A Dynamic Bayesian Network for Diagnosing Ventilator-Associated Pneumonia in ICU Patients

Theodore Charitos, Linda C. van der Gaag

Stefan Visscher, Karin A.M. Schurink

Peter J.F. Lucas

Department of Information and Computing Sciences,
Utrecht University

Technical Report UU-CS-2007-013

www.cs.uu.nl

ISSN: 0924-3275

A Dynamic Bayesian Network for Diagnosing Ventilator-Associated Pneumonia in ICU Patients

Theodore Charitos[1], Linda C. van der Gaag[1], Stefan Visscher[2], Karin Schurink[2], and Peter Lucas[3]

[1] Dept. of Inform. and Comp. Sciences, Utrecht University,The Netherlands
[2] Dept. of Internal Medic. and Infect. Diseases, University Medical Center Utrecht, The Netherlands
[3] Inst. for Comp. and Inform. Sciences, Radboud University, Nijmegen, The Netherlands

Abstract

Diagnosing ventilator-associated pneumonia in mechanically ventilated patients in intensive care units is seen as a clinical challenge. The difficulty in diagnosing ventilator-associated pneumonia stems from the lack of a simple yet accurate diagnostic test. To assist clinicians in diagnosing and treating patients with pneumonia, a decision-theoretic network had been designed with the help of domain experts. A major limitation of this network is that it does not represent pneumonia as a dynamic process that evolves over time. In this paper, we construct a dynamic Bayesian network that explicitly captures the development of the disease over time. We discuss how probability elicitation from domain experts served to quantify the dynamics involved and how the nature of the patient data helps reduce the computational burden of inference. We evaluate the diagnostic performance of our dynamic model for a number of real patients and report promising results.

1 Introduction

Many patients admitted to an intensive care unit (ICU) need respiratory support by a mechanical ventilator; in addition, many of these patients are affected by severe disease which may result in depression of their immune system. Both conditions promote the development of ventilator-associated pneumonia (VAP) in these patients. Because of the wide-spread dissemination of multiresistant bacteria at the ICU, effective and fast treatment of VAP is seen as an issue of major significance. The difficulty of the diagnosis of VAP is in the lack of a gold standard; VAP is therefore diagnosed by taking a number of clinical features into account [13]. To support ICU clinicians in diagnosing and treating VAP, a probabilistic and decision-theoretic network, representing the uncertainties and preferences involved, was constructed by Lucas et al. [8]. The network was developed with the help of two infectious disease experts, who assessed both its qualitative structure and its numerical part. The goal of the network was to prescribe an optimal antimicrobial therapy for treating patients with VAP.

Two stochastic processes play a prominent role in the domain of pneumonia: the *colonisation* of the laryngotracheobronchial tree by pathogens and the onset and development of *pneumonia*. Although both processes evolve dynamically, these dynamics were not explicitly modelled by means of temporal transitions in the network of Lucas et al. Instead, the dynamics of the processes were modelled implicitly by additional interactions between the duration of hospital stay and the duration of mechanical ventilation of a patient with the colonisation by pathogens. The main motivation for this simplification was the large amount of data needed to specify the probability distribution underlying the stochastic processes and the increase in computational requirements. The network of Lucas et al. thus constitutes a *static* simplification of the domain. The static network was used for every patient for each day on the ICU separately, without taking into account the patient's characteristics from earlier days. Consequently, its diagnostic performance was suboptimal and even confusing for patients without VAP. As the development of VAP is a dynamic process, we feel that time needs to be modelled in a more explicit way to improve the diagnosis.

In this paper, we alleviate the problems associated with the static representation of the domain by modelling VAP as a dynamic process. More specifically, we develop a dynamic Bayesian network (DBN) that explicitly captures the temporal relationships between the variables [9]; our focus thereby is initially on the diagnostic part of the network. We use the method of Van der Gaag et al. [14, 15] for the elicitation, from domain experts, of the probability distribution of the underlying stochastic process. This method transcribes probabilities and uses a scale with both numerical and verbal anchors that allows experts to assess many probabilities in little time. Moreover, we discuss how the computational burden of inference with our model can be eased by exploiting the nature of the observations involved and the properties of the transitional relationships of the model with just a small loss in accuracy. We evaluated our dynamic network on a group of patients, drawn from the files of the ICU of the University Medical Center Utrecht in the Netherlands. Our results indicate that the dynamic model is capable of distinguishing between patients with VAP and without VAP. By exploiting all available past information of a patient, it in fact yields at least as good or even better predictions than the static model. Specifically for patients without VAP, we noticed that the use of previous information leads to lower estimates for VAP than the ones obtained from the static network.

The paper is organised as follows. In Section 2, we briefly describe the static decision-theoretic network that had been developed before for the management of VAP. In Section 3, we discuss the construction of a dynamic network for VAP and present computational methods for performing efficient inference with the model. In Section 4 we present the results of an experimental evaluation of our network. Conclusions and directions for further research are given in Section 5.

2 A static network for VAP

Ventilator-associated pneumonia is a low-prevalence disease occurring in mechanically-ventilated patients in critical care and involves infection of the lower respiratory tract [2]. In contrast to infections of more frequently involved organs (such as the urinary tract), for which mortality is low, ranging from 1 to 4%, the mortality rate for VAP ranges from 24 to 50% and can reach 76% for some high-risk pathogens. VAP therefore has been associated with increased morbidity, attributable mortality and increased health care costs. Important causes related to the development of VAP include the duration of *hospitalisation* and of *mechanical ventilation* of the patient; important symptoms that indicate the presence of VAP include an increased *body temperature*, an abnormal amount of coloured *sputum*, *signs* on the chest X-ray, an abnormal ratio between the amount of oxygen in the arterial blood and the fractional inspired oxygen concentration, that is, pO_2/FiO_2, and an abnormal number of *leukocytes*.

As diagnosing VAP and deciding upon treatment can be a hard task for clinicians, a decision-theoretic network had been constructed as part of a decision-support system to assist clinicians in their task in the ICU [8, 13]. Figure 1(left) illustrates the global structure of the network, which we call the static VAP network, or sVAP network for short. Dashed arcs denote temporal probabilistic relationships; solid arcs represent stochastic dependency without a special temporal meaning. Boxes in the figure indicate collections of stochastic variables, where the collection of therapy variable is shown by thick lines; ellipses indicate single stochastic variables. Clear shapes refer to hidden variables, while shaded shapes mark observable ones. As an example, colonisation by pathogens is modelled as a biological process, in which it is assumed that colonisation by different pathogens occurs independently. The relationship between the *colonisation* by different pathogens and the development of *pneumonia* is captured in the sVAP as shown in Figure 2. The seven groups of microorganisms that appear most frequently in critically ill patients and cause colonisation, are modelled in the diagnostic part of the network. Only a small percentage of pathogens colonising a patient can cause an actual infection. Therefore, there exists a relation in the network between colonisation and pneumonia. The figure now depicts the probabilistic relation between the seven groups of microorganisms from colonisation to pneumonia. Information about which bacterium or bacteria are currently present in a patient in combination with the current signs and symptoms constitute the basis for choosing optimal antimicrobial treatment on resistant bacteria and is considered best practice. The signs and symptoms included in the sVAP network are shown in more detail in Figure 1(right). In the sVAP network, the temporal nature of the processes is expressed by the interaction between the duration of the stay (*hospitalisation*) at the ICU and the duration of the mechanical ventilation: both the duration

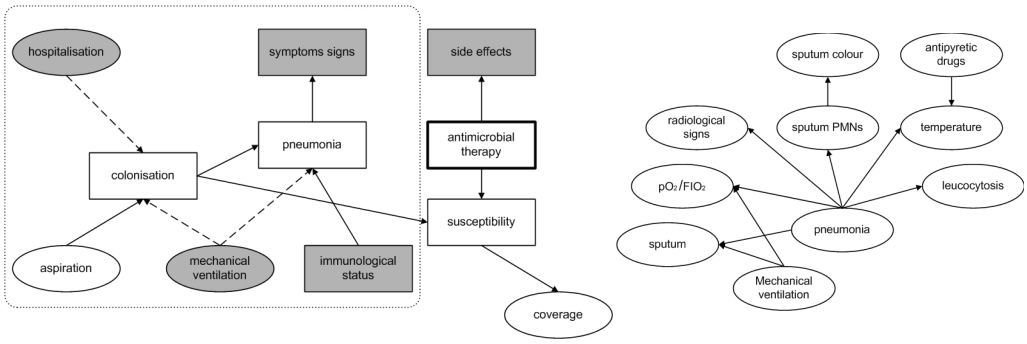

Figure 1: (left) Global structure of the sVAP network. The dashed box indicates the network's diagnostic part. (right) Symptoms and signs of pneumonia.

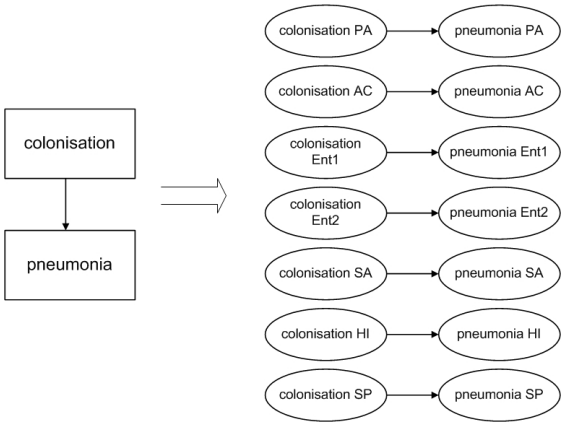

Figure 2: Detailed structure of the influence of colonisation on pneumonia. Abbreviations: PA: Pseudomonas aeruginosa, AC: Acinetobacter, Ent1: Enterobacteriaceae1, Ent2: Enterobacteriaceae2, SA: Staphylococcus aureus, HI: Haemophilus influenzae, SP: Streptococcus pneumoniae.

of the stay and the duration of the ventilation are correlated to the process of colonisation by pathogens. Hence, time is modelled implicitly by these two variables; for example, the mechanical ventilation variable can take one of the six values $\{0, 0-24, 24-48, 48-96, 96-144, >144\}$ which indicate the number of hours that the patient has been mechanically ventilated. The model thus hides the temporal nature of the development of the processes of colonisation and pneumonia in conditioning variables, instead of handling time explicitly.

In the present sVAP network, no history is captured and possible changes in a patient's condition cannot be taken into consideration. Since the network constitutes a rough representation of time, only rough estimates can be obtained upon diagnostic evaluation. For a patient without VAP for instance, a positive symptom observed at a specific day can increase the probability of VAP significantly even though on the previous days only negative symptoms were observed. A fine-grained and meticulous representation of the processes underlying the development of pneumonia can considerably improve the diagnostic performance of the network, as will be demonstrated in the next section.

3 A dynamic network for VAP

In this section, we describe the construction of a DBN that explicitly represents the development of pneumonia in mechanically ventilated patients.

3.1 Preliminaries

A DBN is a graphical model that encodes a joint probability distribution on a set of stochastic variables, explicitly capturing the temporal relationships between them. More formally, let $\mathcal{V}_n = (V_n^1, \ldots, V_n^m)$, $m \geq 1$, denote the set of variables at time n. Then, a DBN is a tuple (B_1, B_2), where B_1 is a Bayesian network that represents the prior distribution for the variables at the first time slice $\mathcal{V}_1$, and B_2 defines the transitional relationships between the variables in two consecutive time slices, so that for every $n \geq 2$

$$p(\mathcal{V}_n \mid \mathcal{V}_{n-1}) = \prod_{i=1}^{m} p(V_n^i \mid \pi(V_n^i))$$

where $\pi(V_n^i)$ denotes the set of parents of V_n^i, for $i = 1, \ldots, m$.

We distinguish between two types of relationship in a DBN: *transitional* relations that capture a dependence among variables between different time slices, and *local* relations that capture a dependence between variables within the same time slice. If a relationship exists between the same variable over different time slices, this variable is called *persistent*. Based on the two types of relationship, per time slice, the set of variables $\mathbf{V}_n$ is split into three mutually exclusive and collectively exhaustive sets $\mathbf{I}_n, \mathbf{X}_n, \mathbf{Y}_n$, where the sets $\mathbf{I}_n, \mathbf{Y}_n$ constitute the input and observable variables and $\mathbf{X}_n$ consists of the hidden variables for the time slice under study. Usually, $\mathbf{I}_n$ includes observable variables that affect the probability distribution of $\mathbf{X}_n$, while $\mathbf{Y}_n$ includes observable variables whose probability distribution is affected by $\mathbf{X}_n$. The set $\mathbf{X}_n$ includes the variables that represent the stochastic processes of the network and whose values are never observed. Later in the paper, we will need the notion of *forward interface* of a dynamic network, which is the set of variables at time n that affect some variables at time $n + 1$.

DBNs are usually assumed to be time invariant, which means that the topology and the parameters of the network per time slice and across time slices do not change. Moreover, the Markov property for transitional dependence is assumed, which means that $\pi(V_n^i)$ can include variables either from the same time n or from the previous time $n-1$, but not from earlier time slices [9]. Then, by unrolling B_2 for N time slices, a joint probability distribution $p(\mathbf{V}_1, \ldots, \mathbf{V}_N)$ is defined for which the following decomposition property holds:

$$p(\mathbf{V}_1, \ldots, \mathbf{V}_N) = \prod_{n=1}^{N} \prod_{i=1}^{m} p(V_n^i \mid \pi(V_n^i))$$

Applying a DBN usually amounts to computing the marginal probability distributions of the hidden variables at different times. The computations involved constitute the *inference*. Three types of inference are distinguished. *Monitoring* is the task of computing the probability distribution for $\mathbf{X}_n$ at time n given the observations that are available up to and including time n. *Smoothing* is the task of computing the marginal probability distribution for $\mathbf{X}_n$ at time n given the observations available up to time N where $N > n$. Finally, *forecasting* is the task of predicting the probability distribution of $\mathbf{X}_n$ at time n given the observations that are available about the past up to time N where $N < n$.

For this purpose, we use the *interface algorithm* with the dVAP network [9]. The interface algorithm is an extension of the *junction-tree algorithm* for inference with Bayesian networks in general [5], efficiently exploiting the forward interface of a dynamic network. The complexity of the algorithm is exponential in the number of variables belonging to the forward interface and can thus be infeasible for large networks. Furthermore, the algorithm is linear in the total number of time slices and for large time scopes, the computation time can prove to be prohibitive for practical purposes.

3.2 Constructing the dynamic network

A natural extension of the diagnostic part of the sVAP network is a network that represents time explicitly [8]. Figure 3 gives an overview of the structure of the dynamic network that we constructed for the diagnosis of VAP, which we call the dVAP network. The major difference with the sVAP network is the explicit representation of two processes that evolve over time. The dVAP network includes two interacting dynamic hidden processes, modelled by the compound variables *colonisation* and *pneumonia*. There are no transitional influences between these variables, but both are persistent and hence belong to the forward interface of dVAP. The process of colonisation is influenced by three input variables, *hospitalisation*,

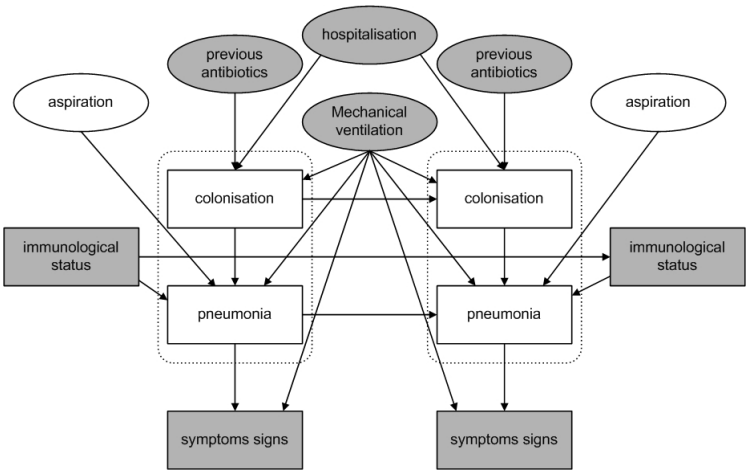

Figure 3: The dVAP network for the diagnosis of VAP; clear nodes are hidden, shaded nodes are observable. The dashed boxes indicate the hidden processes of the network.

Input variables $\mathbf{I}_n$	Hidden variables $\mathbf{X}_n$	Observable variables $\mathbf{Y}_n$
hospitalisation	aspiration	**symptoms-signs** (8)
mechanical ventilation	**colonisation*** (7)	
previous antibiotics	**pneumonia*** (8)	
immunological status* (3)		

Table 1: The variables and compound variables (boldface) per time slice, with their number of variables included in parenthesis, of the dVAP network. The variables marked with an asterisk $*$ belong to the forward interface.

mechanical ventilation and *previous antibiotics*, which in essence control its dynamics. The process of pneumonia is influenced by the hidden yet not persistent variable *aspiration* and by the input compound variable *immunological status* that is persistent and represents the current condition of the patient. We note that both the variables *hospitalisation* and *mechanical ventilation* are observed for a period that is longer than the transition interval of the model. The variables thus are modelled as affecting adjacent time slices. The variable *previous antibiotics* is an additional variable with respect to the sVAP network and represents the effect of previous medication to the patient on the process of colonisation. Finally, similarly to the sVAP network, the compound variable *symptoms-signs* represents the observable variables whose values influence probabilistically the two hidden processes.

The model includes 30 variables per time slice, 6 of which are input variables, 16 are hidden variables and 8 are observable variables; the model thus includes one additional variable (previous antibiotics) per time slice in comparison to the diagnostic part of the sVAP network. Table 1 gives and overview of the variables. The number of values per variable ranges between two and thirty, with an average of 3.2. The number of incoming arcs per variable ranges between zero and eight with an average of 2.6. In total, the model includes 1637 parameter probabilities, 1044 of which concern the transitional relationships.

One of the first difficulties in constructing the dVAP network was to define the length of the transition interval. It may seem trivial in general to decide upon an interval length, but in our case it proved to be rather difficult since there was no *a-priori* commonly acknowledged interval length that appropriately represents the evolution of the unobserved disease. Also, there was not a standard interval over which observations were collected in our data files. The latter can be attributed to most of the measurements being collected by nurses; for example, observable variables such as *body temperature* and *sputum colour*

Suppose a patient has been mechanically ventilated for 48 hours and now has pneumonia caused by *s.aureus*. If this patient after 24 hours is *not mechanically ventilated*, but is *colonized with s.aureus* and has *phagocyte dysfunction*, then how likely is it that the patient will still have pneumonia caused by s.aureus ?

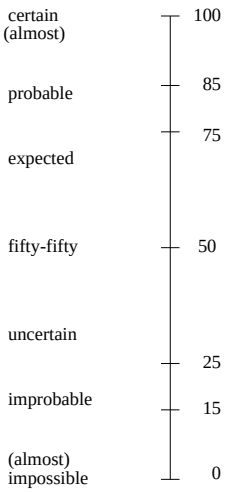

Figure 4: The fragment of text and probability scale for the assessment of a conditional probability.

were measured frequently (approximately every two or three hours), while variables such as *radiological signs* and *leucocytosis* were measured once per day. In cooperation with the expert, we decided to use a transition interval of one day (24 hours) for the dVAP network. Within this interval, the network *aggregates* the observations in a way similar to the previously constructed static network. For each observable variable, the value most frequently observed during the day was chosen as representative for that day; in cases where there was no prevalent value in the data, the worst value observed for the patient was chosen, to allow for *conservative* conclusions from the network. The chosen transition interval appeared to be compatible with the application characteristics and admissible by the domain experts.

A subsequent issue in building the dVAP network was the acquisition of all conditional probabilities required. We recall that the difficulty in acquiring all probabilities involved was one of the reasons that Lucas et al. initially chose to build a static model [8]. Although the three ICUs that acted as a setting for this study used the same shared computer-based patient record system, it appeared very hard to select relevant patient cases from the collected data. The main reason was that VAP is always a concomitant disease. As a consequence, clinicians tend to not report the presence of VAP in a patient. We thus found that only in a very small proportion of cases, a patient was reported as having VAP. Also, for the same reason, hardly any results reported in the literature were usable for our model. Since we could not exploit the data for estimating the probabilities for our network, the single remaining source of probabilistic information was the knowledge and personal clinical experience of the domain expert involved in this study.

Compared to the sVAP network, the new parameters to be assessed for the dVAP network concerned the dynamics of the stochastic processes of colonisation and pneumonia. To estimate those probabilities from the domain expert we used the elicitation method proposed by Van der Gaag et al. [14, 15]. This method is tailored to eliciting a large number of probabilities in a short time. Its main characteristic is the idea of presenting conditional probabilities as fragments of text and of providing a scale for marking assessments with both numerical and verbal anchors; for every conditional probability that needs to be assessed the domain expert is provided with a separate figure with the text and associated scale. Figure 4 shows, as an example, the figure pertaining to the conditional probability

$$p(\mathit{pneum.aureus}{=}\mathit{yes} \,|\, \mathit{pneum.aureus} = \mathit{yes}, \mathit{mech.ventilation} = \mathit{no}, \\ \mathit{colonisation.aureus} = \mathit{yes}, \mathit{phagocytes.dysfunction} = \mathit{yes})$$

for the dVAP network. On the left of the figure is a fragment of text that transcribes the conditional probability to be assessed. Using a fragment of text to denote a probability circumvents the need to use mathematical notation. The fragment is stated in terms of likelihood rather than in terms of frequency to forestall difficulties with the assessment of a conditional probability for which the conditioning text is quite rare. To facilitate the assessment of a required probability, a vertical scale is depicted to the right of the text fragment. Indicated on this scale are various different numerical and verbal anchors. With this method,

we elicited in a few hours from the domain expert the conditional probabilities required for the part of the dVAP network that pertains to the transitional relations of the two hidden processes.

3.3 Computational issues

The practicability of the dVAP network depends to a large extent on the computational burden of inference with the network. For diagnosing patients with VAP, we monitor them at each time slice. In total, there are 17 variables that belong to the forward interface of the model and there are also 17 binary hidden variables per time slice. The runtime complexity of the interface algorithm for exact inference can therefore be quite time consuming if not infeasible. We recall that this problem was one of the reasons why Lucas et al. [8] preferred to use a static model instead of a dynamic one. In our application, however, and in fact in many other applications, the nature of the observations obtained may help reduce the computational requirements involved. More specifically, in case consecutive similar observations are obtained, the probability distribution of the hidden process converges to a limit distribution within a given level of accuracy [4]. After some number of time slices, therefore, there is no need for further inference as long as similar observations are obtained. The phenomenon of consecutive similar observations was particularly evident for several patients in the ICU files. For example, for many patients we found that the same combination of values was observed for all or almost all of the observable variables for a number of consecutive days.

As an example, we consider a patient who has been mechanically ventilated for six days and is observed with a *high* body temperature, an *abnormal* amount of sputum, an *abnormal* ratio pO_2/FiO_2, and a *normal* number of leukocytes. These observations cause the probability of VAP to be at that day $p(\mathit{VAP}_6) = 0.4321$. The same values for these observable variables are obtained for the next three days. According to our model, we find that $p(\mathit{VAP}_7) = 0.5516, p(\mathit{VAP}_8) = 0.7301, p(\mathit{VAP}_9) = 0.8341$. If we continue to obtain similar observations for the following three days we find that $p(\mathit{VAP}_{10}) = 0.8658, p(\mathit{VAP}_{11}) = 0.8734, p(\mathit{VAP}_{12}) = 0.8751$. We notice that the probability distribution for VAP does not change much after a number of time slices and further inference can be forestalled.

Using the *relative entropy* distance measure for distributions, we can show that it suffices to use just the most recent data for monitoring [3]. Based upon this result, we define the *backward acceptable window* $\omega^{\phi}_{n,\epsilon}$for the present time n given a specified level of accuracy ϵ, to be the minimal number of time slices that we need to use from the past to compute the probability distribution of the hidden variable at the present time within the level of accuracy ϵ. The scheme below illustrates the concept of the backward acceptable window :

$$\underbrace{\{1, \ldots, n_\phi, \ldots, n\}}_{\text{total time scope}} \longrightarrow \underbrace{\{n_\phi, \ldots, n\}}_{\omega^{\phi}_{n,\epsilon}}$$

We now perform inference for time n by considering only the backward acceptable window $\omega^{\phi}_{n,\epsilon}$ without losing too much in accuracy. Note that by doing so, we perform inference for $n - n_\phi$ time slices instead of for the n slices that would be taken into consideration by an exact algorithm. In the next section we report promising results from applying the backward acceptable window to speed up inference with our model. The main conclusion from the above considerations is that monitoring in the dVAP network can be eased considerably by exploiting the characteristics of the observations for a patient and by using the backward acceptable window.

4 Diagnostic performance

Monitoring a patient on an ICU ward is performed as follows. The clinician examines the results of diagnostic tests and the symptoms observed during the day and, taking into account the number of days the patient is hospitalised and mechanically ventilated, assess whether or not the patient has VAP. Based on this assessment, the clinician can prescribe antibiotic treatment for a series of days, while continuing to monitor the patient. The primary goal of the dVAP network is to assist the clinician in this process by explicitly considering the history of the patient. The clinician can of course be aware of the past diagnostic observations, but the effect of those observations on the current diagnosis is hard to assess. The dVAP network serves to explicitly model this effect via its transition probabilities.

symptoms	VAP $n = 5$	no VAP $n = 15$
abnormal temperature	60%	7%
mech. ventilation (mean)	10d	10d
abnormal leukocytes	80%	53%
abnormal pO_2/FiO_2	60%	27%
abnormal sputum	80%	73%
coloured sputum	60%	60%
colonised	40%	13%
antipyretic drugs	100%	87%
positive X chest	40%	0%

Table 2: Data summary

We evaluated the performance of the dVAP network, focusing on its diagnostic prediction per day. At our disposal we had a temporal database with data from 2410 patients. Each record contains data collected for a patient during a one day stay in the ICU. The source of these data is the clinical management system used at the Intensive Care Units of the University Medical Center Utrecht in the Netherlands. The conclusions obtained from the dVAP network were examined on a group of 20 patients in total, 5 of which were diagnosed with VAP as established by two infectious-disease specialists. This group of patients was chosen from a total of 487 patients who were admitted for a period of 10 days or longer. For these 5 patients we used the data from the day of admission to the ICU until the day they were diagnosed with VAP, which was confirmed to be at day 10. For each of these 5 patients, we selected from the database three patients for whom it was known that they did not develop VAP over time. These patients were matched on three criteria: gender, number of mechanically ventilated days, and ICU ward. Table 2 summarises the data for the 5 patients with VAP and for the 15 patients without VAP on the tenth day of admission.

To compare the diagnostic performance of the dVAP network to that of the original sVAP network, we used the Brier score best known from the field of statistical forecasting [10, 16]. We illustrate the Brier score for our dVAP network. For each patient i, the network yields a probability distribution p_i over the two values $j = 1, 2$ (yes, no) of VAP. The Brier score B_i for this distribution is defined as

$$B_i = \sum_{j=1,2} (p_{ij} - s_{ij})^2$$

where $s_{ij} = 1$ if the medical record of the patient states the value j, and $s_{ij} = 0$ otherwise. If the network would yield the correct value with certainty for a patient, then the associated Brier score would be equal to 0. Conversely, if the network would yield an incorrect value for a patient, then the associated Brier score would be equal to 2. For the probability distribution computed for any patient, therefore, the Brier score ranges between 0 and 2, and the better the prediction is, the lower the score. The Brier scores for all patients on day 10, for the dVAP and the sVAP networks respectively, are shown in Table 3. We note that for 15 patients of the total of 20, the computed Brier score was lower with the dVAP network than with the sVAP network. The overall quality of the two networks can be expressed in an overall score

$$B = \frac{1}{m} \sum_{i=1,\ldots,m} B_i$$

where m is the number of patients. The overall Brier score for the sVAP network can be readily computed from Table 3 and equals 0.3370, while the overall Brier score for the dVAP network is 0.2376. Although the lower score suggests that the dVAP network is better informed, the number of patients is too small to arrive at valid statistical conclusions concerning which of the two models performs better. One way to gain additional insight into the comparative quality of the two networks is to apply bootstrapping. The bootstrap technique performs sampling with replacement from the original data set to create replicates of this data set [6]. We generated different patient groups by bootstrapping, computed the overall Brier score for each one of these groups for both models and subsequently computed equi-tailed 95% confidence intervals. Using 100 replicates, the confidence interval for the sVAP network was computed to be $[0.026, 0.680]$; for the dVAP network, it was $[0, 0.517]$. This result now demonstrates the ability of the dVAP network to obtain on average lower Brier scores than the sVAP network.

		True class: yes	True class: no
Hypothesized class	yes	True Positives	False Positives
	no	False Negatives	True Negatives

Figure 5: A confusion matrix.

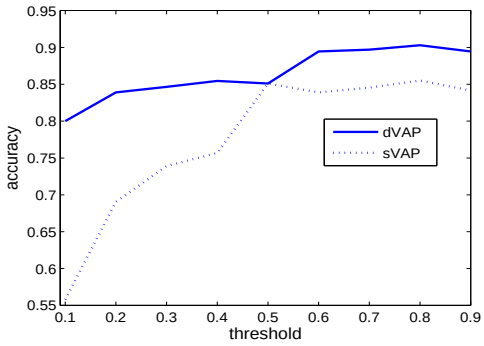

Figure 6: Accuracy vs threshold for the dVAP and sVAP networks.

patient	VAP	sVAP	sBrier	dVAP	dBrier
1	yes	0.9969	$1.9059 \cdot 10^{-5}$	0.9987	$3.3801 \cdot 10^{-6}$
2	no	0.0203	$8.2432 \cdot 10^{-4}$	0.1395	0.0389
3	no	0.1672	0.0559	0.0558	0.0062
4	no	0.0028	$1.5276 \cdot 10^{-5}$	0.0002	$8.0002 \cdot 10^{-8}$
5	yes	0.0097	1.9613	0.0002	1.9992
6	no	0.4309	0.3713	0.0316	0.0019
7	no	0.0203	0.0008	0.0003	$1.8002 \cdot 10^{-7}$
8	no	0.1934	0.0748	0.0309	0.0019
9	yes	0.9999	$3.3620 \cdot 10^{-9}$	0.9987	$3.3801 \cdot 10^{-6}$
10	no	0.0227	0.0010	0.0015	$4.5001 \cdot 10^{-6}$
11	no	0.0457	0.0042	0.0005	$5.0001 \cdot 10^{-7}$
12	no	0.2977	0.1772	0.0325	0.0021
13	yes	0.0348	1.8632	0.0033	1.9868
14	no	0.0203	0.0008	0.0005	$5 \cdot 10^{-7}$
15	no	0.4364	0.3809	0.099	0.0196
16	no	0.0099	$2.2231 \cdot 10^{-5}$	$7.0001 \cdot 10^{-8}$	$9.8001 \cdot 10^{-15}$
17	yes	0.9966	$2.2231 \cdot 10^{-5}$	0.9035	0.0186
18	no	0.1752	0.0614	0.0218	0.0009
19	no	0.0740	0.0109	0.0013	$3.3801 \cdot 10^{-6}$
20	no	0.9421	1.7750	0.5810	0.6751

Table 3: Brier scores for the sVAP network and for the dVAP network, respectively.

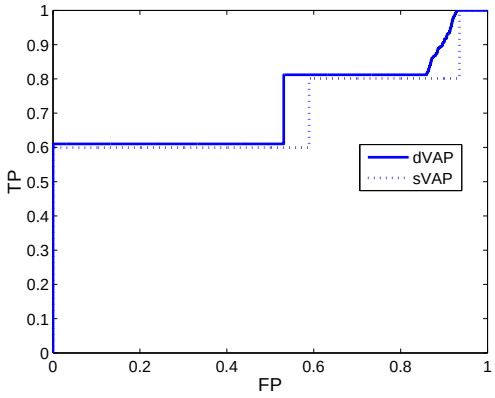

Figure 7: Average ROC curves for the dVAP and sVAP networks.

To further compare the performance of both models, we computed their accuracy in distinguishing between patients with VAP and without VAP, using various threshold probabilities. More formally, if the probability of VAP for a patient is greater than a specific threshold probability, we decide that this patient has VAP; otherwise we decide that this patient does not have VAP. A patient who has been actually diagnosed with VAP and is classified as having VAP, then is called a *true positive* (TP); a patient who has been diagnosed as not having VAP and is also classified as being negative for VAP, then is a *true negative* (TN). Similarly, patients can be classified as *false positives* (FP) or *false negatives* (FN). Based upon the counts TP, TN, FP and FN, a two-by-two confusion matrix can be computed [7] as shown in Figure 5. The accuracy now for each model can be computed as

$$\frac{TP + TN}{TP + FP + TN + FN}$$

Using again 100 replicates and various thresholds, we computed the average accuracy for each threshold for both models. Figure 6 illustrates the results. We observe that the accuracy of the dVAP network is higher than the one of the sVAP network for all thresholds except for threshold 0.5 where both models show the same performance. Finally, using 200 replicates we plotted the average receiver operating characteristics (ROC) curve for both models as shown in Figure 7. These results again support the observation that the dVAP network is more informed than the sVAP network and can arrive at relatively good estimates for diagnosing VAP.

Observing in more detail the results from Table 3, we notice that for the patients $\mathbf{6}, \mathbf{12}, \mathbf{15}, \mathbf{18}$ and $\mathbf{20}$ for example, who were diagnosed not to have VAP, the dVAP network derived low probabilities for the presence of VAP. It arrived at these low probabilities by exploiting all previous information. The sVAP network, in contrast, used just the current information and produced much higher probabilities. For the patients diagnosed with VAP, the two models behave more or less similarly, with the highest absolute discrepancy observed in patient $\mathbf{17}$, to whom the sVAP network assigned a probability of VAP of 0.997 and the dVAP network assigned a probability of VAP of 0.904.

To study the performance of the dVAP network over time, we computed the probability of VAP for each day and compared it to the respective probability established from the sVAP network. In Figure 8 we plot, for two separate groups of four related patients, the probability of VAP for patient $\mathbf{9}$ and the mean probability of VAP for the matched patients $\mathbf{10}, \mathbf{11}, \mathbf{12}$ (left), and for patient with VAP $\mathbf{17}$ and matched patients $\mathbf{18}, \mathbf{19}, \mathbf{20}$ (right), from both networks. We observe in the first group that for the patient with VAP the trend in both networks is more or less the same for all days considered. However, for the second group the trend is similar only after the fifth time slice; in fact, we notice that the dVAP network assigns a low probability of VAP for the first five time slices for patient $\mathbf{17}$ and then assigns a higher probability of VAP confirming the diagnosis by the expert. In contrast, the sVAP network assigns a high probability of VAP throughout the whole period of ten days. Although it may be argued that for this patient the sVAP

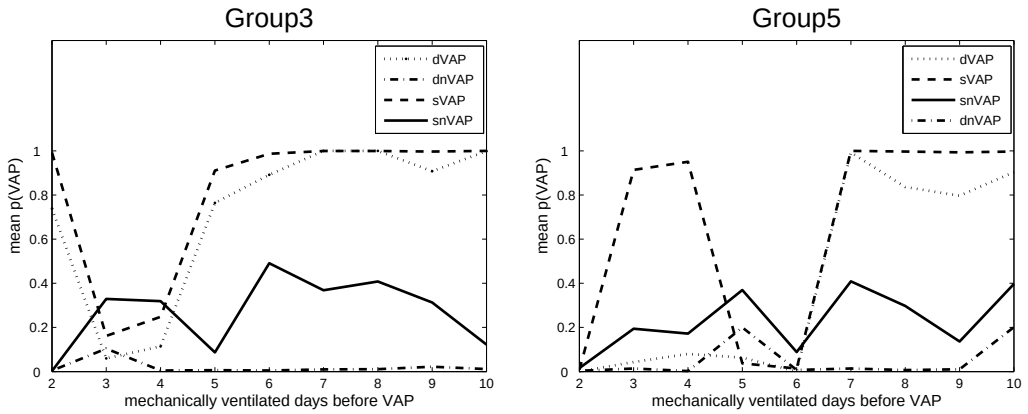

Figure 8: The dVAP and sVAP performance over time for two groups of matched patients; dnVAP and snVAP represent the average performance for the three patients without VAP.

patient	**9**	**10**	**11**	**12**
exact	0.9987	0.0015	0.0005	0.0325
$\omega^{\phi}_{10,0.003}$	0.9987	0.0013	0.0005	0.0347

Table 4: Exact and approximate probabilities for VAP for a group of matched patients.

network outperforms the dVAP network, a plausible explanation is as follows. This specific patient was diagnosed by the expert as having VAP at the tenth day of observation. The findings for the first days do not advocate the presence of VAP, which was developed later and finally diagnosed. We see that the dVAP actually confirms this scenario by assigning a low probability of VAP in the first days; in contrast, the sVAP network performed wrongly for these time slices. Concerning the patients without VAP, we note that the dVAP network assigns consistently lower probabilities than the sVAP network for both groups.

A preliminary conclusion from our experiments is that the dVAP is better able to distinguish between VAP and non-VAP patients. In the dVAP network, the transition model carries the information of the patient from previous time slices into the current time slice, where diagnosis is performed. For a patient who has constantly been monitored throughout time not to have VAP, the dVAP network will reduce the effect of new positive findings on the current diagnosis. On the other hand, if a patient has been monitored to have VAP with high probability, then he/she will continue to do so but with lower probability even if new observations are negative. The dynamic nature of the model is thus responsible for conveying the history of the patient to the present and thereby increases the diagnostic performance of the model. In the sVAP network, in contrast, diagnosis is merely performed on the current observations of the patient disregarding past knowledge from previous time slices.

To conclude, we performed the computations in the dVAP network using different values for the backward acceptable window $\omega^{\phi}_{n,\epsilon}$. For a particular group of matched patients, the computed exact and approximate probabilities of VAP are shown in Table 4. We conclude that instead of using the observations for all 10 days in the ICU to compute the probability of VAP, we can use the observations for just the last 5 days with an average error for all patients smaller than $\epsilon = 0.003$. We can thus use this backward acceptable window to decrease the computational burden involved in inference and obtain results with an almost negligible error.

5 Discussion

In this paper, we discussed the construction of a probabilistic model that is aimed at assisting ICU clinicians in diagnosing ventilator-associated pneumonia. In contrast to previous approaches that used a static

decision-theoretic network for this low-prevalence disease [8, 13], we focused on its dynamic evolution and used a dynamic Bayesian network as the primary tool for representation and inference. We detailed various modelling steps in the construction of our dynamic network and described the use of an efficient procedure for expert elicitation of the probabilities required. We further argued that a number of convergence properties of dynamic Bayesian networks can be exploited to arrive at feasible algorithms that restrict the computational burden of inference with such a model. In this way, we ameliorated two important problems that were considered impervious in the past [8]: the specification of the probabilities underlying the stochastic process modelled in the network and the computational burden of inference.

In the past, a number of dynamic models have been developed for medical applications, that accord with the basic theoretical framework underlying the dVAP network. Examples of such models include a dynamic network for insulin adjustment by Andreassen et al. [1], an influence diagram for diagnosis and treatment of acute abdominal pain by Provan [12], and a decision-theoretic network for therapy planning in the domain of paediatric cardiology by Peek [11]. The referenced articles focus primarily on the structure of the model and its performance but do not address in detail such issues as the determination of the transition interval to be used, the estimation of the transition probabilities, and the development of algorithms for inference that exploit various characteristics of the model at hand. In developing the dVAP network, we found that these issues also needed careful attention.

We evaluated our dVAP network on a set of ICU patients to examine its diagnostic performance. The lower overall Brier score of the dynamic network in comparison to the static one, indicated that representing time explicitly and taking into consideration the history of the patient, increases diagnostic performance. In our experiments, the dynamic network proved to exhibit better performance at distinguishing between VAP and non-VAP patients than the static network, especially by assigning lower probabilities of VAP to the non-VAP patients. Future research includes improvement of the dVAP network by use of the available data for parameter learning of the conditional tables for the observable variables of the model, and an extensive evaluation study using data from more patients. The overall aim is to ultimately embed it in the clinical information system of the ICU.

References

[1] S. Andreassen, R. Hovorka, J. Benn, K.G. Olesen, and E.R. Carson. A model-based approach to insulin adjustment. *Proceedings of the 3rd Conference on Artificial Intelligence in Medicine*, pp. 239-248, 1991.

[2] M.J. Bonten. Prevention of infection in the intensive care unit. *Current Opinion in Critical Care*, 10(5):364-368, 2004.

[3] T. Charitos and L.C. van der Gaag. Sensitivity analysis of Markovian models. *Proceedings of the 19th International Florida Artificial Intelligence Research Society Conference*, pp. 806-811, 2006.

[4] T. Charitos, P. de Waal, and L.C. van der Gaag. Convergence in Markovian models with implications for efficiency of inference. *International Journal of Approximate Reasoning*, 2007 (in press).

[5] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, and D.J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer, 1999.

[6] B. Efron and R.J. Tibshirani. *An introduction to the Bootstrap*. Chapman and Hall: New York, 1993.

[7] T. Fawcett. *ROC graphs: Notes and practical considerations for data mining researchers*. Technical report, HP Laboratories, 2004.

[8] P.J.F. Lucas, N.C. de Bruijn, C. Schurink, and A. Hoepelman. A probabilistic and decision theoretic approach to the management of infectious disease at the ICU. *Artificial Intelligence in Medicine* 19(3): 251-279, 2000.

[9] K. Murphy. *Dynamic Bayesian networks: Representation, Inference and Learning*. PhD diss., University of California Berkeley, 2002.

[10] H.A. Panofsky and G.W. Brier. *Some applications of statistics to meteorology*. The Pennsylvania State University, University Park, Pennsylvania, 1968.

[11] N.B. Peek. Explicit temporal models for decision-theoretic planning of clinical management. *Artificial Intelligence in Medicine* 15(2): 135-154, 1999.

[12] G.M. Provan. Tradeoffs in constructing and evaluating temporal influence diagrams. *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pp. 40-47, 1993.

[13] C.A.M. Schurink. *Ventilator Associated Pneumonia: a Diagnostic Challenge*. Ph.D. diss, Utrecht University, 2003.

[14] L.C. van der Gaag, S. Renooij, C.L.M. Witteman, B. Aleman, and B.G. Taal. How to elicit many probabilities. *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, pp. 647-654, 1999.

[15] L.C. van der Gaag, S. Renooij, C.L.M. Witteman, B.M.P. Aleman and B.G. Taal. Probabilities for a probabilistic network: a case study in oesophageal cancer. *Artificial Intelligence in Medicine* 25: 123148, 2002.

[16] L.C. van der Gaag and S. Renooij. Probabilistic networks as probabilistic forecasters. *Proceedings of the 9th Conference on Artificial Intelligence in Medicine in Europe* pp. 294-298, 2003.