RPH, and Williams R. Use of a sequential bayesian model in the diagnosis of jaundice. *British Medical Journal*, i:530–533, 1973.

[8] de Dombal FT, Dallos V, and McAdam WA. Can computer-aided teaching packages improve clinical care in patients with acute abdominal pain? *British Medical Journal*, 302:1495–1497, 1991.

[9] Lucas PJF and van der Gaag LC. *Principles of Expert Systems*. Wokingham: Addison-Wesley, 1991.

[10] Oniśko A, Lucas PJF, and Druzdzel MJ. Comparison of rule-based and bayesian network approaches in medical diagnostic systems. In S. Quaglini, P. Barahona, and S. Andreassen, editors, *LNAI2101, Proceedings of the Artificial Intelligence in Medicine Europe (AIME) conference*, pages 283–292. Springer-Verlag, Berlin, 2001.

[11] Shortliffe EH. *Computer-based Medical Consultations: MYCIN*. Elsevier, New York, 1976.

[12] Shortliffe EH and Buchanan BG. A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23:351–379, 1975.

[13] Lucas PJF. Certainty-like structures in Bayesian belief networks. *the Knowledge-based Systems journal*, 14:327–335, 2001.

[14] Bayes T (1702-1761). An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a letter to john canton, a. m. f. r. s. *Philosophical Transactions, Giving Some Account of the Present Undertakings, Studies and Labours of the Ingenious in Many Considerable Parts of the World*, 53:370–418, 1763. Online: `http://www.stat.ucla.edu/history/essay.pdf`.

[15] van Rijsinge WP, van der Gaag LC, Visseren FL, and van der Graaf Y. Compliance with the hyperlipidaemia consensus: Clinicians versus the computer. In M. Dojat, E. Keravnou, and P. Barahona, editors, *Lecture Notes in Computer Science, Proceedings of the 9^{th} conference on Artificial Intelligence in Medicine in Europe*, pages 340–344. Springer Verlag, 2003.

[16] Witten IA and Frank E. *Data mining*. Morgan Kaufmann, San Francisco, CA, USA, 2000.

[17] Mitchell TM. *Machine learning*. McGraw-Hill, 1997.

[18] Husmeier D, Dybowski R, and Robert S. *Probabilistic Modeling in Bioinformatics and Medical Informatics*. Springer-Verlag, London, 2005.

[19] Leavis HL. *Evolutionary insights into CC17 Enterococcus faecium*. PhD thesis, University of Utrecht, 2007.

[20] Bellazzi R, Larizza Chr, Magni P, and Bellazi R. Temporal data mining for the quality assessment of hemodialysis services. *Artificial Intelligence in Medicine*, 34(1):25–39, 2005.

[21] Lucas PJF, van der Gaag LC, and Abu-Hanna A. Bayesian networks in biomedicine and health-care. *Artificial Intelligence in Medicine*, 30(3):201–214, 2004.

[22] Friedman N, Murphy K, and Russell S. Learning the structure of dynamic probabilistic networks. In *Proceedings of the 14th UAI*, pages 139–147, 1998.

[23] Donders AR, van der Heijden GJ, Stijnen T, and Moons KG. Review: a gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, 59:1087–1091, 2006.

[24] Dempster AP, Laird NM, and Rubin DB. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[25] Knaus WA, Draper EA, Wagner DP, and Zimmerman JE. Apache ii: a severity of disease classification system. *Critical Care Medicine*, 13:818–829, 1985.

[26] Schurink CA, Van Nieuwenhoven CA, Jacobs JA, Rozenberg-Arska M, Joore HC, Buskens E, Hoepelman AI, and Bonten MJ. Clinical pulmonary infection score (cpis) for ventilator-associated pneumonia: accuracy and inter-observer variability. *Intensive Care Medicine*, 30:217–224, 2004.

[27] Bonten MJM, Kollef MH, and Hall JB. Risk factors for ventilator-associated pneumonia: from epidemiology to patient management. *Clinical Infectious Diseases*, 38(8), 2004.

[28] Schurink CAM. *Ventilator-associated pneumonia: a diagnostic challenge*. PhD thesis, University of Utrecht, 2004.

[29] Bonten MJM and Bergmans CJJ. Risk factors for pneumonia, and colonization of respiratory tract and stomach in mechanically ventilated icu patients. *American Journal of Respiratory and Critical Care Medicine*, 154:1339–1346, 1996.

[30] Bergmans CJJ and Bonten MJM. Indications for antibiotic use in icu patients: a one-year prospective surveillance. *Journal of Antimicrobial Chemotherapy*, 39:527–535, 1997.

[31] Bigger JW. Treatment of staphylococcal infections with penicillin. *Lancet*, ii:497–500, 1944.

[32] Kussel E, Kishony R, Balaban NQ, and Leibler S. Bacterial persistence: a model of survival in changing environments. *Genetics*, 169:1807–1814, 2005.

[33] Keren I, Kaldalu N, Spoering A, Wang Y, and Lewis K. Persister cells and tolerance to antimicrobials. *FEMS Microbiology Letters*, 230:13–18, 2004.

[34] Schurink CA, Visscher S, Lucas PJ, van Leeuwen HJ, Buskens E, Hoff RG, Hoepelman A, and Bonten MJ. A Bayesian decision-support system for diagnosing ventilator-associated pneumonia. *Intensive Care Medicine*, 33(8):1379–1386, 2007.

[35] Murray TS, Egan M, and Kazmierczak BI. Pseudomonas aeruginosa chronic colonization in cystic fibrosis patients. *Current Opinion in Pediatrics*, 19:83–88, 2007.

[36] Bergmans DC, Bonten MJ, De Leeuw PW, and Stobberingh EE. Reproducibility of quantitative cultures of endotracheal aspirates from mechanically ventilated patients. *Journal of Clinical Microbiology*, 35:796–798, 1997.

[37] SPSS Inc., Chicago USA, `http://www.spss.com`.

[38] Gnuplot, interactive data and function plotting utility. `http://www.gnuplot.info`.

[39] Trouillet JL, Chastre J, Vuagnat A, Joly-Guillou ML, Combaux D, Dombret MC, and Gibert C. Ventilator-associated pneumonia caused by potentially drug-resistant bacteria. *American Journal of Respiratory and Critical Care Medicine*, 157:531–539, 1998.

[40] Verhamme KM, De Coster W, De Roo L, De Beenhouwer H, Nollet G, Verbeke J, Demeyer I, and Jordens P. Pathogens in early-onset and late-onset intensive care unit-acquired pneumonia. *Infection Control and Hospital Epidemiology*, 28:389–397, 2007.

[41] Sirvent JM, Torres A, El-Ebiary M, Castro P, de Batlle J, and Bonet A. Protective effect of intravenously administered cefuroxime against nosocomial pneumonia in patients with structural coma. *American Journal of Respiratory and Critical Care Medicine*, 155:1729–1734, 1997.

[42] Ewig S, Torres A, El-Ebiary M, Fegas N, Hernez C, Gonzz J, NicolJM, and Soto L. Bacterial colonization patterns in mechanically ventilated patients with traumatic and medical head injury. *American Journal of Respiratory and Critical Care Medicine*, 159:188–198, 1999.

[43] Chastre J and Fagon JY. Ventilator-associated pneumonia. *American Journal of Respiratory and Critical Care Medicine*, 165:867–903, 2002.

[44] CDC. CDC definitions for nosocomial infections. *American Review of Respiratory Disease*, 139:1058–1059, 1988.

[45] Pugin J, Auckenthaler R, Mili N, Janssens JP, Lew PD, and Suter PM. Diagnosis of ventilator-associated pneumonia by bacteriologic analysis of bronchoscopic and nonbronchoscopic "blind" bronchoalveolar lavage fluid. *merican Review of Respiratory Disease*, 143:1121–1129, 1991.

[46] Fagon JY, Chastre J, Wolff M, Gervais C, Parer-Aubas S, Stephan F, Similowski T, Mercat A, Diehl JL, Sollet JP, and Tenaillon A. Invasive and noninvasive strategies for management of suspected ventilator-associated pneumonia. a randomized trial. *Annals of Internal Medicine*, 132:621–630, 2000.

[47] Sanchez Garcia M, Cambronero Galache JA, Lopez Diaz J, Cerda Cerda E, Rubio Blasco J, Gomez Aguinaga MA, Nunez Reiz A, Rogero Marin S, Onoro Canaveral JJ, and Sacristan del Castillo JA. Effectiveness and cost of selective decontamination of the digestive tract in critically ill intubated patients. a randomized, double-blind, placebo-controlled, multicenter trial. *American Journal of Respiratory and Critical Care Medicine*, 158:908–916, 1998.

[48] Ruiz M, Torres A, Ewig S, Marcos MA, Alcon A, Lledo R, Asenjo MA, and Maldonaldo A. Noninvasive versus invasive microbial investigation in ventilator-associated pneumonia: evaluation of outcome. *American Journal of Respiratory and Critical Care Medicine*, 162:119–125, 2000.

[49] Schurink CAM, Lucas PJF, Hoepelman IM, and Bonten MJM. Computer-assisted decision support for the diagnosis and treatment of infectious diseases in intensive care units. *Lancet Infectious Diseases*, 5:305–312, 2005.

[50] Lucas PJF, de Bruijn NC, Schurink K, and Hoepelman A. A probabilistic and decision-theoretic approach to the management of infectious disease at the icu. *Artificial Intelligence in Medicine*, 19(3):251–279, 2000.

[51] Friedman N, Geiger D, and Goldszmidt M. Bayesian network classifiers. *Machine Learning*, 29:131–163, 1997.

[52] Stockley RA, Bayley D, Hill SL, Hill AT, Crooks S, and Campbell EJ. Assessment of airway neutrophils by sputum colour: correlation with airways inflammation. *Thorax*, 56:366–372, 2001.

[53] Lucas PJF. Restricted bayesian-network structure learning. In G.A. Gamez, S. Moral, and A. Salmerón, editors, *Advances in Bayesian Networks, Studies in Fuzziness and Soft Computing*, volume 146, pages 217–232, Berlin, 2004. Springer-Verlag.

[54] Pingleton SK, Fagon JY, and Leeper KV Jr. Patient selection for clinical investigation of ventilator-associated pneumonia. criteria for evaluating diagnostic techniques. *Chest*, 102:553S–556S, 1992.

[55] Metz CE. Basic principles of roc analysis. *Seminars in Nuclear Medicine*, 8, 1978.

[56] Sox HC, Blatt A, Higgins MC, and Marton KI. *Medical Decision Making*. Butterworth-Heinemann, Woburn, 1998.

[57] Bonten MJM, Froon AH, Gaillard CA, Greve JW, de Leeuw PW, Drent M, Stobberingh EE, and Buurman WA. The systemic inflammatory response in the development of ventilator-associated pneumonia. *American Journal of Respiratory and Critical Care Medicine*, 156:1105–1113, 1997.

[58] Froon AH, Bonten MJM, Gaillard CA, Greve JW, Dentener MA, de Leeuw PW, Drent M, Stobberingh EE, and Buurman WA. Prediction of clinical severity and outcome of ventilator-associated pneumonia. comparison of simplified acute physiology score with systemic inflammatory mediators. *American Journal of Respiratory and Critical Care Medicine*, 158:1026–1031, 1998.

[59] Bonten MJM, Bergmans DC, Ambergen AW, de Leeuw PW, van der Geest S, Stobberingh EE, and Gaillard CA. Risk factors for pneumonia, and colonization of respiratory tract and stomach in mechanically ventilated icu patients. *American Journal of Respiratory and Critical Care Medicine*, 154:1339–1346, 1996.

[60] Bonten MJM, Gaillard CA, de Leeuw PW, and Stobberingh EE. Role of colonization of the upper intestinal tract in the pathogenesis of ventilator-associated pneumonia. *Clinical Infectious Diseases*, 24:309–319, 1997.

[61] Kollef MH. Prolonged use of ventilator circuits and ventilator-associated pneumonia: a model for identifying the optimal clinical practice. *Chest*, 113:267–269, 1998.

[62] Kristensen B, Andreassen S, Leibovici L, Riekehr C, Kjaer AG, and Schonheyder HC. Empirical treatment of bacteraemic urinary tract infection. evaluation of a decision support system. *Danish Medical Bulletin*, 46:349–353, 1999.

[63] Leibovici L, Alpert G, Laor A, Kalter-Leibovici O, and Danon YL. A clinical model for diagnosis of urinary tract infection in young women. *Archives of Internal Medicine*, 149:2048–2050, 1989.

[64] Paul M, Andreassen S, Nielsen AD, Tacconelli E, Almanasreh N, Fraser A, Yahav D, Ram R, and Leibovici L: TREAT study group. Prediction of bacteremia using treat, a computerized decision-support system. *Clinical Infectious Diseases*, 42:1274–1282, 2006.

[65] Aronsky D and Haug PJ. Diagnosing community-acquired pneumonia with a bayesian network. In *Proceedings of the AMIA Symposium*, pages 632–636, 1998.

[66] Aronsky D and Haug PJ. An integrated decision support system for diagnosing and managing patients with community-acquired pneumonia. In *Proceedings of the AMIA Symposium*, pages 197–201, 1999.

[67] Grimes DA and Schulz KF. Uses and abuses of screening tests. *Lancet*, 359:881–884, 1999.

[68] Cook DJ, Walter SD, Cook RJ, Griffith LE, Guyatt GH, Leasa D, Jaeschke RZ, and Brun-Buisson C. Incidence of and risk factors for ventilator-associated pneumonia in critically ill patients. *Annals of Internal Medicine*, 129:433–440, 1998.

[69] Jensen FV and Nielsen ThD. *Bayesian Networks and Decision Graphs, Second Edition*. Springer, New York, 2007.

[70] Flesch I, Lucas PJF, and Visscher S. On the modularisation of independence in dynamic Bayesian networks. In *Proceedings of BNAIC-2006*, pages 133–140, 2006.

[71] Augusto JC. Temporal reasoning for decision support in medicine. *Artificial Intelligence in Medicine*, 33(1):1–24, 2005.

[72] Dagum P and Galper A. Forecasting sleep apnea with dynamic models. In *Proceedings of UAI-1994*, pages 64–71, 1994.

[73] Adlassnig K-P, Combi C, Das AK, Keravnou ET, and Pozzi G. Temporal representation and reasoning in medicine: Research directions and challenges. *Artificial Intelligence in Medicine*, 38(2):101–13, 2006.

[74] Provan G and Clarke JR. Dynamic network construction and updating techniques for the diagnosis of acute abdominal pain. *PAMI*, 15(3):299–307, 1993.

[75] Geiger D and Heckerman D. Knowledge representation and inference in similarity networks and Bayesian multinets. *Artificial Intelligence*, 82:45–74, 1996.

[76] Labatut V, Pastor J, Ruff S, Démonet JF, and Celsis P. Cerebral modeling and dynamic Bayesian networks. *Artificial Intelligence in Medicine*, 39:119–139, 2004.

[77] Tucker A and Liu X. Learning dynamic Bayesian networks from multivariate time series with changing dependencies. In *Proceedings of the 5th International Symposium on Intelligent Data Analysis, Berlin, Germany*, 2003.

[78] Boutilier C, Friedman N, Goldszmidt M, and Koller D. Context-specific independence in Bayesian networks. In *Proceedings of UAI-1996*, pages 64–72, 1996.

[79] Steck H. *Constraint-Based Structural Learning in Bayesian Networks using Finite Data Sets*. PhD thesis, Institut für Informatik der Technischen Universität München, 1999.

[80] Verma TS and Pearl J. Equivalence synthesis of causal models. In *Proceedings of UAI-1990*, pages 220–227. Morgan Kaufmann, 1990.

[81] Madsen AL andLang M, Kjærulff UB, and Jensen J. The hugin tool for learning Bayesian networks. In *Proceedings of 7th ECSQARU*, pages 594–605, 2003.

[82] Wellman MP. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44:257–303, 1990.

[83] Safdar N, Dezfulian C, Collard HR, and Saint S. Clinical and economic consequences of ventilator-associated pneumonia: a systematic review. *Critical Care Medicine*, 33:2184–2193, 2005.

[84] van Nieuwenhoven CA, Buskens E, Bergmans DC, van Tiel FH, Ramsay G, and Bonten MJ. Oral decontamination is cost-saving in the prevention of ventilator-associated pneumonia in intensive care units. *Critical Care Medicine*, 32:126–130, 2004.

[85] Berdal JE, Bjœrnholt J, Blomfeldt A, Smith-Erichsen N, and Bukholm G. Patterns and dynamics of airway colonisation in mechanically-ventilated patients. *Clinical microbiology and infection*, 13:476–480, 2007.

[86] Visscher S, Schurink CA, Melsen WG, Lucas PJ, and Bonten MJ. Effects of systemic antibiotic therapy on bacterial persistence in the respiratory tract of mechanically ventilated patients. *Intensive Care Medicine*, 34:692–699, 2008.

[87] Schwarz G. Estimating the dimension of a model. *Annals of Statistics*, 6:461–466, 1978.

[88] Akaike H. Information theory and the maximum likelihood principle. In *International Symposium on Information Theory*, 1973. Budapest: Akademiai Kiádo.

[89] Bonten MJ, Slaughter S, Ambergen AW, Hayden MK, van Voorhis J, Nathan C, and Weinstein RA. The role of "colonization pressure" in the spread of vancomycin-resistant enterococci: an important infection control variable. *Archives of Internal Medicine*, 158:1127–1132, 1998.

[90] George DL, Falk PS, Wunderink RG, Leeper KV Jr, Meduri GU, Steere EL, Corbett CE, and Mayhall CG. Epidemiology of ventilator-acquired pneumonia based on protected bronchoscopic sampling. *American Journal of Respiratory and Critical Care Medicine*, 158:1839–1847, 1998.

[91] Koulenti D and Rello J. Gram-negative bacterial pneumonia: aetiology and management. *Current Opinion in Pulmonary Medicine*, 12:198–204, 2006.

[92] Hernández G, Rico P, Díaz E, and Rello J. Nosocomial lung infections in adult intensive care units. *Microbes and infection*, 6:1004–1014, 2004.

[93] Langer M, Cigada M, Mandelli M, Mosconi P, and Tognoni G. Early onset pneumonia: a multicenter study in intensive care units. *Intensive Care Medicine*, 13:342–346, 1987.

[94] Luna CM, Aruj P, Niederman MS, Garzón J, Violi D, Prignoni A, R F, Baquero S, and Gando S: Grupo Argentino de Estudio de la NeumonAsociada al Respirador group. Appropriateness and delay to initiate therapy in ventilator-associated pneumonia. *European Respiratory Journal*, 27:158–164, 2006.

[95] Dupont H, Mentec H, Sollet JP, and Bleichner G. Impact of appropriateness of initial antibiotic therapy on the outcome of ventilator-associated pneumonia. *Intensive Care Medicine*, 27:355–362, 2001.

[96] Teixeira PJ, Seligman R, Hertz FT, Cruz DB, and Fachel JM. Inadequate treatment of ventilator-associated pneumonia: risk factors and impact on outcomes. *The Journal of Hospital Infection*, 65:361–367, 2007.

[97] Paul M, Nielsen AD, Goldberg E, Andreassen S, Tacconelli E, Almanasreh N, Frank U, Cauda R, and Leibovici L: TREAT Study Group. Prediction of specific pathogens in patients with sepsis: evaluation of treat, a computerized decision-support system. *Journal of Antimicrobial Chemotherapy*, 59:1204–1207, 2007.

[98] Paul M, Andreassen S, Tacconelli E, Nielsen AD, Almanasreh N, Frank U, Cauda R, and Leibovici L: TREAT Study Group. Improving empirical antibiotic treatment using treat, a computerized decision support system: cluster randomized trial. *Journal of Antimicrobial Chemotherapy*, 58:1238–1245, 2006.

[99] Bonten MJM. Prevention of infection in the intensive care unit. *Current Opinion in Critical Care*, 10(5), 2004.

[100] Genesereth MR and Nilsson NJ. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann, 1987.

[101] Sox HC, Blatt MA, Higgins MC, and Marton KI. *Medical Decision Making*. Butterworth-Heinemann, 1988.

[102] Lucas PJF. Bayesian network modelling through qualitative patterns. *Artificial Intelligence*, 163:233–263, 2005.

[103] Heckerman D and Breese JS. Causal independence for probabilistic assessment and inference using bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics*, 26(6), 1996.

[104] Reiter R. On closed world databases. In H. Gaillaire and J. Minker, editors, *Logic and Data Bases*, pages 55–76. New York: Plenum Press, 1978.

[105] Jurgelenaite R, Lucas PJF, and Heskes T. Exploring the noisy threshold function in designing Bayesian networks. In *Proceedings of AI-2005, the Twenty-fifth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pages 133–146, 2005. Cambridge.

[106] Zhang LH and Poole D. Exploiting causal independence in bayesian networks inference. *Journal of Artificial Intelligence Research*, 5:301–328, 1996.

[107] Heckerman D and Breese JS. Causal independence for probabilistic assessment and inference using bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics*, 26(6), 1996.

[108] Jurgelenaite R and Lucas PJF. Exploiting causal independence in large bayesian networks. *Knowledge Based Systems Journal*, 18(4–5), 2005.

[109] Enderton HB. *A Mathematical Introduction to Logic*. Academic Press, 1972.

[110] Wegener I. *The Complexity of Boolean Functions*. John Wiley & Sons, 1987.

[111] Salgado CD, O'Grady N, and Farr BM. Prevention and control of antimicrobial-resistant infections in intensive care patients. *Critical Care Medicine*, 33:2373–2382, 2005.

[112] Charitos C, van der Gaag LC, Visscher S, Lucas PJ, and Schurink CA. A dynamic Bayesian network for diagnosing ventilator-associated pneumonia in ICU patients. Expert systems with applications, 2008, in press.

[113] Friedman C and Wyatt J. *Evaluation methods in medical informatics*. Springer-Verlag, New York, 1997.

[114] van Bemmel JH and Musen MA. *Handbook of Medical Informatics*. Bohn Staffleu Van Loghum, The Netherlands, 1997.

[115] Eclipsys `http://www.eclipsys.com/about/default.asp`.

[116] Williams EW and Lane D. *Web Database Applications*. O'reilly, Sebastopol, CA USA, 2002.

[117] Ammenwerth E and de Keizer N. A viewpoint on evidence-based health informatics, based on a pilot survey on evaluation studies in health care informatics. *Journal of the American Medical Informatics Association*, 14(3), 2007, pages = 368–371.

[118] Hugin `http://www.hugin.com`.

Nederlandse Samenvatting

Dit hoofdstuk vat het onderzoek en haar resultaten beschreven in dit proefschrift samen. Verder worden mogelijke alternatieven voor het ontwikkelen van probabilistische modellen voor de diagnose en behandeling van beademingsgerelateerde longontsteking (ventilator-associated pneumonie (VAP) in het Engels) gepresenteerd en daarbij suggesties gedaan voor verder onderzoek.

VAP Diagnosticeren

De zorg voor ernstig zieke Intensive Care (IC) patiënten brengt specifieke aspecten met zich mee. Zo worden, om verschillende redenen, frequent (soms dagelijks) röntgenfoto's van de longen gemaakt, monsters genomen van verschillende delen van het lichaam (zoals sputum, wonden, urine en bloed) voor bacteriologische kweek en worden dagelijks biochemische analyses van bloedmonsters verricht om infectieparameters te vervolgen. Patiënten die kunstmatig beademd worden zijn vatbaarder voor het oplopen van een longontsteking, doordat de natuurlijke barrière, het slijmvlies van de luchtpijp met bewegende trilharen die het slijm met bacteriën richting mond vervoeren, wordt omzeild door de beademingsbuis. De bacteriën die in de patiënt en de IC omgeving aanwezig zijn hebben hierdoor bijna vrij toegang tot het longweefsel. Het type bacterie waarmee de luchtpijp gekoloniseerd is, is afhankelijk van de duur van IC opname ofwel de duur van intubatie. Gedurende de eerste zes dagen zijn patiënten vooral gekoloniseerd met zogenaamde "early-onset" pathogenen, zoals *Staphylococcus aureus*, *Haemophilus influenzae* en *Streptococcus pneumoniae*. Deze bacteriën nemen patiënten mee van huis. "Late-onset" pathogenen zoals *Pseudomonas aeruginosa*, *Acinetobacter* species en Enterobacteriaceae, daarentegen, koloniseren deze patiënten voornamelijk als ze meer dan zes dagen op de IC te hebben gelegen (Hoofdstuk 2). Deze bacteriën zijn typische 'ziekenhuis bacteriën'.

Voor het onderzoek maakten we gebruik van een prospectief verzamelde dataset van meer dan 2000 beademde IC patiënten. Bij 70% van deze patiënten waren één of meer endotracheale kweekresultaten beschikbaar. Bij deze patiënten onderzochten we hoelang de bacteriën in de luchtpijp aanwezig bleven (persistentie), tijdens hun verblijf op de IC. Persistentie werd gedefinieerd als de duur van kolonisatie met een bepaalde bacterie van een patiënt. De persistentie bleek sterk afhankelijk te zijn van het soort bacterie. De mediane duur van kolonisatie was 4–5 dagen voor early-onset pathogenen en 5–8 dagen voor late-onset pathogenen. Aangezien veel IC patiënten antibiotica krijgen, onderzochten we de effecten van antibiotica op het verkrijgen (acquisitie) en de persistentie van de bacteriële tracheakolonisatie. Antibiotica reduceerden de kans op de acquisitie van endotracheale kolonisatie met early-onset pathogenen en Enterobacteriaceae (een groep van late-onset pathogenen) zolang deze bacteriën gevoelig waren voor de antibiotica. Daarnaast was de kans dat deze bacteriën ten tijde van antibiotica toediening verdwenen significant hoger vergeleken met andere late-onset pathogenen. Het werd duidelijk dat pathogenen zoals *P. aeruginosa* lang persisteren, ook al worden er antibiotica toegediend die in het laboratorium actief zijn tegen deze pathogeen. In een optimale onderzoekssetting zouden we persistentie van pathogenen analyseren in de afwezigheid van antibiotische behandeling. Met een dergelijke analyse zouden we het effect van 'tijd' op het kolonisatie- en persistentieproces kunnen kwantificeren. Helaas was het aantal patiënten in onze studiepopulatie dat gekoloniseerd was zonder dat antibiotica toegediend werden, te laag voor het uitvoeren van een dergelijke analyse.

Slechts een deel van de patiënten die gekoloniseerd zijn met bacteriën ontwikkelt VAP. Meestal is het ontstaan van VAP een geleidelijk proces van asymptomatische kolonisatie tot een klinisch-manifeste infectie. Helaas is er geen nauwkeurige gouden standaard voor de diagnose VAP voorhanden, wat het stellen van de diagnose moeilijk maakt. Als alternatief hebben we een referentie standaard gebruikt, die gebaseerd is op een combinatie van diagnostische tests, samengevat in een diagnostische beslisboom (Hoofdstuk 3). In de studiepopulatie werd er voor 238 episodes van verdenking op VAP antibiotica door de behandelend intensivisten voorgeschreven. Echter, door de klinische parameters op het tijdstip van verdenking op VAP en de klinische respons op de behandeling retrospectief te onderzoeken, concludeerden twee infectiologen uiteindelijk dat er in 157 van deze episodes daadwerkelijk sprake was geweest van VAP.

Het stellen van de diagnose VAP gebeurt dus op basis van onzekere parameters. In dit proefschrift beredeneren we dat het paradigma van Bayesiaanse netwerken geschikt is voor het omgaan met deze onzekere diagnostische processen en zo de arts kan helpen in het stellen van de diagnose VAP. De structuur van een Bayesiaans netwerk dat een complexe ziekte als VAP modelleert, kan het beste door een domeinexpert gegeven worden. Zeker omdat deze structuur cruciaal is voor het aanduiden van causale verbanden en onafhankelijkheden binnen het betreffende domein. Tegelijkertijd geldt dat hoe meer knopen geïntroduceerd worden in een model, hoe complexer dit model wordt. Zeker als het aankomt op het schatten van alle conditionele kansen door de expert (op basis van ervaring en/ of uit literatuur) is een minimale structuur gewenst, die toch alle belangrijke onafhankelijkheden bevat. In hoofdstuk 3 lieten we zien dat het Bayesiaanse netwerk dat het best onderscheid maakt tussen VAP en geen VAP een simpele structuur had en dat een uitgebreid model op basis van deze simpele structuur beter onderscheidde dan alle andere onderzochte modellen. De conclusie dat een Bayesiaans netwerk met een simpele structuur vaak het best presteert blijkt ook uit andere studies, ook buiten de geneeskunde.

Omdat we niet blind wilden varen op alleen de expertschattingen, hebben we ook een longitudinale database met daarin gegevens van kunstmatig beademde patiënten gebruikt om deze – zogenaamde – subjectieve kansschattingen te updaten. In het proces van het construeren van een klinisch beslissingsondersteunend systeem zouden artsen echter nooit uitgesloten mogen worden, omdat hun ervaring net zo belangrijk, of zelfs belangrijker kan worden geacht, dan prospectief verzamelde data. Om het model te optimaliseren werden de data daarom opgesplitst: de conditionele kansen werden vergeleken en – waar nodig – geüpdate in de ene helft van dataset en getest op de andere helft om overfitting te voorkomen. Deze procedure had tot gevolg dat de prestatie van het model verbeterde terwijl de testkarakteristieken goed bleven. De area under the ROC curve (AUC) was 0.857 [0.827–0.888 95% CI], sensitiviteit en specificiteit werden beiden zo optimaal mogelijk gekozen en waren 80%; de positief en negatief voorspellende waarden waren respectievelijk 6.1% en 99.6%. De lage positief voorspellende waarde hangt gedeeltelijk samen met de lage dagelijkse prevalentie van VAP. De optimale wisselwerking tussen sensitiviteit en specificiteit resulteerde in een drempelwaarde van 0.46: dat wil zeggen, als het model een kans op pneumonie tussen 0.46 en 1.0 voorspelt, zou de intensivist er vanuit kunnen gaan dat de patiënt VAP heeft en kunnen overwegen om te starten met het toedienen van antibiotica. Een a posteriori kans tussen 0 en 0.46, daarentegen, zou een lage waarschijnlijkheid op VAP betekenen.

Dynamiek in Diagnostische Parameters van VAP

In het tweede gedeelte van dit proefschrift hebben we getracht de temporele evolutie van VAP te ontrafelen. Hiervoor werd standaard beschrijvende statistiek aangevuld door het leren van dynamische Bayesiaanse netwerken uit data van patiënten met en zonder VAP. Het meest gebruikte instrument hiervoor was het concept van contextspecifieke afhankelijkheid.

Het Bayesiaans netwerkmodel, beschreven in Hoofdstuk 3, voorziet in dagelijkse voorspellingen van de kans op VAP, zonder daarbij rekening te houden met eventuele tijdstrends. In hoofdstuk 5 hebben we onderzocht in hoeverre waarden van bijvoorbeeld lichaamstemperatuur, sputum productie, leukocyten, en pO_2/ FiO_2 ratio enig verschil in verloop van tijd vertoonden. Hiervoor werd elke patiënt met VAP gematcht met drie niet-VAP patiënten; matchende criteria waren geslacht, type IC (algemene interne/ neurochirurgisch) en duur van beademing. In samenwerking met Charitos en Van der Gaag [112] ontwikkelden we een repetitief temporeel Bayesiaans netwerk voor het diagnosticeren van VAP. Aangenomen werd dat er een vaste transitie bestond tussen twee tijdseenheden, die twee opeenvolgende dagen representeerden, met een vaste structuur. We berekenden de waarschijnlijkheid op VAP op een bepaalde dag door de waarschijnlijkheid op VAP op de voorgaande dag mee te nemen in deze berekening. Deze aanpak leidde tot goede testresultaten, afgemeten aan AUC en Brier score. In dit proefschrift, echter, hebben we een alternatieve aanpak beschreven. We hebben onderzocht in hoeverre niet-repetitieve temporele Bayesiaanse netwerken kunnen worden geconstrueerd als een instrument voor het onderzoeken van de temporele evolutie van VAP bij mechanisch beademende IC patiënten. We ontdekten dat associaties tussen variabelen veranderlijk aan- en afwezig waren op verschillende tijdstippen. Dit bevestigde onze hypothese over niet-repetitief gedrag in dit specifieke domein. De temporele modellen die geconstrueerd werden aan de hand van het Necessary Path Condition (NPC) algoritme voor het automatisch leren van structuur uit data lieten verschillende (a)temporele patronen tussen VAP en niet-VAP patiënten zien. Op basis van deze bevindingen concludeerden we dat het NPC algoritme nuttig is, omdat het gebruik van achtergrondinformatie hierdoor mogelijk wordt gemaakt. Het algoritme combineert automatisch leren van onafhankelijkheden uit data met oplossen van onzekerheid over de aanwezigheid en de richting van afhankelijkheden door de gebruiker. Uitgaande van deze resultaten zouden 'normale' repetitieve dynamisch Bayesiaanse netwerken onvoldoende geschikt kunnen zijn voor het modelleren van de karakteristieken van een domein.

Selectie van Antibiotische Behandeling voor VAP

In het derde deel van dit proefschrift hebben we de modellen getest op het voorspellen van de bacteriële verwekker van VAP. Dit is van belang, omdat op het moment dat de diagnose VAP is gesteld, gerichte antibiotische therapie zo snel mogelijk gestart moet worden. In de klinische praktijk echter, zijn kweekresultaten afkomstig van tracheaal aspiraat niet direct voorhanden. Definitieve uitslagen kunnen soms wel twee dagen op zich laten wachten. Daarom wordt er meestal empirisch gestart met behandelen met onzekerheid over (gevoeligheid van) verwekkers, resulterend in (te) brede antimicrobiële dekking.

Het Bayesiaans netwerkmodel gebruikt een andere aanpak voor het selecteren van antibiotische behandeling voor VAP. Zoals beschreven in Hoofdstuk 6 worden de duur van de opname op de IC en duur van beademing gebruikt als voorspellers voor de aanwezigheid van het soort kolonisatie. Daarnaast worden kweekresultaten van tracheaal aspiraat in voorafgaande dagen en eerder antibioticagebruik (gelinkt aan de – *in-vitro* – gevoeligheid van de pathogenen) gebruikt als parameters. Gebaseerd op deze parameters werden a posteriori kolonisatie kansen berekend en de performance van het model werd geanalyseerd door middel van ROC analyse. De optimale drempelwaarde (maximaliseren van zowel sensitiviteit als specificiteit) werd aan de hand van deze analyse bepaald voor elk pathogeen. Voor *P. aeruginosa* (PA) bleek deze drempelwaarde 0.289 en die zou alsvolgt geïnterpreteerd moeten worden: wanneer de a posteriori kans op PA-kolonisatie groter is dan 0.289, is het aannemelijk dat de trachea van de patiënt gekoloniseerd is door PA. Bij een kans die lager is dan 0.289 kan men aannemen dat dit pathogeen niet aanwezig is. Deze procedure werd herhaald voor alle andere 6 pathogeen groepen in het netwerk en het bleek dat het Bayesiaanse netwerkmodel goede testkarakteristieken had voor het voorspellen van pathogenen die VAP veroorzaakten. Een nadeel van het dichotomiseren van a posteriori kolonisatiekansen is dat 20 van de in totaal 153 VAP episodes volgens de modelpredicties werden veroorzaakt door meer dan 3 pathogenen, wat in de praktijk onwaarschijnlijk is. Tevens zou dit leiden tot het onnodig voorschrijven van breed-spectrum antibiotica. Het nut van de modelpredicties voor de overige 133 episodes weegt echter op tegen deze 20.

In Hoofdstukken 7 en 8 worden verschillende modelleringmethoden beschreven die invloed hebben op de keuze van antibiotica voor VAP. Daarbij hebben we met verschillende modellen voor endotracheale kolonisatie onderzocht hoe antibiotica deze koloniserende bacteriën dekken. Zoals al eerder gesteld, is het belangrijk dat er snel gestart wordt met antibiotica als VAP wordt gediagnosticeerd. Het uitgangspunt is dan de kolonisatiestatus van de patiënt. Doorgaans controleert de arts de resultaten van eerdere kweken en zal op basis hiervan antibiotica voorschrijven. Het is aannemelijk dat wanneer uit eerdere kweken blijkt dat PA de trachea van de patiënt koloniseerde, deze op de dag van VAP verdenking, enkele dagen later, zeer waarschijnlijk nog steeds aanwezig is, zodat de arts de antibiotica op de PA richt (Hoofdstuk 2).

Zoals ook blijkt uit Hoofdstuk 2, is er verschil tussen de duur van persisterende kolonisatie, zodat de kolonisatiestatus van een patiënt nog steeds onzeker kan zijn. Met andere woorden: gegevens over eerdere kolonisatie kunnen nuttig zijn, maar zeggen niet alles over de kolonisatiestatus van een patiënt. Daarbij moet worden opgemerkt dat kweken niet alle pathogenen detecteren, bijvoorbeeld wanneer hun aantal onder een bepaald detectieniveau ligt. In deze situaties zou een Bayesiaans netwerkmodel een betrouwbaar beeld geven van de kolonisatiestatus van de patiënt, gebaseerd op duur van hospitalisatie en mechanische ventilatie.

In Hoofdstuk 7 lieten we zien dat toepassen van de closed-world assumption (gesloten wereld aanname; CWA), ofwel de aanname dat onbekende testresultaten als negatief worden aangenomen, bruikbaar bleek bij het interpreteren van uitslagen van endotracheale kweken. Een nadeel is dat de CWA geen rekening houdt met het hypothetische geval dat een ander pathogeen de patiënt koloniseert dan degene waarvan op basis van eerdere kweken bekend is dat deze aanwezig of afwezig is. Daarnaast houdt CWA geen rekening met de mogelijkheid dat kolonisatie verdwijnt.

In Hoofdstuk 8 beschrijven we drie verschillende manieren van modelleren van antibiotische gevoeligheid. Deze worden 'gevoeligheidsmodellen' genoemd:

I antibiotische dekking van zowel door middel van microbiologische kweken geïdentificeerde en niet-geïdentificeerde pathogenen wordt als adequaat beschouwd;

II antibiotische dekking van door middel van microbiologische kweken geïdentificeerde pathogenen wordt als adequaat beschouwd, terwijl dit niet geldt voor dekking van niet-geïdentificeerde pathogenen;

III alleen antibiotische dekking van door middel van microbiologische kweken geïdentificeerde pathogenen wordt als adequaat beschouwd.

De kansdistributie van de knoop die de algehele gevoeligheid ofwel de dekking representeert, is uitgedrukt in het conjunctieve effect van de zeven verschillende pathogeen groepen. Deze vorm van modelleren is beter bekend als de noisy-AND causaal onafhankelijkheidsmodel. Het nadeel van het gebruik van de noisy-AND in deze setting is dat het model probeert om alle pathogenen te dekken, terwijl uit de data blijkt dat patiënten met hooguit drie pathogenen tegelijkertijd gekoloniseerd zijn. Dit resulteerde in het advies om onnodig breed-spectrum antibiotica voor te schrijven. Daarom onderzochten we het nut van het gebruik van Booleaanse drempelwaarde functies als generalisaties van de Booleaanse AND (100% dekking voor alle pathogenen resulteert in 100% algehele dekking) en OR (100% dekking voor een of meerdere pathogenen resulteert in 100% algehele dekking). Het is mogelijk om door gebruik van deze drempelwaarde functies minder dan zeven pathogenen te dekken. We concluderen dat gevoeligheidsmodel III in combinatie met drempelwaarde functie τ_k met $k = 1, 2$ het beste presteert in het voorspeld dekken van bacteriën die VAP veroorzaken, zowel bij monobacteriële als bij polymicrobiële VAP episodes.

Verder Onderzoek

De diagnose en behandeling van VAP is een voorbeeld van complexe patiëntenzorg op de IC. Sommige facetten daarvan werden onderzocht in dit proefschrift, maar vele vragen zijn nog niet, of slechts gedeeltelijk, beantwoord. We kozen het formalisme van Bayesiaanse netwerken, omdat dit één van de weinige statistische formalismen is dat integratie van expertkennis en kennis uit data mogelijk maakt. Gebruikmakend van dit formalisme voor het construeren van een model dat gebruikt kan worden voor diagnostische en behandelingsdoeleinden, kwamen we een aantal fundamentele problemen tegen die opgelost moesten worden. Sommigen problemen hebben de revue reeds gepasseerd, terwijl anderen hierna besproken zullen worden.

Modelleren van kolonisatiedynamiek

Nu we de dynamiek van endotracheale kolonisatie in IC patiënten hebben beschreven en gekwantificeerd, zou een logische volgende stap zijn om een dynamisch Bayesiaans netwerk (DBN) te construeren dat deze kolonisatiepatronen zou modelleren. Een dergelijk DBN zou de aan- en afwezigheid van de meest voorkomende pathogenen die IC patiënten koloniseren kunnen voorspellen, gebaseerd op de duur van opname en duur van mechanische ventilatie.

Temporele Bayesiaanse netwerken

Het gebruik van temporele Bayesiaanse netwerken (TBNs) neemt toe, aangezien er vele problemen zijn waar tijd een rol speelt. TBNs zijn handige instrumenten voor de constructie van compacte probabilistische modellen waarin tijd een rol speelt. Deze TBNs zijn repetitive Bayesiaanse netwerken met kansdistributies waarvan de transities tijdsinvariant zijn. Deze TBNs worden ook wel dynamisch Bayesiaanse netwerken genoemd, aangezien ze kunnen worden geconstrueerd uit een specificatie bestaande uit twee tijdseenheden die elkaar opvolgen, wat een nog compactere specificatie biedt. Ondanks dat DBNs krachtige representaties voor verscheidene vormen van beslissingsproblemen bieden, zijn ze niet geschikt voor het nauwkeurig bepalen van subtiele veranderingen van afhankelijkheid en onverwachte informatie in de tijd. Voor dit doel zijn niet-repetitieve TBNs meer geschikt en inclusie van contextspecifieke afhankelijkheid zou het formalisme zelfs krachtiger maken.

Op dit moment zijn er geen instrumenten beschikbaar die het exploreren van de contextspecifieke onafhankelijkheid in de tijd mogelijk maken en daardoor waren we genoodzaakt terug te vallen op het gebruik van standaard tools voor Bayesiaanse netwerken. Uitbreiding van bestaande softwarepakketten voor Bayesiaanse netwerken door faciliteiten voor non-repetitieve TBNs toe te voegen, lijkt een gepast vervolg op het onderzoek beschreven in dit proefschrift.

Modelleren van interacties

Tijdens het modelleren van interacties tussen variabelen die invloed hebben op een bepaalde uitkomst, kunnen probabilistische modellen snel uitgroeien en log worden. Het aantal kansen dat moet worden gespecificeerd voor een discrete conditionele kansdistributie is namelijk exponentieel in het aantal variabelen dat geassocieerd is met de ouders van een variabele. Lineaire benaderingsmethoden, zoals logistische regressie, worden daardoor veel gebruikt. Het probleem dat probabilistische modellen exponentieel kunnen groeien heeft zich ook tijdens ons onderzoek voorgedaan en werd aangepakt door het concept van causale onafhankelijkheid te gebruiken. Dit is één van de vele typen probabilistische modellen, waarvoor kansdistributies worden gegeneerd op basis van het aantal parameters. We hebben Booleaanse functies, in het bijzonder Booleaanse drempelwaardefuncties, gebruikt voor dit doeleinde. Aangezien er 2^{n^n} verschillende Booleaanse functies en 2^{n+1} symmetrische Booleaanse functies zijn, waarbij n het aantal Booleaanse variabelen in het domein is, kunnen vele andere functies nog onderzocht worden voor een meer generaliserende manier van modelleren van interacties dan voor de noisy-AND het geval is.

Modelleren van utiliteit en selectie van behandeling

Veel aandacht is in dit proefschrift uitgegaan naar gestructureerde kansdistributies voor het stellen van de diagnose VAP. Minder aandacht werd besteed aan het modelleren van utiliteitsinformatie die nodig is voor de selectie van een geschikte behandeling voor VAP. Vervolgonderzoek naar het bepalen van utiliteiten voor de selectie van antimicrobiële behandeling ligt daarom voor de hand.

Rol van beslissingondersteuning bij de diagnose en behandeling van VAP

Aangezien VAP geassocieerd wordt met een hoge mortaliteit zou beslissingsondersteuning, waardoor VAP sneller gediagnosticeerd en beter behandeld kan worden, de patiënt ten goede kunnen komen. Afgezien van het probleem van het stellen van de diagnose VAP, zijn we erin geslaagd een diagnostisch deel van een Bayesiaans netwerk te construeren dat in staat is onderscheid te maken tussen VAP en niet-VAP patiënten (vergeleken met de geconstrueerde referentie standaard). De positief voorspellende waarde is nog steeds laag, maar de extreem hoge negatief voorspellende waarde zou kunnen bijdragen aan reductie van onnodig voorgeschreven antibiotica. De invloed en praktische meerwaarde van het geconstrueerde Bayesiaanse beslissingondersteunde systeem (BDSS) in de kliniek is vooralsnog onbekend en zal daartoe in de toekomst vastgesteld en geëvalueerd moeten worden.

Klinische evaluatie van beslissingsondersteunende systemen

Het Bayesiaans netwerkmodel beschreven in dit proefschrift is inmiddels geïntegreerd in een infrastructuur met onder andere een Linux server. Dit maakt het gebruik van het model op een Internet browser via Ethernet mogelijk. En daarmee is ook een klinische evaluatie van het systeem mogelijk. Het verfijnen van het model kostte echter veel meer tijd dan oorspronkelijk gepland, omdat het stellen van de diagnose en het selecteren van adequate behandeling bij IC patiënten moeilijker bleek dan verwacht. Helaas is deze gang van zaken veel voorkomend bij het ontwikkelen van klinische beslissingsondersteunende systemen, waardoor weinig van deze systemen al geëvalueerd zijn in een klinische setting [117]. Het is echter belangrijk dat het BDSS geëvalueerd wordt door gebruikers in de kliniek.

Voordat een BDSS geïntegreerd kan worden in de kliniek, zouden de effecten van een dergelijk systeem op de patiëntenzorg en de bereidwilligheid van artsen om een dergelijke vorm van beslissingsondersteuning te gebruiken vastgesteld moeten worden.

Om de prestaties van de BDSS op het gebied van het stellen van de diagnose VAP en het selecteren van geschikte antibiotische behandeling te evalueren stellen we een gerandomiseerde studie voor. Het domein van de studie zou 'alle patiënten langer dan 48 uur beademd met een klinische verdenking op VAP volgens de behandelend arts' zijn. De determinant zal het advies van de BDSS zijn. Accepterend dat er geen geschikte gouden standaard voor VAP voorhanden is, zullen er verscheidene klinische eindpunten gekozen kunnen worden, zoals 'uitkomst voor de patiënt' (bijvoorbeeld overleving, duur van IC en/ of ziekenhuisopname) of antibiotica gebruik (bijvoorbeeld totale antibiotica gebruik of proportie van patiënten dat de juiste antibiotische behandeling krijgt).

Voor elke patiënt die in aanmerking komt (dit is 'beademd' en 'klinische verdenking VAP') zal de arts vervolgens gevraagd worden te beoordelen in hoeverre er sprake is van VAP (door middel van een simpel ja/ nee of een waarschijnlijkheid in het bereik van 0 tot 100%). Daarnaast wordt de arts gevraagd wat volgens hem de meest geschikte antibioticabehandeling zou zijn. Vervolgens presenteert het systeem de klinische symptomen van de patiënt (Zie Figuur 9.1), direct afgeleid uit de klinische database. Indien nodig kunnen er bepaalde criteria handmatig toegevoegd of veranderd worden door de arts. Nadat dit formulier is gepresenteerd, voorziet het systeem in het geven van een advies (waarschijnlijkheid van VAP en voorstel van antibiotica) dat gerandomiseerd gepresenteerd wordt aan de arts. Voor de gevallen waarvoor het advies van het de BDSS is getoond (i.e., in ongeveer 50% van de gevallen), zal de arts worden gevraagd – aan de hand van het advies van het systeem – in hoeverre hij blijft bij de aanvankelijke mening over de waarschijnlijkheid van VAP en antibioticakeuze. Dit studieontwerp zou de effecten van de BDSS op de behandeling van VAP kunnen kwantificeren.

Evaluatie biases

Bij het evalueren van de BDSS in de IC zouden resultaten wel eens een vertekend beeld kunnen geven, ten gevolge van onbekende systematische effecten. Met deze effecten moet rekening gehouden worden bij het ontwerp van een dergelijke studie. Bekende typen biases zijn onder andere [113, 114]:

- Het vrijwilligers effect: artsen die vrijwillig meedoen aan het gebruik van het systeem zouden beter kunnen presteren dan anderen. Dit zou kunnen worden voorkomen door alle artsen te laten participeren.
- De assessment bias: kennis van de gouden standaard zou de beslissingen van de gebruikers kunnen beïnvloeden. Maar aangezien er geen echt geschikte gouden standaard voor VAP beschikbaar is, lijkt dit geen rol te spelen.
- Het Hawthorne effect: klinische prestaties verbeteren wanneer clinici weten dat ze worden geobserveerd of bestudeerd. Dit is een effect dat bijna niet te voorkomen valt. Echter door alle artsen het eerste deel van de interventie te laten doorlopen (waarin zij hun mening geven over de waarschijnlijkheid van VAP en patiëntkarakteristieken aanvullen) zullen zowel degenen die wel als zij die geen terugkoppeling van de BDSS krijgen dezelfde vorm van 'bestudeerd worden' ervaren.
- Het checklist effect: prestaties kunnen verbeteren wanneer clinici een checklist gebruiken, doordat er herinnerd wordt aan de factoren die relevant zijn in het specifieke domein. Bij VAP zou dit neerkomen op bijvoorbeeld koorts, verhoogde productie van sputum, leukocytose, etc.

Resulterende BDSS voor VAP

In deze paragraaf beschrijven we de technische aspecten van het uitvoeren van de voorgestelde evaluatie studie in de IC's van het UMC Utrecht. Ten eerste zal het systeem moeten worden geïntegreerd in het klinisch informatie systeem dat gebruikt wordt in de IC's. Dit systeem fungeert als elektronisch patiënten dossier (EPD) en bevat een rijkdom aan informatie dat gebruikt kan worden voor beslissingsondersteuning.

De voorste laag van de BDSS bestaat uit een grafische user-interface met pop-up menu's, tabellen en grafieken, die het de arts mogelijk maken de gegevens van de patiënt in te zien. De onderste laag van het systeem is een relationeel database management systeem, dat verbonden is met het relationele database systeem van het EPD. Hierdoor wordt niet alleen op een moderne manier beveiligde gegevensopslag en het bijwerken en ontvangen ervan gefaciliteerd, maar voorziet het in een Structured Query Language (SQL) interface die externe systemen toegang tot de gegevens verschaft.

In het UMCU werd besloten tot een afzonderlijke user-interface voor de BDSS om de afzonderlijke ontwikkeling van het onderliggende Bayesiaanse netwerk en de user-interface mogelijk te maken. Om dit te verwezenlijken gebruikten we moderne gedistribueerde informatiesysteem technologieën, zoals Hypertext Preprocessor (PHP). PHP is een server-side platformonafhankelijke scriptingtaal (inclusief hypertext markup language (HTML)) dat het genereren van de dynamische inhoud van HTML pagina's mogelijk maakt.

PHP wordt gebruikt als taal om van verschillende onderdelen van het systeem te verbinden. Het merendeel van de functionaliteit van het systeem wordt verzorgd door een commercieel beschikbaar softwarepakket voor Bayesiaanse en beslisnetwerken [118]. Het pakket is in ons project gebruikt om de inference engine (inferentie motor) voor de BDSS te implementeren. Dit gedeelte van het systeem verwerkt de patiëntgegevens en biedt verscheidene vormen van advies op basis van de resultaten die berekend zijn na het invoeren van deze gegevens in het Bayesiaans netwerk model. Het berekende advies wordt vervolgens aan de gebruiker getoond.

De laatste componenten van het systeem zijn een hypertext transfer protocol (HTTP) Apache server en een Web browser. Deze Web browser fungeert als user-interface voor de BDSS. De gehele architectuur is gevisualiseerd in Figuur 9.2.

Journal Publications

[1] Stefan Visscher, Lisa Kruisheer, Peter Lucas, Karin Schurink, and Marc Bonten. Predicting pathogens causing ventilator-associated pneumonia using a Bayesian network model. *Journal of Antimicrobial Chemotherapy*, 62(1):184–188, 2008.

[2] Stefan Visscher, Karin Schurink, Marianne Melsen, Peter Lucas, and Marc Bonten. Effects of systemic antibiotic therapy on bacterial persistence in the respiratory tract of mechanically ventilated patients. *Intensive Care Medicine*, 34(4):692–699, 2008.

[3] Theodore Charitos, Linda van der Gaag, Stefan Visscher, Peter Lucas, and Karin Schurink. A dynamic Bayesian network for diagnosing ventilator-associated pneumonia in ICU patients. Expert systems with applications, 2008, in press.

[4] Stefan Visscher, Peter Lucas, Karin Schurink, and Marc Bonten. Modelling treatment effects in a clinical Bayesian network using Boolean threshold functions. *Artificial Intelligence in Medicine*, 2008. In press.

[5] Stefan Visscher, Peter Lucas, Ildikó Flesch, Karin Schurink, , and Marc Bonten. The exploratory analysis of disease processes using temporal context-specific independence. Submitted, 2008.

[6] Karin Schurink, Stefan Visscher, Peter Lucas, Henk van Leeuwen, Erik Buskens, Reinier Hoff, Andy Hoepelman, and Marc Bonten. A Bayesian decision-support system for diagnosing ventilator-associated pneumonia. *Intensive Care Medicine*, 33(8):1379–1386, 2007.

Conference Proceedings

[1] Stefan Visscher, Peter Lucas, Ildikó Flesch, and Karin Schurink. Using temporal context-specific independence information in the exploratory analysis of disease processes. In R. Bellazzi, A. Abu-Hanna, and J. Hunter, editors, *11th Conference on Artificial Intelligence in Medicine, AIME 2007 proceedings*, pages 87–96. Lecture Notes in Artificial Intelligence 4594, Springer-Verlag, 2007.

[2] Stefan Visscher, Karin Schurink, Peter Lucas, and Marc Bonten. Bacterial persistence of respiratory tract in ICU patients. Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC), 2006.

[3] Theodore Charitos, Stefan Visscher, Linda van der Gaag, Peter Lucas, and Karin Schurink. A dynamic model for therapy selection in ICU patients with VAP. In Niels B. Peek and Carlo Combi, editors, *Proceedings of the 11th Intelligent Data Analysis in Medicine And Pharmacology (IDAMAP) workshop*, pages 71–76, 2006.

[4] Ildikó Flesch, Peter Lucas, and Stefan Visscher. On the modularisation of independence in dynamic Bayesian networks. BeNeLux conference on Artificial Intelligence: **Best paper award**, 2006.

[5] Stefan Visscher, Karin Schurink, Marc Bonten, and Peter Lucas. Application of the close-world assumption to a Bayesian network for optimal treatment selection. 3rd International Federation for Medical and Biological Engineering (IFMBE), Proceedings of the 3rd European Medical and Biological Engineering Conference (EMBEC 2005), 2005.

[6] Stefan Visscher, Peter Lucas, Marc Bonten, and Karin Schurink. Improving the therapeutic performance of a medical Bayesian network using noisy threshold functions. In José Louis Oliveira, Víctor Maojo, Fernando Martin-Sanchez, and António Sousa Pereira, editors, *6st International Symposium on Biological and Medical Data Analysis, ISBMDA 2005 proceedings*, pages 161–172. Lecture Notes in Bioinformatics 3745, Springer-Verlag, 2005.

[7] Stefan Visscher, Karin Schurink, Marc Bonten, and Peter Lucas. Using a Bayesian-network model for the analysis of clinical time-series data. In Silvia Miksch, Jim Hunter, and Elpida Keravnou, editors, *10th Conference on Artificial Intelligence in Medicine, AIME 2005 proceedings*, pages 48–52. Lecture Notes in Artificial Intelligence 3581, Springer-Verlag, 2005.

[8] Theodore Charitos, Linda van der Gaag, Stefan Visscher, Karin Schurink, and Peter Lucas. A dynamic bayesian network for diagnosing ventilator-associated pneumonia in ICU patients. In John H. Holmes and Niels B. Peek, editors, *Proceedings of the 10th Intelligent Data Analysis in Medicine And Pharmacology (IDAMAP) workshop*, pages 32–37, 2005.

[9] Stefan Visscher, Karin Schurink, Marc Bonten, and Peter Lucas. Improving the diagnostic performance of a Bayesian network for the management of ventilator-associated pneumonia using data. In the proceedings of the Medical Informatics Congress (MIC 2005), 2005.

[10] Stefan Visscher, Karin Schurink, Marc Bonten, and Peter Lucas. Incorporating evaluation into the design of a decision-support system. *Studies in Health care technology and Informatics*, 110:54–60, 2004.

SIKS Dissertations

1998

1998-1 Johan van den Akker (CWI)
DEGAS - An Active, Temporal Database of Autonomous Objects

1998-2 Floris Wiesman (UM)
Information Retrieval by Graphically Browsing Meta-Information

1998-3 Ans Steuten (TUD)
A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective

1998-4 Dennis Breuker (UM)
Memory versus Search in Games

1998-5 Eduard Oskamp (RUL)
Computerondersteuning bij Straftoemeting

1999

1999-1 Mark Sloof (VU)
Physiology of Quality Change Modelling; Automated modelling of Quality Change of Agricultural Products

1999-2 Rob Potharst (EUR)
Classification using decision trees and neural nets

1999-3 Don Beal (UM)
The Nature of Minimax Search

1999-4 Jacques Penders (UM)
The practical Art of Moving Physical Objects

1999-5 Aldo de Moor (KUB)
Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems

1999-6 Niek Wijngaards (VU)
Re-design of compositional systems

1999-7 David Spelt (UT)
Verification support for object database design

1999-8 Jacques Lenting (UM)
Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation

2000

2000-1 Frank Niessink (VU)
Perspectives on Improving Software Maintenance

2000-2 Koen Holtman (TUE)
Prototyping of CMS Storage Management

2000-3 Carolien Metselaar (UvA)
Sociaal-organisatorische gevolgen van kennistechnologie; een procesbenadering en actorperspectie

2000-4 Geert de Haan (VU)
ETAG, A Formal Model of Competence Knowledge for User Interface Design

2000-5 Ruud van der Pol (UM)
Knowledge-based Query Formulation in Information Retrieval

2000-6 Rogier van Eijk (UU)
Programming Languages for Agent Communication

2000-7 Niels Peek (UU)
Decision-theoretic Planning of Clinical Patient Management

2000-8 Veerle Coupé (EUR)
Sensitivity Analyis of Decision-Theoretic Networks

2000-9 Florian Waas (CWI)
Principles of Probabilistic Query Optimization

2000-10 Niels Nes (CWI)
Image Database Management System Design Considerations, Algorithms and Architecture

2000-11 Jonas Karlsson (CWI)
Scalable Distributed Data Structures for Database Management

2001

2001-1 Silja Renooij (UU)
Qualitative Approaches to Quantifying Probabilistic Networks

2001-2 Koen Hindriks (UU)
Agent Programming Languages: Programming with Mental Models

2001-3 Maarten van Someren (UvA)
Learning as problem solving

2001-4 Evgueni Smirnov (UM)
Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets

2001-5 Jacco van Ossenbruggen (VU)
Processing Structured Hypermedia: A Matter of Style

2001-6 Martijn van Welie (VU)
Task-based User Interface Design

2001-7 Bastiaan Schonhage (VU)
Diva: Architectural Perspectives on Information Visualization

2001-8 Pascal van Eck (VU)
A Compositional Semantic Structure for Multi-Agent Systems Dynamics

2001-9 Pieter Jan 't Hoen (RUL)
Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes

2001-10 Maarten Sierhuis (UvA)
Modeling and Simulating Work Practice BRAHMS: a multiagent modeling and simulation language for work practice analysis and design

2001-11 Tom van Engers (VU)
Knowledge Management: The Role of Mental Models in Business Systems Design

2002

2002-01 Nico Lassing (VU)
Architecture-Level Modifiability Analysis

2002-02 Roelof van Zwol (UT)
Modelling and searching web-based document collections

2002-03 Henk Ernst Blok (UT)
Database Optimization Aspects for Information Retrieval

2002-04 Juan Roberto Castelo Valdueza (UU)
The Discrete Acyclic Digraph Markov Model in Data Mining

2002-05 Radu Serban (VU)
The Private Cyberspace Modeling Electronic Environments inhabited by Privacy-concerned Agents

2002-06 Laurens Mommers (UL)
Applied legal epistemology; Building a knowledge-based ontology of the legal domain

2002-07 Peter Boncz (CWI)
Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications

2002-08 Jaap Gordijn (VU)
Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas

2002-09 Willem-Jan van den Heuvel (KUB)
Integrating Modern Business Applications with Objectified Legacy Systems

2002-10 Brian Sheppard (UM)
Towards Perfect Play of Scrabble

2002-11 Wouter Wijngaards (VU)
Agent Based Modelling of Dynamics: Biological and Organisational Applications

2002-12 Albrecht Schmidt (UvA)
Processing XML in Database Systems

2002-13 Hongjing Wu (TUE)
A Reference Architecture for Adaptive Hypermedia Applications

2002-14 Wieke de Vries (UU)
Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems

2002-15 Rik Eshuis (UT)
Semantics and Verification of UML Activity Diagrams for Workflow Modelling

2002-16 Pieter van Langen (VU)
The Anatomy of Design: Foundations, Models and Applications

2002-17 Stefan Manegold (UvA)
Understanding, Modeling, and Improving Main-Memory Database Performance

2003

2003-01 Heiner Stuckenschmidt (VU)
Ontology-Based Information Sharing in Weakly Structured Environments

2003-02 Jan Broersen (VU)
Modal Action Logics for Reasoning About Reactive Systems

2003-03 Martijn Schuemie (TUD)
Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy

2003-04 Milan Petkovic (UT)
Content-Based Video Retrieval Supported by Database Technology

2003-05 Jos Lehmann (UvA)
Causation in Artificial Intelligence and Law - A modelling approach

2003-06 Boris van Schooten (UT)
Development and specification of virtual environments

2003-07 Machiel Jansen (UvA)
Formal Explorations of Knowledge Intensive Tasks

2003-08 Yongping Ran (UM)
Repair Based Scheduling

2003-09 Rens Kortmann (UM)
The resolution of visually guided behaviour

2003-10 Andreas Lincke (UvT)
Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture

2003-11 Simon Keizer (UT)
Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks

2003-12 Roeland Ordelman (UT)
Dutch speech recognition in multimedia information retrieval

2003-13 Jeroen Donkers (UM)
Nosce Hostem - Searching with Opponent Models

2003-14 Stijn Hoppenbrouwers (KUN)
Freezing Language: Conceptualisation Processes across ICT-Supported Organisations

2003-15 Mathijs de Weerdt (TUD)
Plan Merging in Multi-Agent Systems

2003-16 Menzo Windhouwer (CWI)
Feature Grammar Systems - Incremental Maintenance of Indexes to Digital Media Warehouses

2003-17 David Jansen (UT)
Extensions of Statecharts with Probability, Time, and Stochastic Timing

2003-18 Levente Kocsis (UM)
Learning Search Decisions

2004

2004-01 Virginia Dignum (UU)
A Model for Organizational Interaction: Based on Agents, Founded in Logic

2004-02 Lai Xu (UvT)
Monitoring Multi-party Contracts for E-business

2004-03 Perry Groot (VU)
A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving

2004-04 Chris van Aart (UVA)
Organizational Principles for Multi-Agent Architectures

2004-05 Viara Popova (EUR)
Knowledge discovery and monotonicity

2004-06 Bart-Jan Hommes (TUD)
The Evaluation of Business Process Modeling Techniques

2004-07 Elise Boltjes (UM)
Voorbeeldig onderwijs; voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes

2004-08 Joop Verbeek(UM)
Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politiële gegevensuitwisseling en digitale expertise

2004-09 Martin Caminada (VU)
For the Sake of the Argument; explorations into argument-based reasoning

2004-10 Suzanne Kabel (UVA)
Knowledge-rich indexing of learning-objects

2004-11 Michel Klein (VU)
Change Management for Distributed Ontologies

2004-12 The Duy Bui (UT)
Creating emotions and facial expressions for embodied agents

2004-13 Wojciech Jamroga (UT)
Using Multiple Models of Reality: On Agents who Know how to Play

2004-14 Paul Harrenstein (UU)
Logic in Conflict. Logical Explorations in Strategic Equilibrium

2004-15 Arno Knobbe (UU)
Multi-Relational Data Mining

2004-16 Federico Divina (VU)
Hybrid Genetic Relational Search for Inductive Learning

2004-17 Mark Winands (UM)
Informed Search in Complex Games

2004-18 Vania Bessa Machado (UvA)
Supporting the Construction of Qualitative Knowledge Models

2004-19 Thijs Westerveld (UT)
Using generative probabilistic models for multimedia retrieval

2004-20 Madelon Evers (Nyenrode)
Learning from Design: facilitating multidisciplinary design teams

2005

2005-01 Floor Verdenius (UVA)
Methodological Aspects of Designing Induction-Based Applications

2005-02 Erik van der Werf (UM)
AI techniques for the game of Go

2005-03 Franc Grootjen (RUN)
A Pragmatic Approach to the Conceptualisation of Language

2005-04 Nirvana Meratnia (UT)
Towards Database Support for Moving Object data

2005-05 Gabriel Infante-Lopez (UVA)
Two-Level Probabilistic Grammars for Natural Language Parsing

2005-06 Pieter Spronck (UM)
Adaptive Game AI

2005-07 Flavius Frasincar (TUE)
Hypermedia Presentation Generation for Semantic Web Information Systems

2005-08 Richard Vdovjak (TUE)
A Model-driven Approach for Building Distributed Ontology-based Web Applications

2005-09 Jeen Broekstra (VU)
Storage, Querying and Inferencing for Semantic Web Languages

2005-10 Anders Bouwer (UVA)
Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments

2005-11 Elth Ogston (VU)
Agent Based Matchmaking and Clustering - A Decentralized Approach to Search

2005-12 Csaba Boer (EUR)
Distributed Simulation in Industry

2005-13 Fred Hamburg (UL)
Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen

2005-14 Borys Omelayenko (VU)
Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics

2005-15 Tibor Bosse (VU)
Analysis of the Dynamics of Cognitive Processes

2005-16 Joris Graaumans (UU)
Usability of XML Query Languages

2005-17 Boris Shishkov (TUD)
Software Specification Based on Re-usable Business Components

2005-18 Danielle Sent (UU)
Test-selection strategies for probabilistic networks

2005-19 Michel van Dartel (UM)
Situated Representation

2005-20 Cristina Coteanu (UL)
Cyber Consumer Law, State of the Art and Perspectives

2005-21 Wijnand Derks (UT)
Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics

2006

2006-01 Samuil Angelov (TUE)
Foundations of B2B Electronic Contracting

2006-02 Cristina Chisalita (VU)
Contextual issues in the design and use of information technology in organizations

2006-03 Noor Christoph (UVA)
The role of metacognitive skills in learning to solve problems

2006-04 Marta Sabou (VU)
Building Web Service Ontologies

2006-05 Cees Pierik (UU)
Validation Techniques for Object-Oriented Proof Outlines

2006-06 Ziv Baida (VU)
Software-aided Service Bundling - Intelligent Methods & Tools for Graphical Service Modeling

2006-07 Marko Smiljanic (UT)
XML schema matching – balancing efficiency and effectiveness by means of clustering

2006-08 Eelco Herder (UT)
Forward, Back and Home Again - Analyzing User Behavior on the Web

2006-09 Mohamed Wahdan (UM)
Automatic Formulation of the Auditor's Opinion

2006-10 Ronny Siebes (VU)
Semantic Routing in Peer-to-Peer Systems

2006-11 Joeri van Ruth (UT)
Flattening Queries over Nested Data Types

2006-12 Bert Bongers (VU)
Interactivation - Towards an e-cology of people, our technological environment, and the arts

2006-13 Henk-Jan Lebbink (UU)
Dialogue and Decision Games for Information Exchanging Agents

2006-14 Johan Hoorn (VU)
Software Requirements: Update, Upgrade, Redesign - towards a Theory of Requirements Change

2006-15 Rainer Malik (UU)
CONAN: Text Mining in the Biomedical Domain

2006-16 Carsten Riggelsen (UU)
Approximation Methods for Efficient Learning of Bayesian Networks

2006-17 Stacey Nagata (UU)
User Assistance for Multitasking with Interruptions on a Mobile Device

2006-18 Valentin Zhizhkun (UVA)
Graph transformation for Natural Language Processing

2006-19 Birna van Riemsdijk (UU)
Cognitive Agent Programming: A Semantic Approach

2006-20 Marina Velikova (UvT)
Monotone models for prediction in data mining

2006-21 Bas van Gils (RUN)
Aptness on the Web

2006-22 Paul de Vrieze (RUN)
Fundaments of Adaptive Personalisation

2006-23 Ion Juvina (UU)
Development of Cognitive Model for Navigating on the Web

2006-24 Laura Hollink (VU)
Semantic Annotation for Retrieval of Visual Resources

2006-25 Madalina Drugan (UU)
Conditional log-likelihood MDL and Evolutionary MCMC

2006-26 Vojkan Mihajlovic (UT)
Score Region Algebra: A Flexible Framework for Structured Information Retrieval

2006-27 Stefano Bocconi (CWI)
Vox Populi: generating video documentaries from semantically annotated media repositories

2006-28 Borkur Sigurbjornsson (UVA)
Focused Information Access using XML Element Retrieval

2007

2007-01 Kees Leune (UvT)
Access Control and Service-Oriented Architectures

2007-02 Wouter Teepe (RUG)
Reconciling Information Exchange and Confidentiality: A Formal Approach

2007-03 Peter Mika (VU)
Social Networks and the Semantic Web

2007-04 Jurriaan van Diggelen (UU)
Achieving Semantic Interoperability in Multi-agent Systems: a Dialogue-based Approach

2007-05 Bart Schermer (UL)
Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance

2007-06 Gilad Mishne (UVA)
Applied Text Analytics for Blogs

2007-07 Natasa Jovanovic (UT)
To Whom It May Concern - Addressee Identification in Face-to-Face Meetings

2007-08 Mark Hoogendoorn (VU)
Modeling of Change in Multi-Agent Organizations

2007-09 David Mobach (VU)
Agent-Based Mediated Service Negotiation

2007-10 Huib Aldewereld (UU)
Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols

2007-11 Natalia Stash (TUE)
Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System

2007-12 Marcel van Gerven (RUN)
Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty

2007-13 Rutger Rienks (UT)
Meetings in Smart Environments; Implications of Progressing Technology

2007-14 Niek Bergboer (UM)
Context-Based Image Analysis

2007-15 Joyca Lacroix (UM)
NIM: a Situated Computational Memory Model

2007-16 Davide Grossi (UU)
Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems

2007-17 Theodore Charitos (UU)
Reasoning with Dynamic Networks in Practice

2007-18 Bart Orriens (UvT)
On the development an management of adaptive business collaborations

2007-19 David Levy (UM)
Intimate relationships with artificial partners

2007-20 Slinger Jansen (UU)
Customer Configuration Updating in a Software Supply Network

2007-21 Karianne Vermaas (UU)
Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005

2007-22 Zlatko Zlatev (UT)
Goal-oriented design of value and process models from patterns

2007-23 Peter Barna (TUE)
Specification of Application Logic in Web Information Systems

2007-24 Georgina Ramírez Camps (CWI)
Structural Features in XML Retrieval

2007-25 Joost Schalken (VU)
Empirical Investigations in Software Process Improvement

2008

2008-01 Katalin Boer-Sorbán (EUR)
Agent-Based Simulation of Financial Markets: A modular,continuous-time approach

2008-02 Alexei Sharpanskykh (VU)
On Computer-Aided Methods for Modeling and Analysis of Organizations

2008-03 Vera Hollink (UVA)
Optimizing hierarchical menus: a usage-based approach

2008-04 Ander de Keijzer (UT)
Management of Uncertain Data - towards unattended integration

2008-05 Bela Mutschler (UT)
Modeling and simulating causal dependencies on process-aware information systems from a cost perspective

2008-06 Arjen Hommersom (RUN)
On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective

2008-07 Peter van Rosmalen (OU)
Supporting the tutor in the design and support of adaptive e-learning

2008-08 Janneke Bolt (UU)
Bayesian Networks: Aspects of Approximate Inference

2008-09 Christof van Nimwegen (UU)
The paradox of the guided user: assistance can be counter-effective

2008-10 Wauter Bosma (UT)
Discourse oriented summarization

2008-11 Vera Kartseva (VU)
Designing Controls for Network Organizations: A Value-Based Approach

2008-12 Jozsef Farkas (RUN)
A Semiotically Oriented Cognitive Model of Knowledge Representation

2008-13 Caterina Carraciolo (UVA)
Topic Driven Access to Scientific Handbooks

2008-14 Arthur van Bunningen (UT)
Context-Aware Querying; Better Answers with Less Effort

2008-15 Martijn van Otterlo (UT)
The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains

2008-16 Henriette van Vugt (VU)
Embodied agents from a user's perspective

2008-17 Martin Op 't Land (TUD)
Applying Architecture and Ontology to the Splitting and Allying of Enterprises

2008-18 Guido de Croon (UM)
Adaptive Active Vision

2008-19 Henning Rode (UT)
From Document to Entity Retrieval: Improving Precision and Performance of Focused Text Search

2008-20 Rex Arendsen (UVA)
Geen bericht, goed bericht. Een onderzoek naar de effecten van de introductie van elektronisch berichtenverkeer met de overheid op de administratieve lasten van bedrijven

2008-21 Krisztian Balog (UVA)
People Search in the Enterprise

2008-22 Henk Koning (UU)
Communication of IT-Architecture

2008-23 Stefan Visscher (UU)
Bayesian network models for the management of ventilator-associated pneumonia

Dankwoord

Na al het wetenschappelijke, nu ook tijd voor een meer luchtig maar zeker niet minder belangrijk deel. Hier sta ik even stil bij de totstandkoming van het proefschrift en de mensen bedanken die hier op wat voor manier dan ook aan hebben bijgedragen.

Mijn promotor, prof.dr. M.J.M. Bonten. Beste Marc, op mijn eerste werkdag bleek je poli te hebben en dropte je me bij op 'de vrouwenvleugel'. Daar bleek ik in goede handen, maar daarover later meer. Je hebt me wegwijs gemaakt in de wondere wereld van de geneeskunde en was nooit te druk om even met me te overleggen. Over het algemeen waren er merkbaar weinig irritaties over en weer, hoewel ik van mijn kant herinner dat we eens weken over 1 tabel hebben gebakkeleid; we zaten toen duidelijk niet op een lijn. Toen ik een keer een dag vrij nam en er toch nog een abstract voor de ICAAC geschreven moest worden, mailde ik je "...als jij dit na controle nu even voor mij submit..." en ik meteen terug op mijn plaats gezet werd door een in hoofdletters: "HET IS JOUW ABSTRACT!!". Je hebt een heldere blik; zo kon je ook nuttig commentaar geven op mijn informatica papers. Ik heb veel van je geleerd (en ik hoop andersom ook), waarvoor ik je wil bedanken. Ik zal voor de toekomst onthouden dat een congresposter altijd een zekere 'feminine touch' behoeft.

Dr. P.J.F. Lucas, mijn eerste co-promotor. Beste Peter, jij bent de drijvende kracht achter het informatica onderzoek beschreven in dit proefschrift. Ik had in 1996 in de krant gelezen dat de studie medisch-technische informatica zou starten in Utrecht. Informatica toegepast in de gezondheidszorg leek voor mij de perfecte combinatie. Daartoe bezocht ik de opendag van de universiteit in het seizoen 1997/1998 met mijn vader en ontmoette je daar. Later in 2000 volgde ik met veel plezier het vak expertsystemen bij je, terwijl je al deels in Schotland werkte. Blijkbaar waren jullie toen al gestart met het construeren van het huidige onderzoek naar een computersysteem voor VAP. Pas in 2003 kwam ik toen weer met je in contact en begon – nadat ik voor de ingang van de 'verkeerde' Broese had staan te wachten – in januari 2004 aan mijn promotieonderzoek. We hebben veel papers geschreven en hadden daarover veel

contact; ik denk dat je mailbox wel eens verzuchtte of het aantal mails van mij per dag niet iets minder kon. Ik heb erg prettige werkbesprekingen bij je thuis gehad; onder het genot van een door jou persoonlijk gemaakte typisch Italiaanse lunch. Op de achtergrond klonk dan vaak prachtige muziek van een Italiaanse componist als Scarlatti en waande ik me wel eens in Italië, een klavecimbelconcert bijwonend. Ook van jou heb ik veel geleerd; bedankt daarvoor.

Dr. C.A.M. Schurink, mijn tweede co-promotor. Beste Karin, toen ik eind 2003 solliciteerde en je ontmoette, was je nog steeds enthousiast over het onderzoek, ondanks dat je later liet doorschemeren dat de eindfase van je eigen proefschrift 'de slechtste fase in je leven' was geweest. Ondanks dat je altijd druk was met patientenzorg, wist ik je toch vaak te strikken voor overleg ("Ik moet nog even een patiënt zien, maar daarna heb ik tijd voor je"), ook al was dit pas buiten de reguliere werktijden. Je hielp me veel met het medisch-inhoudelijke, ook voor artikelen. Je bent altijd aardig, laagdrempelig en geduldig. Geduld met computers heb je echter niet en als ik weer eens kwam met de vraag of je toevallig database X nog ergens op je computer had staan, leek dat soms een onmogelijke exercitie. Toen je je overstap naar Rotterdam maakte, heb ik je aanwezigheid en winegum-pitstops nog lange tijd gemist. Gelukkig kon ik je per telefoon of email altijd bereiken. Bedankt voor alles. Hopelijk hebben we nog eens (werk)overleg terwijl jij op een zonovergoten terras in Frankrijk zit.

Prof.dr. Hoepelman, beste Andy, zonder jou als hoofd en alle medewerkers van de afdeling infectieziekten had dit proefschrift niet tot stand kunnen komen. Bedankt.
Mijn collega AiO's van het TimeBayes project, Theodore Charitos en
Rasa Jurgelénaité, wil ik bedanken voor de prettige samenwerking.
NWO wil ik bedanken dat de mogelijkheid werd geboden me te ontplooien op wetenschappelijk gebied, onder anderen door internationale congressen te bezoeken en zo ook vakidioten te kunnen ontmoeten. Zo heb ik op congressen in Schotland, Tsjechië, Portugal en de VS mijn werk gepresenteerd.

Lisa Kruisheer, Marianne Melsen en Ildikó Flesch wil ik voor hun werk en coauteurschap bedanken.
Ook wil ik Jacq Berk en Hans van der Brugge bedanken voor hun tomeloze support als het ging om al mijn vragen inzake de Sybase server en de historie van 'het systeem'.
Jeroen van Wolffelaar heeft een grote rol gespeeld in het ontwerpen van de PHP module en de user-interface. Daarnaast had ik via zijn prive computer, die als server fungeerde, toegang tot mijn eigen werkcomputer in het UMC. Via een zg. tunnelverbinding werd de firewall van het UMC simpelweg omzeild. Het aantal uren dat ik remote ben ingelogd geweest is ontelbaar. Bedankt!
Hetty Blok, dank voor je hulp met het samenstellen en opzoeken van alle kweekdata.

Dan terugkomend op de vrouwenvleugel: Irene, Marianne en Fieke, de vaste harde kern, regelmatig versterkt door Ilja, Anne Marie, Jan Jelrik, Emily en Anke (en voorheen ook Saskia, Ruby en Mirelle). Altijd tijd voor koffie en een kletspraatje; met recht ook wel kletspraat te noemen, aangezien een koffiemoment (met of zonder opgeklopte melk) regelmatig uitliep. We hadden serieuze, maar voornamelijk ook minder serieuze gesprekken. Jullie zijn de beste collega's die je je maar kunt wensen.
Verder niet te vergeten secretaresses Els den Tex en Jeannette Westerbeek.

Mijn paranimfen. Jan Jelrik, mijn steun en toeverlaat in (en ook buiten) het UMC en Frank, al jaren m'n maatje. Bedankt.

Mijn vriendenclub de PartyClub bedankt voor welkome
ontspanningsmomenten en zinnige gesprekken over het doen van promotieonderzoek.

Mijn ouders en broers bedankt voor het vrijwillig aanhoren van mijn promotieperikelen.

Doordat Marc verbonden is aan de afdeling Medische Microbiologie (het voormalige EWI) mocht ik ook de 'Koepel' bezoeken; een jaarlijks terugkerend evenement waarbij het wetenschappelijke met het aangename wordt gecombineerd. De collega's van deze afdeling staan bekend als een gezellig stel mensen die wel van een feestje houden (zo weet ik ook uit ervaring van een congres in San Francisco). Op die afdeling van het UMC ontmoette ik jou dan ook, **lieve Stephanie**. We hadden elkaar al wel eens gesproken, maar veel verder dan een kort "Hoi" op de gang kwam het niet. Tot de Koepel in 2007, waar we volgens de geruchten samen een leuke afterparty zouden hebben gehad. Wij wisten zelf echter van niets. Na een tijdje gingen we samenwonen, terwijl ik mijn proefschrift af aan het maken was (terwijl jij ver weg op vakantie was, of gewoonweg allang lag te slapen). Je toonde veel geduld. Gelukkig is het nu af, zodat we eindelijk tijd hebben: tijd voor elkaar, voor vakantie samen en . . . !

Curriculum Vitae

Stefan Visscher werd geboren op 31 maart 1980 te Utrecht. In 1998 behaalde hij het VWO diploma aan het Oosterlicht College te Nieuwegein. Vervolgens studeerde hij Informatica aan de faculteit Bètawetenschappen van de Universiteit Utrecht (UU) en koos voor de specialisatie Medisch-Technische Informatica met als subspecialisatie Medische Beslissingsondersteunende Systemen. Medische vakken als anatomie, fysiologie en pathologie werden gevolgd aan de faculteit Geneeskunde. Na afstudeeronderzoek te hebben gedaan in samenwerking met het Universitair Medisch Centrum Utrecht (UMCU) onder supervisie van mw. prof.dr.ir. L.C. van der Gaag (UU, Faculteit Bètawetenschappen), mw. prof.dr. Y. van der Graaf (UMCU, Julius Centrum) en dr. F.L.J. Visseren (UMCU) met als titel *Validating clinicians against an automated version of the Hyperlipidaemia Consensus*, slaagde hij voor het doctoraal examen in september 2003. In januari 2004 werd gestart met het onderzoek beschreven in dit proefschrift als onderdeel van het NWO-gesubsidieerde TimeBayes project onder supervisie van promotor prof.dr. M.J.M. Bonten (UMCU) en co-promotores dr. P.J.F. Lucas (Faculteit Informatica en Informatiekunde, Radboud Universiteit Nijmegen) en mw. dr. C.A.M. Schurink (UMCU). Per 1 juni 2008 is hij werkzaam als onderzoeker bij het Nederlands instituut voor onderzoek van de gezondheidszorg (NIVEL), waar hij werkt aan het landelijk informatienetwerk huisartsenzorg (LINH) project.