Using hierarchical dynamic Bayesian networks to investigate dynamics of organ failure in patients in the Intensive Care Unit

Linda Peelen [a,d,*], Nicolette F. de Keizer [a], Evert de Jonge [b], Robert-Jan Bosman [c], Ameen Abu-Hanna [a], Niels Peek [a]

[a] Department of Medical Informatics, Academic Medical Center, Amsterdam, The Netherlands
[b] Department of Intensive Care Medicine, Academic Medical Center, Amsterdam, The Netherlands
[c] Department of Intensive Care Medicine, Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands
[d] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands

ARTICLE INFO

ABSTRACT

In intensive care medicine close monitoring of organ failure status is important for the prognosis of patients and for choices regarding ICU management. Major challenges in analyzing the multitude of data pertaining to the functioning of the organ systems over time are to extract meaningful clinical patterns and to provide predictions for the future course of diseases. With their explicit states and probabilistic state transitions, Markov models seem to fit this purpose well. In complex domains such as intensive care a choice is often made between a simple model that is estimated from the data, or a more complex model in which the parameters are provided by domain experts.

Our primary aim is to combine these approaches and develop a set of complex Markov models based on clinical data. In this paper we describe the design choices underlying the models, which enable them to identify temporal patterns, predict outcomes, and test clinical hypotheses. Our models are characterized by the choice of the dynamic hierarchical Bayesian network structure and the use of logistic regression equations in estimating the transition probabilities. We demonstrate the induction, inference, evaluation, and use of these models in practice in a case-study of patients with severe sepsis admitted to four Dutch ICUs.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Time is an important concept in medical care [1]. The evolution of disease is often closely related to treatment policy, and influences the chances of recovery. Physicians therefore closely monitor changes that occur over time. Intensive care medicine aims to care for patients with severe organ dysfunction or failure and to take over the role of the organ systems by means of machinery and medication if necessary. In this domain, insight into changes in organ failure over time is important for several reasons. First, it supports monitoring of the effect of current treatment and considerations concerning adaptations of the treatment regimen for individual patients. Second, it is useful for planning purposes, e.g., to schedule the availability of machinery such as mechanical ventilation in case of respiratory failure. Finally, as these insights describe scenarios of organ failure that are to be expected for pa-

tients with a given profile, they can serve as a basis for evaluation of care, where the expected scenario is compared with the course of disease as it occurred in practice.

In the Intensive Care Unit (ICU) organ failure is measured on a regular basis and generally expressed using organ failure scoring systems such as the Logistic Organ Dysfunction System (LODS) [2] and the Sequential Organ Failure Assessment (SOFA) score [3]. These systems typically express the degree of dysfunction/failure of individual organ systems on a particular day by means of a score (based on physiological values) and combine these individual scores into a total measure of organ failure. This results in daily scores of organ failure throughout the ICU stay until the patient has died or has been discharged from the ICU. One of the major challenges for data analysts is to extract meaningful patterns from these data, which provide insight into factors that influence development and persistence of different types of organ failure.

Most of the research on organ failure focuses on the relation between organ failure and eventual outcome of the ICU stay, often ICU mortality or hospital mortality. Within these studies the temporal aspect of the data is handled in various ways: it is not included in the models [3,4], sequential measurements of organ failure are summarized over time (i.e., horizontal temporal abstraction [5]), resulting

* Corresponding author. Address: Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Mailbox Str. 6.131, P.O. Box 85500, 3508 GA Utrecht, The Netherlands. Fax: +31 88 7555480.
E-mail address: l.m.peelen@umcutrecht.nl (L. Peelen).

in summary measures describing changes in organ failure [6–10], or *patterns* of organ failure are constructed and subsequently related to outcome [11,12]. However, much less is known on the dynamics of organ failure in *successive days* of ICU stay. As indicated before, insight into factors that influence the short-term risk of ICU death or development and persistence of organ failure can be helpful in daily clinical practice and planning and for evaluation of care. Therefore the current study focuses on day-to-day changes in organ failure in ICU patients.

One convenient type of model to investigate those day-to-day changes in organ failure over time is the so-called (discrete-time) *probabilistic state-transition model*, or Markov model. A Markov model identifies a set of possible clinical patient states, and describes the probabilities that states are followed by other states or persist through time [13,14]. In the ICU context, patient states are typically defined by number and type of failing organ systems.

Markov models have been used in various medical applications. In the medical statistical literature Markov models have been used, among others, to analyze patterns of changes in oncology [15] and allergic disease [16]. These applications generally have a univariate state representation with a small number of states (e.g., two to three disease states) and are often built based on clinical data.

In medical artificial intelligence (AI), Markov models often have a complex, multivariate state representation, and the model is expressed as a dynamic Bayesian network [17]. Applications are found in the field of, e.g., bacteraemia [18] and other infectious diseases [19–21]. These networks are often hand-crafted with the help of domain experts, and contain large numbers of subjectively estimated parameters. Several attempts have been made to build this type of model from data [22,23].

In this paper we combine these existing approaches to induce Markov models with complex, multivariate state descriptions from clinical data. We aim to derive models that can identify temporal patterns, predict future course of disease, and test clinical hypotheses. Furthermore, the models should be able to capture the dynamics of organ failure at the ICU on two levels: the processes regarding the eventual outcome of ICU stay (i.e., the short-term risk of dying at the ICU or being discharged from the ICU) and the processes related to development or persistence of organ failure in successive days when the patient stays at the ICU. To fulfill these requirements we have chosen to use a dynamic hierarchical Bayesian network structure in which the transition probabilities are estimated using logistic regression equations.

In this paper we describe these design choices and demonstrate the induction, inference, evaluation, and use of these models in practice by means of a case-study in patients admitted to the ICU with severe sepsis. These patients form an important part of the ICU population, with incidences ranging from 10% to 64% [24]. It is the leading cause of death in adult general ICUs with a mortality of 30–45% [24–26]. Given the large amount of organ failure that is seen in these patients [24,25] and the lack of knowledge regarding the mechanisms underlying organ failure in these patients [27], this is a typical disease area in which more insight into the dynamics of the disease is desired.

In the medical literature much debate has been going on whether it is the *severity* of organ failure that is related to the outcome, or failure in *specific* organ systems [3,28,29]. Therefore three Markov models are developed related to these clinical hypotheses. The first model describes how the severity of organ failure (i.e., the number of failing organ systems) changes over time, whereas the second model focuses on failure in specific organ systems. The third model is an extension of the second one and makes a distinction between persistence of existing organ failure and development of new organ failure on the next day.

The paper is constructed as follows. In Section 2 we introduce the data that are used in this study, describe the design, structure, induction, and inference of the models, and we indicate how the models were evaluated. Section 3 provides the clinical results from the models and the results from the evaluation. In Section 4 we briefly illustrate two possible applications of the models. In Section 5 we discuss our work and further relate it to other literature.

2. Methods

2.1. Data

2.1.1. Patient population

The analysis is based on prospectively collected data of all consecutive patients admitted with severe sepsis to four Dutch ICUs between January 1st, 2002 and December 31st, 2006. The four ICUs were all mixed-type ICUs, one in an academic hospital and three in teaching hospitals. The data has been collected as part of the Dutch National Intensive Care Evaluation (NICE) registry (www.stichting-nice.nl) and has been anonymized in a way that all patient identifying information, such as name and patient identification number, has been removed.

Patients were identified as being admitted with severe sepsis if they fulfilled the following criteria within the first 24 h of ICU admission: confirmed infection with at least two modified Systemic Inflammatory Response Syndrome (SIRS) criteria [30,31] and at least one dysfunctioning organ system. Patients for whom information on the first day of ICU admission was missing were excluded from the analyses as their compliance with the severe sepsis definition could not be verified.

2.1.2. Organ failure

Organ failure was measured on a daily basis using the SOFA scoring system [3]. For each patient, data were collected on the first 24 h of admission, and thereafter on sequential 24-h-periods synchronized with the starting time of the ICU day as used in each of the hospitals. In accordance with the original SOFA scoring system [3] the worst values of each 24 h period were used. Furthermore, data regarding patient characteristics, severity-of-illness in the first 24 h of admission (based on the Simplified Acute Physiology (SAPS) II score [32]), and ICU- and hospital outcome were registered. All data were collected as raw data according to stringent data definitions used within the NICE registry [33]. Based on these raw data the SAPS II score, the SOFA scores per organ system, and the total SOFA score were calculated in the coordinating centre of the NICE registry (Department of Medical Informatics, Academic Medical Center, Amsterdam). Originally, the SOFA scores for the individual organ systems range from 0 (indicating no organ dysfunction) to 4 (indicating most severe organ failure). For the case-study these scores were dichotomized into the categories 'non-failure' (organ system values 0–2) and 'failure' (organ system values 3 or 4) [3].

2.2. Model types

A discrete-time Markov model identifies a set of possible states, a set of time points, and has a function for describing the probabilities of transitions between states at subsequent time points. In medical applications, the set of possible states is the set of possible clinical conditions deemed relevant for the application at hand. A most basic Markov model for ICU patients, depicted in Fig. 1, has only three states (1) the patient is at the ICU; (2) the patient has died; (3) the patient has been discharged from the ICU. Here, the latter two states are *final* states from which no further transitions are possible.[1] We will use this basic model as a conceptual basis to

---

[1] In the models, we will ignore the fact that patients are sometimes readmitted to the ICU after being discharged, basically extending the first ICU stay. Instead, we treat such readmissions as independent ICU admissions.
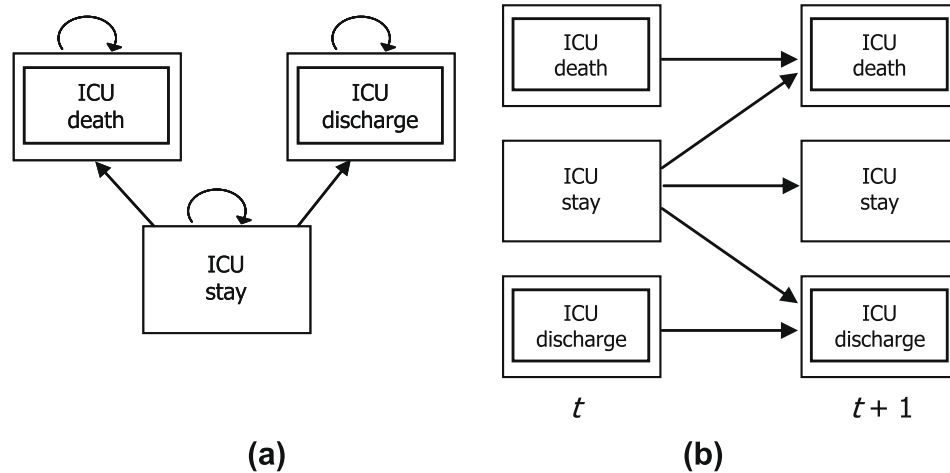
Fig. 1. Basic state-transition model for the ICU with three possible states: 'ICU discharge', 'ICU death' and 'ICU stay'. Arrows indicate possible transitions. ICU death and ICU discharge are depicted by a double box as these are so-called *final* states, i.e., states from which no transitions to other states are possible. (a) Static representation; (b) dynamic representation.

explore more elaborate models. In Fig. 1 the model is depicted using two representations: part (a) shows the 'static' representation in which time is modelled implicitly, and part (b) provides a 'dynamic' view on the process where two subsequent "time slices" are depicted (t and $t + 1$) and the arcs explicitly indicate transitions over time. Below, we discuss three extensions to the basic model, each of which is more complex than the previous one and is therefore equipped to address more refined clinical scenarios. In each of the models, the set of time points is the number of whole days since admission to the ICU, starting with the day of admission itself.

2.2.1. Model I amount of organ failure

Several studies have shown that a higher number of failing organs at ICU admission increases the probability of a prolonged ICU stay and death; the same holds for organ failure experienced at subsequent days during the ICU stay [3,6,7]. These findings were obtained with conventional statistical analysis techniques, such as logistic regression analysis. In Model I we describe day-to-day changes in the amount of organ failure and verify to which extent these relate to the outcome of ICU stay. To this end, the state 'ICU stay' in the basic model is replaced by five states describing the amount of organ failure, namely having zero, one, two, three, or more than three failing organ systems (Fig. 2), resulting in seven possible states in total.

2.2.2. Model II type of organ failure

It is known that problems with different organ systems, especially early in ICU stay, have a different influence on the outcome of the patient [9,10]. In Model II we investigate how failure in specific organ systems relates to organ failure on subsequent days and to the final outcome. In order to do so the state 'ICU stay' in the basic model is replaced by a combination of six binary variables indicating the presence of failure in each of the six organ systems in the body (coag, hepa, circ, neuro, renal, and resp, denoting failure in the coagulatory, hepatic, circulatory, neurological, renal, and respiratory system, respectively). This yields 64 states describing 'ICU stay', resulting in a total of 66 states in the model.

Fig. 3 gives a representation of Model II, using the dynamic representation of the basic model depicted in Fig. 1b. The six ellipses represent the binary variables expressing the presence or absence of failure in each of the individual organ systems if the patient is in the ICU. Fig. 3a highlights the transitions between the three states ICU death, ICU discharge, and ICU stay, similar to Fig. 1b. Fig. 3b describes the situation where the patient remains at the ICU, and

then focuses on the relations between organ failures at 2 subsequent days.

2.2.3. Model III differences between development and persistence of organ failure

The mechanisms involved in the interaction between different organ systems over time are complex as changes in organ failure are a result of the interplay between hemodynamics, oxygen transportation, and metabolic disturbances [27] and different factors are involved in the development and the persistence of specific organ failure [29]. Model II assumes that the same function can be used to model development and persistence of organ failure, i.e., it is assumed that the role of other organ systems remains similar independent from the presence of failure in the organ systems of interest. In Model III this assumption is relaxed, as separate functions are used to describe the mechanisms of development and persistence of organ failure, in which the role of other organ systems is modeled conditional on whether the organ failure of interest was already present at day t. For instance, where Model II assumes that the influence of the cardiovascular system on respiratory functioning is the same in all situations, Model III makes a distinction between normally breathing patients, at risk of increasing respiratory problems, and patients already having these problems, at risk of not recovering from it. The influence of the cardiovascular and other organ systems is known to be different for these two types of patients, and this is accounted for in Model III. We note that the representation of Models II and III is similar, i.e., Fig. 3 also represents Model III; the difference between the models is captured in the transition probabilities.

2.2.4. Additional state information

After establishing the sets of possible states and time points, the final ingredient of a Markov model is the transition probability function (tpf) which defines, for each pair of states s_t and s_{t+1} identified by the model, the probability $P(s_{t+1}|s_t)$ of moving from state s_t at time t to state s_{t+1} at time $t + 1$. In this relationship it is assumed that the state information at a given time point t suffices to describe the probability distribution over states at the subsequent time point $t + 1$ (the first-order Markov assumption). In the ICU domain it is, however, known that a patient's status at the time of ICU admission continues to influence the development of organ failure throughout the entire ICU stay [3,29] and may also influence the eventual outcome (i.e., death or survival) [32]. For this reason we also determine the transition probabilities conditional on the pa-
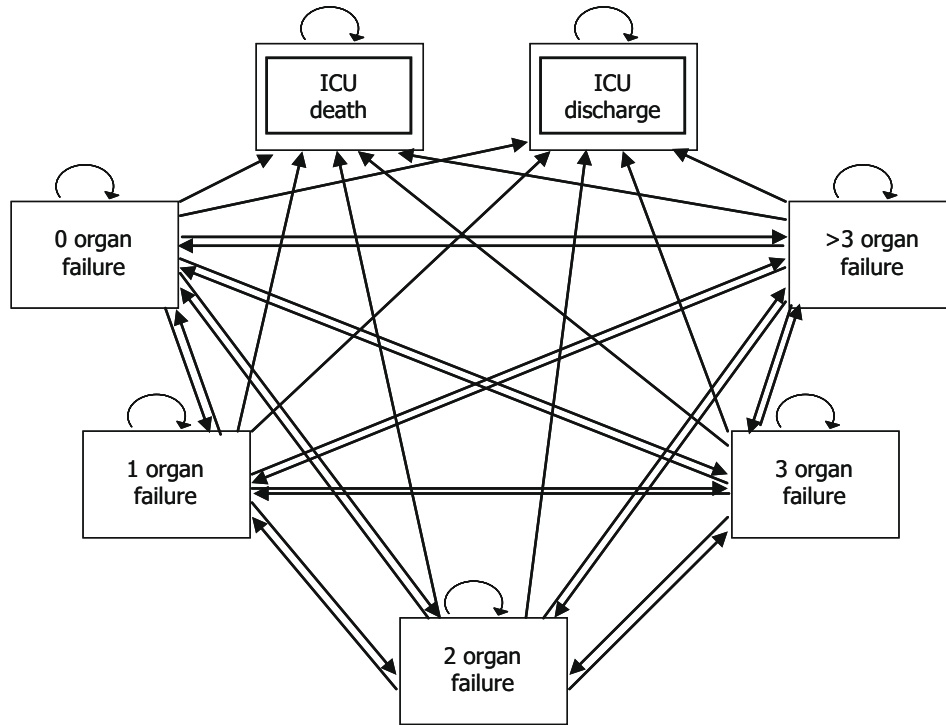
Fig. 2. Representation of Model I. State-transition model with seven possible states (Model I): ICU discharge, ICU death, and five states describing the number of failing organ systems when the patient is still at the ICU. Arrows indicate possible transitions. The final states ICU death and ICU discharge are indicated by a double box. In total 37 transitions are possible.
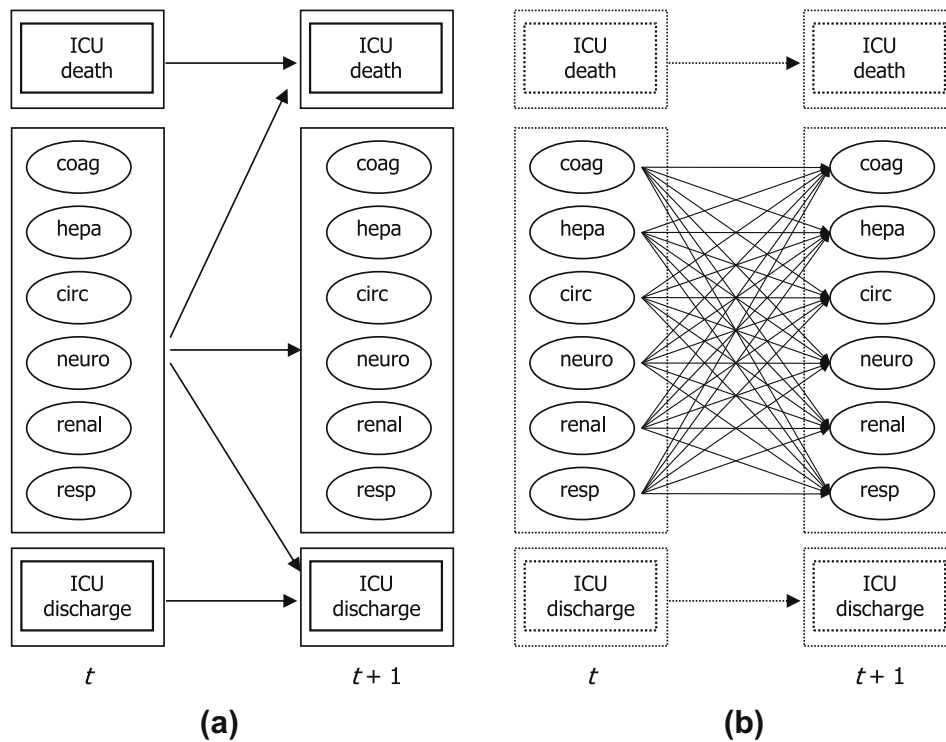
Fig. 3. Representation of Models II and III. Models II and III are represented using two levels. In (a) the upper level of the model is highlighted, focusing on ICU death, ICU stay and ICU discharge on day $t + 1$. In (b) the lower level of the model is highlighted, focusing on the relations between organ systems. Note that arcs run from each organ system on day t to each organ system on day $t + 1$.

tient's severity-of-illness at admission, expressed by means of the SAPS II score [32]. Furthermore, it was recently shown that the probabilities of death and discharge vary during ICU stay, even if

we take the patient's current clinical state into account [34]. Kayaalp et al. also found that non-stationary models performed better than stationary models in the ICU domain [35]. Therefore

time is included as a parameter in the transition probabilities. This means that the transition probabilities are written $P(s_{t+1}|s_t, SAPS, t)$, where *SAPS* denotes the severity-of-illness at admission.

2.3. Model structure

The patient state as described in the above models comprises information at two levels (i) where is the patient (at the ICU, died at the ICU, or discharged from the ICU) and (ii) if the patient is at the ICU, how is his or her condition, expressed in terms of organ failure. In the representation of the state of the patient and in the structure of the models we take this hierarchical nature of the patient state into account, as follows. We express the state of the patient at time point t, s_t, by a pair $s_t = (l_t, OF_t)$, where l_t describes the state of the patient at day t at the upper level, $l_t \in$ ('ICU death', 'ICU discharge', 'ICU stay'), and OF_t describes the organ failure at day t (lower level). Note that OF_t is only relevant if l_t equals 'ICU stay'. When l_t equals 'ICU death' or 'ICU discharge', the transition probabilities default to 1 for $l_{t+1} = l_t$ and zero otherwise. The transition probabilities that are to be estimated from the data therefore can be written as $P(s_{t+1}|OF_t, SAPS, t)$.

Theoretically, it is possible to estimate all transition probabilities of a Markov model directly from the data, in which case the tpf is a matrix of transition probabilities. In the current study, however, the large number of possible states (especially in Models II and III) hinders direct estimation of the transition probabilities because many cells in the tpf matrix will be empty or represented by very few data instances. A reduction in the number of probability estimates can be achieved by assuming a structured or parametric form for the tpf. Here we choose a combination of the two, using sets of additive logistic regression models (Models I, II, and III) as parametric form and hierarchical dynamic Bayesian networks (Models II and III) to model structure.

Additive logistic regression models have the advantage that they quantify the strengths of the relations between variables, and therefore enhance clinical interpretability of the Markov model. Furthermore, they allow for relatively straightforward inclusion of the additional state information, in our case the values of t and *SAPS*. The general form of the additive logistic regression models used in this study is

$$\log \frac{P(s_{t+1}|OF_t, SAPS, t)}{1 - P(s_{t+1}|OF_t, SAPS, t)} = \beta_0 + \beta_1 g_1(OF_t) + \beta_2 g_2(t)$$
$$+ \beta_3[g_2(t) * SAPS] \quad (1)$$

where g_1 is a parametric function specific for each of the three Markov models (see below), and g_2 is a discretization function that transforms the time parameter t into four categories (based on [34]), relating to 1–3, 4–7, 8–14, and more than 14 days of ICU stay, respectively. The SAPS II score is included as a continuous parameter, interacting with the discretized time parameter to allow for a varying influence of severity-of-illness at admission throughout the ICU stay. For each model the regression parameters $\beta_0, \ldots, \beta_3$ are estimated from the data using regression analysis (see also Section 2.4).

An important difference exists between Model I on the one hand and Models II and III on the other in the elaboration of the left-hand side of Eq. (1). While Models II and III employ a structured, multi-dimensional state description, in Model I the state description is 'flattened' to a single categorical variable with seven possible values.

When focusing on the right-hand side of Eq. (1), in Model I OF_t is a numerical variable that counts the *amount* of organ failure, i.e., the number of failing organ systems. Because we cannot assume a linear relationship between this number and log odds of the transition probabilities, OF_t is represented by an categorical variable. Furthermore, as relatively few patients experience organ failure

in four, five, or six organ systems, these patients are grouped into one category. So $OF_t \in$ ('0', '1', '2', '3', '>3'). Within the regression equations this is implemented by means of four binary dummy variables denoting failure in one, two, three or more than three organ systems, with absence of organ failure serving as the reference category and represented by each of the four dummy variables being set to zero. This results in the following equation for $g_1(OF_t)$:

$$g_1(OF_t) = \beta_{10}I(OF_t = 1) + \beta_{11}I(OF_t = 2) + \beta_{12}I(OF_t = 3)$$
$$+ \beta_{13}I(OF_t = ' > 3') \quad (2)$$

where again parameters $\beta_{10}, \ldots, \beta_{13}$ are estimated from the data when the model of Eq. (1) is fitted.

As indicated above, in Model I the future state of the patient, s_{t+1}, is represented by one categorical variable with seven possible values, one for each state represented in Fig. 2. In estimating the transition probabilities in Model I we thus estimate $P(s_{t+1}|OF_t, SAPS, t)$ using six regression equations (see Section 2.4 for more details on the statistical analysis). As in each regression equation eleven parameters are being estimated (four in $g_1(OF_t)$, three for the categorical variable describing t, and four for the interaction between t and SAPS), in total Model I contains $6 \cdot 11 = 66$ estimated parameters.

In Models II and III organ failure is described by six dimensions representing failure in each of the six organ systems. So OF_t itself is a tuple: $OF_t = (coag_t, hepa_t, circ_t, neuro_t, renal_t, resp_t)$, where $coag_t$, $hepa_t$, $circ_t$, $neuro_t$, $renal_t$, $resp_t$ are binary variables denoting failure in the coagulatory, hepatic, circulatory, neurological, renal and respiratory system at day t, respectively. In these models the large number of possible states prevents combining the two levels as was done in Model I. Therefore, $P(l_{t+1}|OF_t, SAPS, t)$ and $P(OF_{t+1}|OF_t, SAPS, t)$ are estimated separately. However, these probabilities do have a hierarchical relation, as $P(OF_{t+1}|OF_t, SAPS, t)$ becomes irrelevant if $P(l_{t+1} = $ 'ICU stay'$|OF_t, SAPS, t) = 0$.

To represent this hierarchical relation we use hierarchical dynamic Bayesian networks. Hierarchical Bayesian networks [36] are a generalization of standard Bayesian Networks which allow their variables to represent structured types. This means that each variable V in the network may be composed of multiple variables that are subsumed by V, and between which independency relations can be expressed. As a result, we can consider the Bayesian network at multiple levels of aggregation – hence the name *hierarchical* Bayesian network. For Models II and III, we used the hierarchical variant of *dynamic* Bayesian networks (DBNs) [37–39], which allow for modelling multivariate stochastic processes.

To restrict the complexity of the DBNs we assume that all state variables are mutually conditionally independent given their values at the preceding time point. This implies that only *diachronic dependencies* (i.e., relations between variables from different time slices) exist in the networks, and that *synchronic dependencies* (i.e., relations between variables within one time slice) are absent.

At both levels of the network, the conditional probability functions associated with variables at time $t + 1$ are all described by additive logistic regression models of the form of Eq. (1), now using the state transformation

$$g_1'(OF_t) = \beta_{11}coag_t + \beta_{12}hepa_t + \beta_{13}circ_t + \beta_{14}neuro_t$$
$$+ \beta_{15}renal_t + \beta_{16}resp_t \quad (3)$$

In each of these regression equations thirteen parameters are estimated, namely six for the organ systems, three for the categorical value representing t, and four for the interaction between t and SAPS. Model II consists of two regression equations on the upper level and six regression equations on the lower level, which implies that in total $8 \cdot 13 = 104$ parameters are to be estimated.

Model III makes a distinction between the development and persistence of organ failure. This occurs at the lower level depicted in Fig. 3. The distinction is implemented by estimating separate

sets of parameters $\beta_{11}, \ldots, \beta_{16}$ for the situations in which failure in the specific organ system is already present (e.g., $coag_t = 1$) or not (e.g., $coag_t = 0$). Note that the parameter related to the organ system of interest, in the example of coagulatory failure β_{11}, is obsolete in these models and is therefore removed. Model III thus requires $2 \cdot 13$ parameter estimates on the higher level (similar to Model II) and $6 \cdot (5 + 5 + 3 + 4) = 102$ parameters on the lower level, making 128 estimated parameters in total.

2.4. Model induction

In order to estimate the parameters for the tpf from the data, the original data were restructured such that each row in the dataset contained information on two subsequent time slices for the same patient. So for a patient who had been at the ICU for n days, the restructured dataset contained n records, of which the last record contained information on ICU day n (first time slice of that record) and on the destination of the patient after he has left the ICU (second time slice of that record). Following the Markov assumption all records of individual patients were considered independent from each other. For the regression equations in the models, the information in the records regarding the first time slice were used as predictor variables, whereas the information on the second time slice were considered the response (outcome) variables.

In Model I the outcome variable has seven possible values (related to the seven possible states in this model), and therefore multinomial logistic regression analysis [40,41], a generalization of binary logistic regression [42], was used. In this approach the six regression equations are jointly estimated, making sure that the resulting probabilities add up to one.

In Models II and III the higher level in the hierarchy has three states and was also modelled using multinomial logistic regression analysis. The lower levels of Models II and III both consist of six binary logistic regression models. In the lower level of Model III different sets of parameters were used to make the distinction between development and persistence of organ failure, as described in Section 2.3. This was obtained by stratifying the analysis on failure of the organ system of interest at day t, by including interaction terms in the regression equations.

For Models II and III, the tpf's on the lower level of the hierarchy (i.e., the function describing $P(OF_{t+1}|OF_t, SAPS, t, l_{t+1} = \text{'ICU stay'})$) were learned based on the records for which the response category was 'ICU stay' only (i.e., the nth record for each patient was not included in this estimation procedure).

2.5. Inference

Following [43,44], we computed predictive inferences on the Markov models using stochastic simulation. Let s_0 and SAPS be a patient's state and severity-of-illness score at the time of ICU admission, respectively. The conditional probability $P(s_t|s_0, SAPS)$ that this patient reaches state s_t at day t was derived as follows. First, we computed the joint posterior distribution on possible states at day 1 using the Eqs. (1)–(3), and randomly drew one state according to this distribution. The hierarchical relation between the levels was reflected by first drawing l_{t+1} based on the distribution for $P(l_{t+1}|OF_t, SAPS, t)$, and subsequently drawing OF_{t+1} based on the distribution for $P(OF_{t+1}|OF_t, SAPS, t, l_{t+1} = \text{'ICU stay'})$ if l_{t+1} equalled 'ICU stay'. This procedure was repeated for days 2, 3, . . ., until one of the final states (death or discharge from ICU) was reached. The series of simulated states was then stored in memory for later computations. The entire simulation process was repeated for N times, after which $P(s_t|s_0, SAPS)$ was estimated as the proportion of simulations where state s_t occurred at day t. To compute the marginal probability $P(s_t)$, we randomly drew s_0 and SAPS from the dataset (with replacement) at the start of each simulation.

2.6. Evaluation

The models in this study describe day-to-day changes in organ failure, which yields insight into factors that influence the dynamics of disease on the short term. For use in clinical practice it is, however, important that the models do not only fit the data on day-to-day transitions, but also properly describe changes in organ failure throughout the entire ICU stay, and the outcome. Therefore we have chosen to focus on this aspect in the evaluation, which has the advantage that it also allows for a comparison with other approaches in this clinical domain. We aim to answer the following two questions (1) Does the long-term predicted distribution of outcomes (ICU death or discharge) match its empirical distribution in the data (calibration)? and (2) Can the models distinguish survivors from non-survivors (discrimination)?

To answer the first question, simulations were conducted as described, where starting values were taken to be the actual values of the 2271 patients in the dataset. For each patient 500 simulations were conducted, resulting in 500 'cohorts' of ICU patients. The average ICU mortality and length of stay of these cohorts were calculated and compared with these outcomes in the original dataset. To obtain more insight whether changes over time were correctly modeled, for each of the cohorts the distribution of the patients over the three states 'ICU death', 'ICU stay' and 'discharged from the ICU' was calculated for the first 30 days after ICU admission. For six time points (days 2, 5, 10, 15, 20, and 30) the distribution over these three states in the simulated cohorts was compared with the distribution in the original dataset. The difference between observed and expected numbers was calculated as $\Sigma((O_{ij} - E_{ij})^2/E_{ij})$, where O_{ij} and E_{ij} are the observed and expected number of patients in state i at day j, respectively, where $i \in (\text{'ICU death', 'ICU discharge', 'ICU stay'})$ and $j \in (2, 5, 10, 15, 20, 30)$. This statistic was tested against a χ^2 distribution with two degrees of freedom [45], using a level of significance of 0.05. This part of the evaluation was performed for the basic model (serving as a reference model) and Models I, II, and III.

To evaluate to which extent the models discriminate between patients who die at the ICU and patients who survive ICU stay the dataset was randomly split into training and test sets of, respectively, 1541 and 757 patients. For each of the three models the following procedure was applied. First, the model was refitted on the training dataset, and subsequently the discriminatory performance of the models was estimated on the test set. To this extent we conducted simulation studies in a similar fashion as described above, again using 500 simulations per patient in the test set. Subsequently, for each patient the probability of ICU death was calculated as the percentage of those 500 simulated courses of disease in which the final state was 'ICU death'. This probability was subsequently compared to the actual outcome by means of the error rate and the area under the ROC curve (AUC).

Model development, simulations and analyses were performed using SPLUS version 6.2 (Insightful Corp., Seattle, WA, USA) extended with the nnet library for multinomial logistic regression analysis (`multinom` function) and the hmisc library to calculate the AUC (`rcorr.cens` function).

3. Results

In the period January 1st, 2002–December 31st, 2006 22,423 patients were admitted to the four ICUs. Of these patients 2271 (10.1%) patients fulfilled the criteria for severe sepsis within the first 24 h of admission. In total 512 patients (22.5%) died at the ICU and 752 patients (34.6%) died during hospitalization. Table 1 provides information on disease characteristics and outcomes of these patients. From the number and types of organ failure we

learn that most of the patients are admitted with failure in one or two organ systems. Almost 90% of the patients have respiratory failure at admission (for which they are being mechanically ventilated), and one-third of the patients have circulatory failure at admission. Hepatic failure is rarely seen, both at admission and during ICU stay. The mean length of ICU stay of these patients was 8.6 days (median value 4.8 days, inter quartile range (IQR) 1.9–9.9). For these 2271 patients in total 18,814 days with SOFA scores were available.

3.1. Results from the models

We will now briefly discuss and compare the three Markov models from a clinical perspective. The models are presented in terms of odds ratios (calculated from the coefficients in the regression equations) and transition probabilities (obtained by applying the tpf).

3.1.1. Model I changes in the amount of failure

Table 2 shows the transition probabilities of Model I for a patient with a SAPS II score equal to the median of these scores in the data (value 46) early during ICU stay (upper panel, current length of ICU stay 1–3 days) and in the second week of ICU stay (lower panel). The first column denotes the state of the patient on the current day; columns two through eight denote the state of the patient on the next day. A larger amount of failing organ systems at day t increases the short-term risk of ICU death and decreases the probability of ICU discharge (except for patients with

more than three failing organ systems, see fifth row, third column, $p = 0.028$). Furthermore, changes in the amount of organ failure occur gradually: in columns four through eight the highest transition probabilities are found around the diagonal, indicating it is most likely to remain in the same state or be at the ICU with one less failing organ system on the next day. When comparing the upper and the lower panel we observe that the time the patient has already spent at the ICU does not influence the probability of imminent death; however, the probability of ICU discharge is lower for patients with a longer ICU stay (second and third columns).

In the first 3 days of ICU stay a higher SAPS II score increases the probability of ICU death (odds ratio (OR) 1.05 with 95% confidence interval 1.05–1.06) and decreases the probability of ICU discharge (OR 0.97 (0.96–0.97)). After the patient has spent more than a week at the ICU, the ORs for ICU death and ICU discharge are 1.00 (0.99–1.02) and 1.00 (0.99–1.01), respectively, indicating that the severity-of-illness at admission is no longer significant for the state of the patient at the next day.

3.1.2. Model II changes in the type of organ failure

Table 3 depicts the results of Model II. Columns two and three relate to the transitions to the states ICU death and ICU discharge on the next day (upper level of the model), whereas columns four through nine focus on the prediction of failure in specific organ systems on the next day (lower level of the model). As indicated by the odds ratios the individual organ systems do play different roles with respect to short-term probabilities of ICU death. For example, neurological failure results in a fourfold risk of death on the next day (second column, OR = 3.9), while coagulatory, circulatory, and renal failure increases the risk of ICU death with a factor 1.6–2. Patients with respiratory failure are most likely to be at the ICU on the next day, as the risk of ICU death and ICU discharge both decrease when respiratory failure is present (lowest row).

The right-hand side of the table shows that respiratory and neurological failure are more likely to occur later during ICU stay, which is reflected by an increasing odds ratio with increasing time (upper block of odds ratios). Patients with a long ICU stay also have a relatively high risk of hepatic failure. Again, the influence of the SAPS II score decreases with time (see middle block of odds ratios), a pattern that is seen in all organ systems. When focusing on the lower right part of the table, which denotes the relations between organ systems, high odds ratios are seen on the diagonal, indicating that once organ failure is present, it is likely to persist. Several combinations of organ systems influencing each other can be distinguished, for example, hepatic failure is associated with coagulatory failure on the next day and vice versa (odds ratios of 2.8 and 3.1, respectively). Similarly, patients with circulatory failure have a 1.3 times higher risk of renal failure on the next day as compared to patients without circulatory failure.

Altogether, information on the type of organ failure on the current day provides an indication of the state of the patient and the type of organ failure that is to be expected on the next day. In comparison with Model I more distinction can be made between the patients. For example, according to Model I a patient who has just arrived at the ICU with a SAPS II score of 46 and one failing organ system has a probability of short-term ICU death of 0.010 (Table 2, second row, second column). From Model II we learn that in fact this probability ranges from 0.010 if the respiratory system shows failure to 0.054 when the neurological system fails. Similarly, Model I assigns patients with two failing organ systems a probability of ICU death of 0.025 (third row, second column), whereas Model II shows a more than 10-fold increase in risk of ICU death when comparing patients with hepatic and respiratory failure with patients who have circulatory and neurological problems (probabilities of 0.011 and 0.125, respectively).

Table 1
Baseline characteristics of the study population.

Severe sepsis patients (no)	2271
Male (no, %)	1337, 58.9
Type of admission (no, %)	
Medical	1363, 60.0
Urgent surgery	528, 23.2
Elective surgery	380, 16.8
SAPS II score (mean ± SD, median)	48.1 ± 17.2, 46
SOFA score (mean ± SD, median)	
Initial	8.6 ± 3.1, 8
Mean	7.7 ± 2.9, 7.1
Max	9.6 ± 3.5, 9
Failing organ systems at admission (no, %)	
0	160, 7.0
1	956, 42.1
2	804, 35.4
3	282, 12.4
4	59, 2.6
5	10, 0.4
6	0, 0.0
Organ failure at admission and during ICU stay (no, %; no, %)	
Coagulatory	207, 9.1; 405, 17.8
Hepatic	27, 1.2; 71, 3.1
Circulatory	796, 35.0; 1054, 46.4
Neurological	338, 14.9; 552, 24.3
Renal	312, 13.7; 568, 25.0
Respiratory	2016, 88.8; 2062, 90.8
Duration of ICU stay (days, median, IQR)	4.8, 1.9–9.9
Duration of hospitalization (days, median, IQR)	16.9, 7.0–35.6
ICU mortality (no, %)	512, 22.5
Hospital mortality (no, %[a])	752, 34.6

SAPS, simplified acute physiology score [32]; SIRS, Systemic Inflammatory Response Syndrome [30]; ICU, Intensive Care Unit; SD, standard deviation; IQR, inter quartile range.

[a] Percentage based on the number of patients, which is 2171 (as 100 of the 2271 admissions were readmissions to the ICU with severe sepsis).

Table 2
Daily transition probabilities based on the number of failing organ systems at day t (Model I).

Number of failing organ systems at day t	State at day $t + 1$						
	ICU death	ICU discharge	ICU stay with number of failing organ systems				
			0 OF	1 OF	2 OF	3 OF	>3 OF
ICU LOS 1–3 days							
0	0.007	0.483	0.406	0.080	0.022	0.002	0.000
1	0.010	0.100	0.098	0.670	0.116	0.006	0.000
2	0.025	0.031	0.024	0.213	0.638	0.066	0.003
3	0.047	0.020	0.004	0.060	0.386	0.452	0.032
>3	0.101	0.028	0.004	0.026	0.166	0.338	0.338
ICU LOS 7–14 days							
0	0.007	0.330	0.543	0.098	0.020	0.002	0.000
1	0.009	0.060	0.115	0.718	0.093	0.005	0.000
2	0.025	0.021	0.032	0.265	0.594	0.059	0.004
3	0.050	0.015	0.006	0.079	0.382	0.431	0.039
>3	0.101	0.019	0.006	0.033	0.154	0.302	0.385

Transition probabilities between states at day t and $t + 1$ for a patient with median SAPS II score at admission (SAPS II score of 46) with a current length of ICU stay of 1–3 days (upper panel) and 7–14 days (lower panel). Results are based on Model I. ICU, Intensive Care Unit; OF, organ failure.

Table 3
Predictive value of length of ICU stay, severity-of-illness at admission and the type of organ failure on the current day for the state of the patient on the next day.

State of the patient at day t	State of the patient at day $t + 1$							
	ICU death	ICU discharge	ICU stay with type of organ failure					
			Coagulatory	Hepatic	Circulatory	Neurological	Renal	Respiratory
ICU LOS (days)								
1–3 (reference)	1	1	1	1	1	1	1	1
4–7	2.8 (1.1–6.9)	0.5 (0.3–0.7)	1.4 (0.6–3.2)	1.8 (0.2–13.3)	0.9 (0.5–1.5)	0.6 (0.3–1.2)	1.3 (0.7–2.5)	1.9 (1.1–3.2)
8–14	5.6 (2.0–15.9)	0.1 (0.1–0.2)	1.2 (0.4–3.1)	0.4 (0.0–4.1)	1.8 (1.0–3.1)	0.9 (0.5–1.8)	1.0 (.5–1.9)	4.6 (2.6–8.1)
>14	6.2 (2.4–16.0)	0.1 (0.0–0.1)	1.8 (0.8–4.5)	16.9 (2.5–112)	0.6 (0.4–1.2)	3.0 (1.6–5.8)	1.8 (1.0–3.3)	5.1 (3.0–8.7)
SAPS (per 10 points) per LOS category[*]								
1–3	1.6 (1.4–1.7)	0.7 (0.7–0.8)	1.3 (1.2–1.4)	1.4 (1.1–1.7)	1.1 (1.0–1.2)	1.3 (1.2–1.4)	1.3 (1.2–1.4)	1.4 (1.3–1.5)
4–7	1.2 (1.0–1.4)	0.9 (0.8–1.0)	1.2 (1.0–1.4)	1.4 (0.9–2.1)	1.0 (0.9–1.2)	1.3 (1.1–1.5)	1.2 (1.1–1.4)	1.1 (1.0–1.3)
8–14	1.0 (0.9–1.3)	1.0 (0.9–1.2)	1.1 (0.9–1.3)	1.7 (1.1–2.6)	1.0 (0.8–1.1)	1.2 (1.1–1.4)	1.3 (1.1–1.5)	1.0 (0.8–1.1)
>14	1.1 (0.9–1.3)	0.9 (0.8–1.1)	1.1 (0.9–1.3)	0.9 (0.6–1.4)	1.1 (1.0–1.2)	1.0 (0.8–1.1)	1.2 (1.1–1.4	1.0 (0.9–1.1)
Type of organ failure at day t								
Coagulatory	2.0 (1.5–2.5)	0.9 (0.7–1.2)	123 (103–146)	3.1 (1.9–5.0)	1.1 (0.9–1.4)	1.0 (0.8–1.3)	1.3 (1.1–1.6)	1.7 (1.3–2.3)
Hepatic	1.0 (0.6–1.7)	0.5 (0.2–0.9)	2.8 (1.9–4.3)	841 (564–1254)	1.5 (1.0–2.1)	1.5 (1.0–2.3)	1.2 (0.9–1.7)	1.3 (0.8–2.0)
Circulatory	1.9 (1.5–2.3)	0.3 (0.2–0.3)	1.3 (1.1–1.6)	1.4 (0.9–2.2)	59 (53–66)	0.9 (0.8–1.1)	1.3 (1.2–1.5)	2.7 (2.3–3.2)
Neurological	3.9 (3.2–4.9)	0.6 (0.5–0.8)	1.1 (0.8–1.4)	1.1 (0.6–2.0)	0.8 (0.6–0.9)	92 (80–105)	0.8 (0.6–1.0)	1.7 (1.4–2.2)
Renal	1.6 (1.2–2.0)	0.7 (0.6–0.8)	1.4 (1.1–1.7)	0.9 (0.5–1.4)	1.6 (1.3–1.8)	0.8 (0.6–1.0)	34 (30–39)	1.1 (0.9–1.3)
Respiratory	0.5 (0.4–0.7)	0.1 (0.1–0.1)	0.7 (0.6–1.0)	1.8 (0.9–3.5)	1.6 (1.3–2.0)	1.6 (1.3–2.0)	0.9 (0.7–1.0)	68 (59–79)

Values denote odds ratios (OR) with 95% confidence interval (CI) and are based on Model II. For the states 'ICU death' and 'ICU discharge' odds ratios are based on the upper level of the model using the state 'ICU stay' as the reference category. The columns related to the state 'ICU stay' describe the odds ratios of having failure in the particular organ system on day $t + 1$ (based on the binary logistic regression equations in the lower level of the model).
ICU, Intensive Care Unit; LOS, length of stay (i.e., the days the patient has already spent at the ICU); SAPS, simplified acute physiology score [32].
[*] Standard errors for the interaction between SAPS II score and ICU LOS are calculated based on [42].

3.1.3. Model III differences in the development and persistence of organ failure

Table 4 resembles the lower right part of Table 3 and denotes the influence of the organ systems on *development* of new organ failure on the next day (upper part of the table) and on *persistence* of existing organ failure (lower part). We now see that the association between hepatic and coagulatory failure found in Model II is restricted to the development of organ failure (odds ratios 3.1 and 6.9, respectively); when focusing on persistence, no significant associations between these organ systems are found (odds ratios 2.5 and 0.8). Likewise, patients with circulatory failure have a 1.6 times higher risk on development of renal failure, whereas there is no association between these two organ systems once renal failure is present. Interestingly, the other way around results point in the opposite direction: renal failure is not associated with the development of circulatory failure, but is involved in the persistence of failure in this body system.

3.2. Evaluation of the models

The first evaluation focused on the calibration of the models with respect to length of ICU stay and eventual outcome. Table 5 presents the results of this evaluation and compares expected ICU mortality and mean length of ICU stay for each of the three models and the basic model with the original data. All models correctly estimate mortality and mean length of ICU stay, as for each of the models the observed mortality and length of stay are contained within the 95% confidence intervals of the estimates. The lower part of the table depicts the χ^2 values for the difference in observed and predicted distribution of the patients over the states 'ICU death', 'ICU discharge' and 'ICU stay' for six time points during the first 30 days of ICU stay, where higher values denote worse fit. Except the first days Models I, II, and III fit the observed data reasonably well. This is also reflected by Fig. 4, which depicts the distribution of the population over the three states ICU stay, ICU

Table 4
Predictive value of organ failure at the current day with respect to the development and persistence of organ failure at the next day (Model III).

Presence of organ failure on day t	Development of new organ failure on day $t + 1$					
	Coagulatory	Hepatic	Circulatory	Neurological	Renal	Respiratory
Development of organ failure						
Coagulatory	–	6.9 (4.0–12.0)	0.9 (0.6–1.3)	1.4 (1.0–1.9)	1.7 (1.3–2.2)	0.6 (0.3–1.2)
Hepatic	3.1 (1.7–5.6)	–	0.9 (0.5–1.7)	2.0 (1.2–3.4)	1.3 (0.8–2.1)	0.6 (0.2–2.0)
Circulatory	1.6 (1.2–2.0)	1.7 (1.0–2.9)	–	0.9 (0.7–1.2)	1.6 (1.3–1.9)	3.7 (2.5–5.7)
Neurological	1.1 (0.7–1.6)	1.2 (0.6–2.5)	0.7 (0.5–0.9)	–	0.7 (0.5–1.0)	1.5 (0.8–2.6)
Renal	1.5 (1.1–2.0)	1.8 (1.0–3.1)	1.2 (0.9–1.5)	1.1 (0.8–1.5)	–	0.8 (0.5–1.1)
Respiratory	0.6 (0.4–0.8)	1.4 (0.6–3.7)	1.7 (1.3–2.2)	2.6 (1.7–4.2)	0.9 (0.7–1.1)	–
Persistence of organ failure						
Coagulatory	–	0.8 (0.4–1.6)	1.4 (1.0–1.9)	0.7 (0.5–1.1)	1.0 (0.7–1.4)	2.4 (1.1–5.2)
Hepatic	2.5 (1.3–4.9)	–	2.5 (1.2–5.2)	0.8 (0.5–1.3)	1.2 (0.7–2.1)	1.5 (0.4–5.9)
Circulatory	1.1 (0.8–1.6)	0.8 (0.4–1.7)	–	0.9 (0.6–1.4)	1.0 (0.9–1.2)	2.6 (1.6–4.3)
Neurological	1.1 (0.7–1.8)	0.8 (0.2–2.4)	0.8 (0.5–1.2)	–	0.9 (0.6–1.3)	1.8 (1.0–3.2)
Renal	1.2 (0.9–1.7)	0.4 (0.2–0.7)	2.0 (1.5–2.5)	0.6 (0.4–0.8)	–	1.2 (0.8–1.8)
Respiratory	1.2 (0.8–1.6)	2.2 (1.4–3.6)	1.5 (1.3–1.7)	0.8 (0.7–1.0)	0.9 (0.8–1.1)	–

Values denote odds ratios (OR) with 95% confidence interval (CI) adjusted for current length of stay (i.e., value of t) and severity-of-illness at admission. Values are based on the lower level of Model III.

death, and ICU discharge in the first 30 days of ICU stay (depicted in blue, green, and red, respectively). The solid lines denote the observed distribution (similar for all three graphs) and the dotted lines indicate the expected distribution with a 95% confidence interval.

The underlying question to be addressed when assessing the calibration of the model is whether the chaining of day-to-day probabilities leads to appropriate predictions for the entire ICU stay. As we learn from Table 5 this is not the case for the basic model, hence one should at least include information on the presence of organ failure to arrive at correct longer-term predictions. Models I, II, and III have well-calibrated predictions over longer time spans, except on day 2 where all models deviate from the observed outcome. Therefore we believe this is not to be attributed to the models, but rather to other reasons, probably related to the nature of the data collection or the SOFA score.

In the second evaluation we focused on discrimination of the models when they are used to predict the outcome of the patient on a separate test set. In prediction of eventual ICU death the error rates were 17.7%, 18.1% and 17.8% for Models I, II, and III, respectively. The AUCs (95% confidence interval) were 0.79 (0.71–0.87), 0.79 (0.71–0.87), and 0.80 (0.72–0.88), respectively. Also in prediction of death over shorter time periods the discriminatory performance of the three models was similar (AUCs of 0.84 (0.74–0.93); 0.83 (0.74–0.92); 0.83 (0.74–0.93) and 0.82 (0.74–0.90); 0.82 (0.73–0.90); 0.82 (0.74–0.90) for death within 3 days and death within a week, respectively).

4. Application

In clinical practice the models can be used for various purposes, two of which we will now briefly illustrate. The first application lies at the patient level and analyzes the various prognostic scenarios for a given patient. These scenarios can be used to inform the patient and his relatives, to see which protocols and guidelines apply to the patient, and to make treatment decisions. The second application lies at the level of the ICU and predicts logistic requirements for a given unit during the forthcoming days. We will briefly describe these applications and provide examples using Model III and real patient data.

4.1. Scenario analysis

Applying simulations on Model III provides us with probabilities of the amount and type of organ failure for all future time points and on survival. Based on this information, probabilities for various prognostic scenarios can be calculated. Table 6 shows a number of (hierarchical) scenarios and provides the related probabilities for two patients. Patient 1 is a 79-year-old man, admitted with severe sepsis from the ward today, with considerable sever-

Table 5
Evaluation results: comparison of observed and expected characteristics of the cohort.

	Original data	Model			
		Basic	I	II	III
ICU mortality (%)	22.6	22.5 (20.8–24.2)	22.5 (21.0–24.0)	22.4 (20.9–24.0)	22.5 (21.0–24.0)
LOS ICU (mean, d)	8.3[a]	8.3 (8.0–8.6)	8.3 (7.8–8.6)	8.4 (8.0–8.8)	8.4 (8.0–8.8)
Observed versus expected distribution (χ^2-value)[b]					
Day 2		28.19[c]	10.39[c]	9.91[c]	9.78[c]
Day 5		24.25[c]	0.09	0.55	0.66
Day 10		23.98[c]	1.89	1.75	2.18
Day 15		2.82	0.28	0.17	0.05
Day 20		3.23	0.08	0.13	0.25
Day 30		21.94[c]	0.17	0.02	0.02

For each of the models numbers are based on 500 simulations for all 2271 patients. For ICU mortality and mean LOS ICU numbers represent the mean and 95% confidence interval.
[a] To allow for a fair comparison with the models we report the mean of the truncated lengths of ICU stay.
[b] The χ^2-value was calculated as $\Sigma((O_{ij} - E_{ij})^2/E_{ij})$, where O_{ij} and E_{ij} are the observed and expected number of patients in state i at day j, respectively, where $i \in$ (ICU death, ICU discharge, ICU stay) and $j \in$ (2, 5, 10, 15, 20, 30).
[c] $p < 0.05$, indicating that predicted distribution of patients over the states 'ICU death', 'ICU stay' and 'ICU discharge is significantly different from the observed distribution. ICU, Intensive Care Unit; LOS, length of stay.
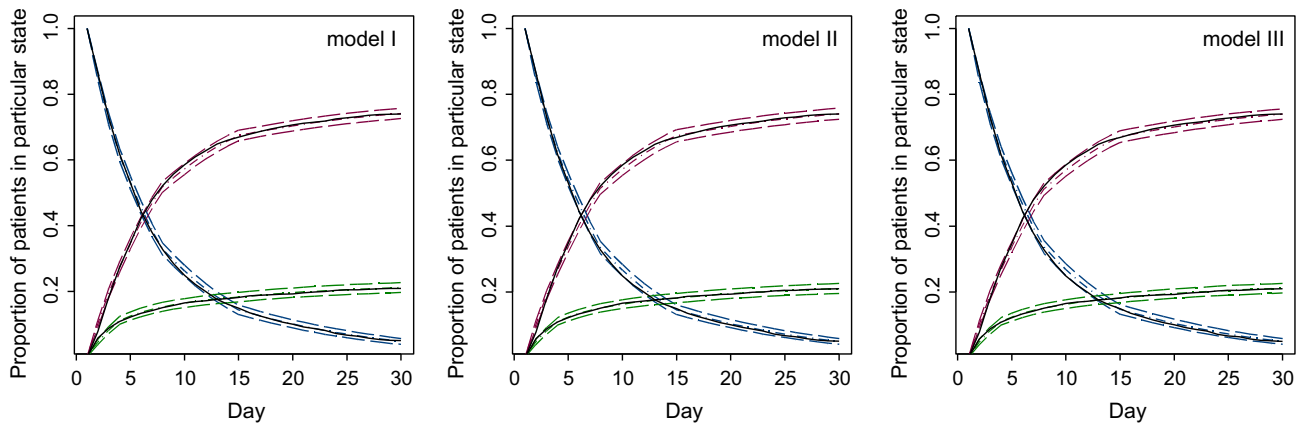
Fig. 4. Distribution of patients over the states 'ICU death', 'ICU discharge' and 'ICU stay': observed versus expected. Results from the 500 simulated cohorts of 2271 patients based on the starting values from the original cohort. Figures display the observed and predicted distribution of the cohort over the states 'ICU stay', 'ICU death' and 'ICU discharge' (in blue, green and red lines, respectively) for the first 30 days after ICU admission. The black lines depict the values for the original cohort; the colored lines indicate predicted probabilities over the 500 simulations with the 95% confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Table 6
Scenarios for future course of disease for two patients based on Model III.

Scenario	Patient 1	Patient 2
1. Worsening in the next 3 days	0.23	0.37
1.1 Death within 3 days	0.10	0.31
1.2 Increase of organ failure in the next 3 days	0.13	0.06
1.2.1 Eventual death	0.03	0.03
1.2.2 Eventual recovery	0.10	0.03
2. Improvement in the next 3 days	0.47	0.54
2.1 Discharge within 3 days	0.06	0.05
2.2 Decrease of organ failure within 3 days	0.41	0.49
2.2.1 Eventual discharge	0.35	0.35
2.2.2 Eventual death	0.06	0.14
3. Stable situation	0.30	0.09

Patient 1 is a 79-year-old man, admitted with severe sepsis from the ward today, with considerable severity-of-illness (SAPS II score of 55). At admission he has respiratory and circulatory failure.
Patient 2 is a 21-year-old woman who has been admitted from the ward 7 days ago with severe sepsis and metabolic problems, and a SAPS score of 33. Currently she has failure in the circulatory, respiratory, and neurological system.

ity-of-illness (SAPS II score of 55). At admission he has respiratory and circulatory failure. Patient 2 is a 21-year-old woman who has been admitted from the ward 7 days ago with severe sepsis and metabolic problems, and a SAPS score of 33. Currently she has failure in the circulatory, respiratory, and neurological system. The probability of change in the condition of the second patient is much higher when compared to the first patient; this is mainly due to a much higher probability of imminent ICU death.

4.2. Prediction of logistic requirements

For the prediction of logistic requirements for a given Intensive Care Unit, simulations are conducted for all patients at the unit, using their current situation as starting values. Given these values, the model can provide information on which patients are likely to be discharged or to die, and on the amount and type of organ failure that is to be expected. This information can subsequently be used for the planning of personnel and machinery, for example to determine how many patients require mechanical ventilation (in case of respiratory failure) or hemodialysis (in case of renal failure). For this application the predicted probabilities summed over patients: from a logistic perspective it is not important to know ex-

actly which patients will be discharged, as long as the number of discharges is known. And in contrast to the scenario analysis application, it is important to obtain information on the exact timing of events to be able to plan the utilization of ICU beds, ICU machinery, and operation theatres.

As an example we have applied Model III to data from the hospitals in our dataset. On March 16, 2005, in two ICUs (which will be considered one unit in this example) 10 patients were present who had been admitted with severe sepsis. Table 7 shows the predicted number of deaths, ICU discharge, and the occurrence of the types of organ failure for the next 3 days and after 1 week. These predicted numbers were obtained by adding the probabilities for each of the patients. Note that these numbers are based on the patients that are currently present at the ICU, the model does not take into account patients that are admitted in between (e.g., in the predictions for March 19 newly admitted patients on March 17 and 18 are not taken into account).

Other directions for the application of this type of models can be found in automatically identifying the sickest patient at the unit (to direct the nurses attention), in the field of health outcomes research, where the observed course of disease is compared to the predicted course of disease (as given by the model), and in the field of logistics and waiting-time analysis (e.g., as in [46]).

5. Discussion

In this paper we have demonstrated how complex hierarchical dynamic Bayesian networks can be derived from temporal clinical data to verify clinical hypotheses and to predict future course of

Table 7
Predictions at the ICU unit level based on Model III.

	March 17 (tomorrow)	March 18	March 19	March 24 (next week)
Died at ICU	0.3	0.5	0.7	1.2
Discharged from ICU	1.2	2.1	2.7	5.0
Multiple organ failure	5.0	4.2	3.6	3.2
Respiratory failure	8.0	7.2	6.4	5.7
Renal failure	1.0	0.9	0.7	0.7

Numbers in the table denote the expected number of patients. Predictions are based on Model III, using the information of 10 patients who were present at the ICU on March 16, 2005.

disease. In the application to ICU patients with severe sepsis we have seen that both the number and type of organ failure can be used to describe day-to-day changes in organ failure. The models describing failure in specific organ systems (Models II and III) provide additional insight into the interplay between the organ systems. From our evaluation studies we learned that increasing the complexity of the models did not reduce their performance, as all models showed more or less similar performance. Therefore, Model III is to be preferred as it provides more information without diminishing the accuracy of predictions.

5.1. Clinical findings

In the models we have developed changes in the amount of organ failure developed gradually. Furthermore, patients who are more severely ill at admission are more likely to develop organ failure during their ICU stay, although the influence of illness severity at admission diminishes over time. When we focus on failure of specific organ systems, Model III shows that organ systems play different roles in the development and persistence of organ failure on the next day. Furthermore, Models II and III show that organ failure does not easily resolve once it exists, a finding which holds especially for neurological and hepatic failure.

Our findings relate to discussions in the clinical scientific literature about the role of organ failure in the future course of disease and outcome of the patient. A higher number of organ failure, both at admission and during ICU stay, has been found to be related to a poor outcome [3,29], and this is reflected in our models by the high short-term risk of ICU death for patients with multiple organ failure and by the influence of the SAPS II score. We have found in Models II and III that information on failure of specific organ systems is also predictive for the outcome of intensive medical care, which was also concluded by Vincent et al. [47], but is in contrast with the findings of Pittet and colleagues [28]. This contrast can be explained by various factors, among which location (US vs. The Netherlands), study period (second half of 1980s vs. 2002–2006), and type of patients (patients who develop sepsis at the ICU vs. patients admitted with sepsis). Another important difference is that the aim of the Pittet study was to arrive at the best prediction model, whereas our focus was to model the role of each organ system more explicitly. In their models the variables related to individual organ failure were all removed in the stepwise statistical analysis, except respiratory failure.

When focusing on particular scenarios using the method described in Section 4, Models II and III confirm previous findings that early failure in the renal or coagulatory system relates to a poor outcome [9,48,49], whereas early improvement of circulatory failure is associated with a better prognosis [10].

5.2. Markov models in critical care medicine

Markov models have been applied in health care for various purposes since the 1970s [13], most prominently in cost-effectiveness analyses [50] and clinical decision analyses [51]. In critical care medicine, Markov models have for example been used to predict length of ICU stay [52] and hospital stay [53]. In both of these approaches a small number of states was used (four and eight, respectively) and clinical data were used to directly estimate the transition probabilities in the probability matrix. Bäuerle and colleagues developed a Markov model to analyze the course of disease of critically ill patients using the three states 'well', 'septic' and 'dead' [54].

A related approach recently used in the domain of critical care is *dynamic microsimulation*. In this approach the underlying process is modeled using a Markov model. However, one is not interested in the value of the parameters of the model, in the transition probabilities, or in accurate predictions for individual patients. Instead, the focus is on simulations on the model, which are used to investigate the influence of particular parameters (e.g., the time already spent at the ICU, severity-of-illness, introduction of new medication) on the cohort level [55,56]. This is in contrast to our approach where we derive clinical knowledge from the parameters of the models (such as described in Section 3) and use the models for individual patient predictions (as described in Section 4). In fact, the approach we took in our strategy to evaluate the models, in particular the first part of the evaluation, closely resembles dynamic microsimulation.

Clermont and colleagues [57] used dynamic microsimulation to describe changes in severity-of-illness in the general ICU population. As part of their results they do provide information on relations between organ systems over time, which show the importance of persistence of organ failure over time and the relation between hepatic and coagulatory failure, which is in accordance with our findings. More recently, Saka et al. used dynamic microsimulation to describe disease progression in patients with pneumonia-related sepsis [34]. Although they focused on a slightly different patient population, we confirmed their findings regarding the importance of length of stay (or duration of the process) in the transition probabilities. Their models are based on an empirical approach (related to k-nearest neighbour prediction) which does not provide direct insight into the strengths of the associations. This impedes a further comparison with our clinical findings.

5.3. Markov models extracted from temporal clinical data

In medical artificial intelligence dynamic Bayesian networks are a common implementation of Markov models. Probabilistic forecasting with dynamic Bayesian networks was first described by Dagum et al. in the early 1990s [38]. They manually assessed the structure of their dynamic Bayesian network on the problem of forecasting sleep apnea episodes; the network parameters were estimated from monitoring data [39]. A related approach was described by Riva and Bellazzi where the network structure was learned from data, using a model-selection approach based on predictive performance [22]. They applied their method to time series of blood glucose measurements from patients with diabetes mellitus.

Kayaalp and colleagues developed various DBNs to describe changes in organ failure in ICU patients, also based on the SOFA score [11,23,35]. In [35] they learned both stationary and non-stationary networks from data using a heuristic search strategy. The optimal model was selected using a scoring metric and turned out to contain only two nodes: the total SOFA score prior to discharge and the outcome (mortality). In [11] Bayesian networks were developed for six outcomes. First patterns of organ failure were identified and their predictive performance was determined in bivariate Bayesian networks. Subsequently the patterns with highest predictive performance (AUC) were included into the final Bayesian network.

A recent application of dynamic Bayesian networks was presented by Charitos and colleagues [20]. Their network is used to predict the development of ventilator-associated pneumonia (VAP) in ICU patients. In this application the structure and the probabilities of the network were determined using physician experts. Recently, this work was extended by Visscher et al. [21] who developed temporal Bayesian Networks for patients who did and those who did not develop VAP, using context-specific independences. In this application the network variables were pre-specified and data was used both to assess the network structure and to estimate the parameters. The network structure was assessed by constraint-based structure learning.

Given the complex processes underlying the dynamics of organ failure in ICU patients we have used a complex hierarchical state representation, implemented by dynamic hierarchical Bayesian networks. This combination of dynamic Bayesian networks and hierarchical Bayesian networks distinguishes our work from the aforementioned approaches. Other important differences are found in the assessment of the structure of the network and the representation of the tpf.

As indicated before Riva et al. and Kayaalp learned the optimal network structure from clinical data by developing multiple networks and subsequently selecting the structure with the best performance [22,23]. Visscher et al. determined the nodes of the network in advance and chose the arcs in the network based on clinical data using statistical significance and physician expert opinion [21]. In contrast, in our study we used a fixed model structure and did not use the data to learn the structure of the models. This could have been implemented by applying stepwise-backward variable selection in the logistic regression procedure (which would amount to removing the arcs in the Bayesian network one-by-one) based on the statistical significance of the parameters. We have, however, explicitly chosen not to use this procedure because in medical statistics variable selection based on significance testing is considered arbitrary, as it is heavily influenced by the size of the data set and the level of significance [58,59]. This choice explicitly distinguishes our work from approaches in Bayesian networks in which the structure is learned from data based on statistical significance.

In most of the related approaches the transition probabilities were expressed using (conditional) transition probability matrices [21–23]. Dagum and Galper used additive models in which the relation between predictors and outcome was modeled for each predictor separately and subsequently combined [39]. In contrast to these approaches we parametrized the tpf by means of multivariate additive logistic regression models, including multiple predictors into one regression equation. This has several advantages in comparison with the aforementioned approaches. First, the number of parameters is rapidly reduced; second, inclusion of 'history information' to relieve the first-order Markov assumption is rather straightforward as it is simply represented by an additional parameter in the model; third, the parameters in the model are easily expressed in terms of odds ratios, which enhances the clinical interpretability of the models; and finally, logistic regression models are extensively studied and applied in medicine [45] and are therefore easily implemented in existing software packages.

5.4. Limitations

Throughout the study we have taken various steps to reduce the complexity of the parameter space. First, we have dichotomized the SOFA score for individual organ systems into 'failure' and 'non-failure' categories. Although this dichotomization is often applied, compared to the original scale this results in a loss of precision when describing clinical states. For example, patients with severe dysfunction (comparable to a SOFA organ score of 2) in several organ systems were categorized together with patients completely free from that organ failure. This is probably the explanation for our finding that changes in organ failure occur only gradually: the patient's condition has to change drastically before it is recognized as such. This might also explain the fact that our models predict that patients without organ failure may still remain at the ICU or even die. A second explanation for the latter observation might be the sampling frequency, as the SOFA score is measured only once a day. The fact that the SOFA score is based on the worst values for a 24 h period might have increased the presence of persistence in our models. These problems could be solved by measuring organ failure using a finer grid, both with respect to time and organ failure measurement. Including the original SOFA

scores into the models is possible within the approach we presented by using multinomial proportional odds logistic regression models [40], however, this comes at the cost of a reduction of clinical interpretability.

Second, in the models presented in this paper we included either the number of failing organ systems (Model I) or failure in specific organ systems (Models II and III) as covariates, but not both. In an attempt to develop a model containing both (results not shown), extreme values for the odds ratios were found, indicating a severe multicollinearity problem. However, when informally comparing the probabilities obtained by the models, Model II showed considerable refinement of the probabilities as compared to Model I, suggesting that using information on the type of organ failure is to be preferred over using the amount of organ failure only. A more extensive evaluation is, however, required to confirm this hypothesis.

Third, in Models II and III we considered failure in each of the organ systems separately and did not take into account that it may be particular combinations of organ failure that are of predictive value for disease progression. A solution to these problems could be to create 'meta-variables' that describe combinations of organ failure (e.g., 'circulatory and respiratory failure') or combine the number and type of organ failure (e.g., 'the patient has respiratory failure only', or 'the patient has at least two failing organ systems, one of which is renal failure'). These combinations could be derived from the data or could be provided as background knowledge by the physician.

Finally, to reduce the number of estimates we have chosen to use additive logistic regression equations in the model, which assumes linear relationships between the predictor variables and the state of the patient on the next day. Using other methods to describe the tpf, e.g., trees, could alleviate the assumptions of the linear model [60], and could at the same time provide a solution to incorporate the aforementioned influence of combinations of organ systems.

In our case-study we performed a general evaluation of the models, focusing on calibration with respect to length of ICU stay and eventual outcome, and on discrimination with respect to the patients who did survive the ICU admission with severe sepsis, and those who did not. The results of these evaluations were promising as all three models performed well both on discrimination and calibration. Developing more complex models generally increases the risk of overfitting the model to the data, and thereby decreasing performance of the model. The results of our evaluation show that the increasing complexity of the three models did not lead to a reduction in performance. During the evaluation we conducted multiple comparisons, for which a level of significance of 0.05 might have been too optimistic. However, using a value 0.01 would not alter the conclusions. Further evaluation of the models should be guided by the purpose for which the model will be used. For example, when the model is used for scenario analysis, evaluation should focus on this aspect. Given the vast number of possible scenarios, the scenarios can be aggregated by temporal abstraction [5] similar to the example presented in Section 4, possibly guided by clinical background knowledge to identify the scenarios of interest to (local) clinical practice. In the evaluation the presence of the scenario then becomes the outcome of interest, and discrimination and calibration with regard to this outcome measure can be assessed.

From a clinical point of view the models have some limitations. First, no information was included into the models regarding treatment of the patient. In fact, currently treatment strategy is modelled implicitly, as changes in organ failure are partly determined by treatment. For example, the observation that patients with more than three failing organ systems are likely to be discharged probably reflects a do-not-resuscitate policy in which patients

are discharged to the ward to die. We cannot exclude the possibility that the included hospitals used different treatment strategies or that treatment strategy has changed over time [61]. This might have diminished the size of the effects (odds ratios) that were found and the extremity of predicted probabilities in the models (due to regression to the mean). When the models are used for planning purposes we therefore suggest adding information regarding the hospital and treatment strategies that are employed in these patients.

Second, we have considered readmissions as separate admissions (which is represented by considering ICU discharge a final state). However, organ damage and treatment effects from the previous admission might influence the transition probabilities in a subsequent ICU stay. In our dataset the patients who were readmitted to the ICU stayed longer at the ICU during the readmission (mean 8.4 vs. 7.1 days) and had a lower ICU mortality (19.4% vs. 23.0%). For these patients the models probably overestimate the probabilities of ICU discharge and ICU death, and underestimate the transition probabilities related to ICU stay and organ failure.

Third, in severe sepsis ICU patients various factors are known to be associated with a poor prognosis. We have included severity-of-illness at admission into the models; in a similar fashion other known predictive factors (e.g., pre-existing comorbidities, type and location of the infection [48,49]) could be included as covariables in the regression equations. However, this comes at the cost of expansion of the parameter space. Finally, our analysis was restricted to patients that were admitted to the ICU with severe sepsis. It is unclear to which extent our clinical findings generalize to patients who develop severe sepsis during ICU stay (as these patients are less severely ill [62]) or to the general ICU population. We do, however, have no reason to assume that the methodological approach cannot be generalized to other populations.

6. Conclusion

In this paper we have presented how complex multivariate Markov models can be induced from clinical data using hierarchical dynamic Bayesian networks and logistic regression modelling. In the application to data from the Intensive Care Unit we have shown that the models provide clinical insight into the dynamics of organ failure and how they can be used in daily clinical practice.

Acknowledgment

References

[1] Shahar Y. Dimensions of time in illness: an objective view. Ann Intern Med 2000;132:45–53.
[2] LeGall JR, Klar J, Lemeshow S, Saulnier F, Alberti C, Artigas A, et al. The Logistic Organ Dysfunction system. A new way to assess organ dysfunction in the intensive care unit. JAMA 1996;276:802–10.
[3] Vincent J-L, De Mendonca A, Cantraine F, Moreno R, Takala J, Suter PM, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Crit Care Med 1998;26:1793–800.
[4] Vosylius S, Sipylaite J, Ivaskevicius J. Sequential organ failure assessment score as the determinant of outcome for patients with severe sepsis. Croat Med J 2004;45:715–20.
[5] Shahar Y. A framework for knowledge-based temporal abstraction. Artif Intell 1997;90:79–133.
[6] Moreno R, Vincent J-L, Matos R, Mendonca A, Cantraine F, Thijs L, et al on behalf of the ESICM working group on sepsis related problems. The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study. Intensive Care Med 1999;25:686–96.
[7] Lopes Ferreira F, Peres Bota D, Bross A, Mélot C, Vincent J-L. Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA 2001;286:1754–8.
[8] Graf J, Janssens U. The early bird catches the worm. Crit Care Med 2005;33:2412–3 [editorial].
[9] Russell JA, Singer J, Bernard GR, Wheeler A, Fulkerson W, Hudson L, et al. Changing patterns of organ dysfunction in early human sepsis is related to mortality. Crit Care Med 2000;28:3405–11.
[10] Levy MM, Macias WL, Vincent J-L, Russell JA, Silva E, Trzaskoma B, et al. Early changes in organ function predict eventual survival in severe sepsis. Crit Care Med 2005;33:2194–201.
[11] Kayaalp M, Cooper G, Clermont G. Predicting with variables constructed from temporal sequences. In: Jaakkola T, Richardson T, editors. Proceedings of the eighth international workshop on artificial intelligence and statistics. Morgan Kaufmann; 2001. p. 220–5.
[12] Toma T, Abu-Hanna A, Bosman RJ. Discovery and inclusion of SOFA score episodes in mortality prediction. J Biomed Inform 2007;40:649–60.
[13] Beck JR, Pauker SG. The Markov-process in medical prognosis. Med Decis Making 1983;3:419–58.
[14] Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making 1993;13:322–38.
[15] Pérez-Ocon R, Ruiz-Castro JE, Gámiz-Pérez ML. A multivariate model to measure the effect of treatments in survival to breast cancer. Biom J 1998;40:703–15.
[16] Saint-Pierre P, Combescure C, Daures JP, Godard P. The analysis of asthma control under a Markov assumption with use of covariates. Stat Med 2003;22:3755–70.
[17] Kjaerulff U. dHugin: a computational system for dynamic time-sliced Bayesian networks. Int J Forecast 1995;11:89–111.
[18] Andreassen S, Riekehr C, Kristensen B, et al. Using probabilistic and decision-theoretic methods in treatment and prognosis modeling. Artif Intell Med 1999;15:121–34.
[19] Lucas PJF, de Bruijn NC, Schurink K, Hoepelman A. A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU. Artif Intell Med 2000;19:251–79.
[20] Charitos T, van der Gaag LC, Visscher S, et al. A dynamic Bayesian network for diagnosing ventilator-associated pneumonia in ICU patients. Working notes of the workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP); 2005. p. 32–7.
[21] Visscher S, Lucas P, Flesch I, Schurink K. Using temporal context-specific independence information in the exploratory analysis of disease processes. In: Bellazi R, Abu-Hanna A, Hunter J, editors. Proceedings of the 11th conference of Artificial Intelligence in Medicine (AIME). 2007. Lecture Notes in Artificial Intelligence 4594. Springer-Verlag; 2007. p. 87–96.
[22] Riva A, Bellazzi R. Learning temporal probabilistic causal models from longitudinal data. Artif Intell Med 1996;8:217–34.
[23] Kayaalp MM. Learning dynamic Bayesian network structures from data. PhD Thesis. University of Pittsburgh; 2003.
[24] Vincent JL, Sakr Y, Sprung CL, Ranieri VM, Reinhart K, Gerlach H, et al on behalf of the Sepsis Occurrence in Acutely Ill Patients (SOAP) investigators. Sepsis in European Intensive Care Units: results of the SOAP study. Crit Care Med 2006;34:344–53.
[25] Angus DC, Linde-Zwirble WT, Lidicker J, et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med 2001;29:1303–10.
[26] Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. N Engl J Med 2003;348:1546–54.
[27] Singh S, Evans TW. Organ dysfunction during sepsis. Intensive Care Med 2006;32:349–60.
[28] Pittet D, Thiévent B, Wenzel RP, Li N, Auckenthaler R, Suter PM. Bedside prediction of mortality from bacteremic sepsis. A dynamic analysis of ICU patients. Am J Respir Crit Care Med 1996;153:684–93.
[29] Knaus WA, Draper EA, Wagner DP, Zimmerman JE. Prognosis in acute organ-system failure. Ann Surg 1985;202:685–93.
[30] Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Chest 1992;101:1644–55.
[31] Bernard GR, Vincent JL, Laterre PF, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. N Engl J Med 2001;344:699–709.
[32] LeGall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS) II based on a European/North American multicenter study. JAMA 1993;270:2957–63.
[33] Arts DGT, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc 2002;9:600–11.
[34] Saka G, Kreke JE, Schaefer A, Chang C-CH, Roberts MS, Angus DC. Using dynamic microsimulation to predict disease progression in patients with pneumonia-related sepsis. Crit Care 2007;11:R65.
[35] Kayaalp MM, Cooper GF, Clermont G. Predicting ICU mortality: a comparison of stationary and non-stationary models. In: Proceedings AMIA symposium; 2002. p. 418–22.
[36] Gyftodimos E, Flach P. Hierarchical Bayesian networks: a probabilistic reasoning model for structured domains. In: Proceedings of the ICML-2002 workshop on development of representations; 2002. p. 23–30.
[37] Provan G, Clarke JR. Dynamic network construction and updating techniques for the diagnosis of acute abdominal pain. IEEE Trans Pattern Anal Mach Intell 1993;15:299–307.

[38] Dagum P, Galper A, Horvitz E. Dynamic network models for forecasting. In: Proceedings of the eighth annual conference on Uncertainty in Artificial Intelligence (UAI-92). San Mateo, CA: Morgan Kaufmann; 1992. p. 41–8.

[39] Dagum P, Galper A. Forecasting sleep apnea with dynamic network models. In: Proceedings of the 10th annual conference on Uncertainty in Artificial Intelligence (UAI-93). Palo Alto: Morgan Kaufmann Publishers; 1993. p. 64–71.

[40] Agresti A. Categorical data analysis. Wiley series in probability and statistics. 2nd ed. Hoboken, New Jersey: John Wiley & Sons Inc.; 2002.

[41] Venables WN, Ripley BD. Modern applied statistics with S. Springer series in statistics and computing. 4th ed. New York: Springer-Verlag; 2002.

[42] Hosmer DW, Lemeshow SL. Applied logistic regression. Wiley series in probability and statistics. New York: John Wiley & Sons Inc.; 1989.

[43] Berzuini C, Bellazzi R, Spiegelhalter D. Bayesian networks applied to therapy monitoring. In: Proceedings of the seventh annual conference on Uncertainty in Artificial Intelligence (UAI-91). Morgan Kaufmann; 1991. p. 35–43.

[44] Dagum P, Chavez RM. Approximating probabilistic inference in Bayesian belief networks. IEEE PAMI 1993;15:246–55.

[45] Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.

[46] McCarthy ML, Zeger SL, Ding R, Levin SR, Desmond JS, Lee J, et al. Crowding delays treatment and lengthens emergency department length of stay, even among high-acuity patients. Ann Emerg Med 2009 [Epub].

[47] Vincent JL, Lopes Ferreira F. Evaluation of organ failure: we are making progress. Intensive Care Med 2000;26:1023–4.

[48] Moreno RP, Metnitz B, Adler L, Hoechtl A, Bauer P. Metnitz PGH for the SAPS 3 investigators. Sepsis mortality prediction based on predisposition, infection and response. Intensive Care Med 2008;34:496–504.

[49] Alberti C, Brun-Buisson C, Goodman SG, Guidici D, Granton J, Moreno R, et al. Le Gall JR for the European Sepsis Group. Influence of systemic inflammatory response syndrome and sepsis on outcome of critically ill infected patients. Am J Respir Crit Care Med 2003;168:77–84.

[50] Gaziano TA, Steyn K, Cohen DJ, Weinstein MC, Opie LH. Cost-effectiveness analysis of hypertension guidelines in South Africa; absolute risk versus blood pressure level. Circulation 2005;112:3569–76.

[51] Magni P, Quaglini S, Marchetti M, Barosi G. Deciding when to intervene: a Markov decision approach. Int J Med Inform 2000;60(3):237–53.

[52] Kapadia AS, Chan W, Sachdeva R, Moye LA, Jefferson LS. Predicting duration of stay in a pediatric intensive care unit: a Markovian approach. Eur J Oper Res 2000;124:353–9.

[53] Pérez A, Chan W, Dennis RJ. Predicting the length of stay of patients admitted for intensive care using a first step analysis. Health Serv Outcomes Res Method 2006;6:127–38.

[54] Bäuerle R, Rücker A, TSchmandra TC, Holzer K, Encke A, Hanisch E. Markov cohort simulation study reveals evidence for sex-based risk difference in intensive care unit patients. Am J Surg 2000;179:207–11.

[55] Ackerman E. Simulation of micropopulations in epidemiology. Tutorial I. Simulation: an introduction. Int J Biomed Comput 1994;36:229–38.

[56] Gerlach H, Toussaint S. Between prediction, education and quality control: simulation models in critical care. Crit Care 2007;11:146 [editorial].

[57] Clermont G, Kalplan V, Moreno R, Vincent J-L, Linde-Zwirble WT, Van Hout B, et al. Dynamic microsomulation to model multiple outcomes in cohorts of critically ill patients. Intensive Care Med 2004;30:2237–44.

[58] Derksen S, Keselman HJ. Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. Br J Math Stat Psychol 1992;45:265–82.

[59] Harrell Jr FE, Lee KL, Mark DB. Tutorial in biostatistics. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361–87.

[60] Friedman N, Goldszmidt M. Learning Bayesian networks with local structure. In: Jordan MI, editor. Learning and inference in graphical models. Dordrecht, The Netherlands: Kluwer; 1998.

[61] Dellinger RP, Carlet JM, Masur H, et aland the Surviving Sepsis Campaign Management Guidelines Committee. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. Crit Care Med 2004;32(3):858–73.

[62] Damas P, Ledoux D, Nys M, Monchi M, Wiesen P, Beauve B, et al. Intensive care unit acquired infection and organ failure. Intensive Care Med 2008;34:856–64.