Prognosis in intensive care: inductive methods using sequential patterns of organ dysfunction scores
Toma, T.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Summary

This thesis investigates novel methods for predicting the survival of patients admitted to the Intensive Care Unit (ICU). It presents inductive methods, stemming from the fields of data mining and statistics, that use patterns of daily organ dysfunction for predicting the vital status of patients at hospital discharge. Mortality prediction models in the ICU are used to perform case-mix adjustment in assessments of care quality, to inform treatment and end-of-life decisions, to inform patients and their families, and to select patients for inclusion in clinical trials.

The introduction of computerized acquisition of physiological data at the bedside has allowed for collecting repeated measurements for patients over time during their stay in the ICU. The type of such “temporal” data that we use in this dissertation is the Sequential Organ Failure Assessment (SOFA) score. The SOFA score is calculated daily from physiological patient data and quantifies the extent of dysfunction for six major organ systems (the respiratory, coagulation, hepatic, circulatory, central nervous and renal system). Higher SOFA score values indicate more severe organ dysfunction or even organ failure (OF). Changes in the overall SOFA score and in the subscores associated with the individual organ systems reflect deterioration or recovery of the patient’s condition over time.

Although the SOFA score was designed to communicate about and report on organ dysfunction in individual patients and patient cohorts, many studies have investigated its utility for predicting ICU and hospital death. Most of these studies use the number of organs failing at the day of prediction, or temporal summaries such as the mean SOFA
score over the patient’s ICU stay thus far, the maximum SOFA thus far, or the difference between the maximum score and the score at the day of ICU admission. No previous studies however have investigated the prognostic value of patterns of organ failure scores.

The methods described in this thesis use data mining algorithms to identify frequent patterns of organ dysfunction. Subsequently, these patterns are included as predictors of mortality in logistic regression models. We hypothesized that these patterns would improve the quality of predictions in comparison to existing prognostic models that only use data from the day of ICU admission.

Our research questions in this thesis are:

1. **How can sequential organ function data be integrated, in a symbolic representation, into models for predicting hospital mortality of ICU patients?**

2. **How can one systematically evaluate methods for prediction, correcting for sources of variation in the model construction process?**

3. **What is the increase in predictive performance when patterns of sequential organ failure assessment score are incorporated into prognostic models, and how do these models compare to the predictive abilities of ICU physicians?**

In Chapter 2 we propose a novel method for mortality prediction that, in addition to data recorded at ICU admission, takes advantage of daily recorded SOFA data as well. The method is characterized by the data-driven discovery of frequent patterns in subsequent, day-to-day SOFA scores (each of which is the sum of six organ dysfunction subscores). These patterns are then represented as binary variables and included as candidate predictors in logistic regression models for predicting the probability of death. Each binary variable indicates the presence or absence of a specific pattern in a patient’s sequence of scores. Our method results in a set of logistic regression models, one for each subsequent day of Intensive Care Unit stay. We evaluated our method in a large dataset of ICU patients from a Dutch teaching hospital (N=6,865) by developing and validating models for the first five days of ICU stay. We compared the five resulting pattern-based models
with ones that only used SAPS-II, which is a score of severity-of-illness at ICU admission. Our results demonstrate that the new models had more accurate predictions on each of the five days, as indicated by Brier scores.

The work described in Chapter 3 extends the method in the previous chapter. In particular, we investigated the added prognostic value residing in frequent patterns obtained from the six individual organ dysfunction scores. Application of this method to a dataset of 2,928 patients from the same hospital showed that model performance improved both in terms of discrimination (using the Area Under the ROC curve, AUC) and accuracy (Brier score).

Chapter 4 addresses multiple organ failure (MOF), i.e., the concurrent failure of multiple organ systems during the same day. Although MOF in general is associated with increased mortality risk, little is known about prevalence of multiple organ failure and its association with mortality. The objective of this study was to identify frequent single and multiple organ failures occurring at the ICU, and investigate their utility to predict risk of death. Based on daily SOFA scores, frequently occurring organ failure patterns were identified. These patterns were subsequently used as candidate covariates in mortality prediction models for the first seven days of ICU stay. Resampling methods were used to select the most predictive patterns for the final models and to evaluate the prognostic performance of the resulting models. We found that replacing the daily number of failing organ systems by specific patterns of frequently occurring organ failures in prognostic models yields a small but statistically significant increase in predictive performance. Patterns of multiple organ failure were found to have better predictive value than single organ failure patterns in terms of both accuracy and discrimination.

While the evaluation of models is important for indicating how well the model in question performs on new data, evaluating inductive methods is equally important as it indicates how consistent a method produces accurate models. Chapter 5 addresses the evaluation and comparison of inductive methods. To account for the inherent sources of variation implied by the experimental design, especially the random selection of develop-
ment and validation datasets, we used a variation of bootstrapping. The purpose of this approach was to allow the execution of the learning algorithm multiple times in order to measure the variation in accuracy of the resulting models. For each bootstrap sample, several competing methods were applied including all their elements such as frequent pattern identification and variable selection. Subsequently, each generated model was evaluated on a test set specific to that bootstrap sample. Finally the results were combined in order to assess and compare the performance of each of the methods. An application of this approach to a real world dataset, previously used for the analysis presented in Chapter 3, showed that an inductive method based on pattern discovery outperformed a standard method that did not use the patterns. In addition we illustrated that this approach provides insights into the discovered patterns, such as the likelihood that a pattern would be selected in a model generated by the method.

Chapter 6 presents a prospective evaluation of our pattern-based mortality prediction models and a comparison with the predictive abilities of ICU physicians. Specifically, we developed prognostic models for day 2 to day 7 of ICU stay by data-driven discovery of patterns of SOFA scores and embedded the patterns in three types of logistic regression models. Type A models included the SAPS-II severity of illness score at ICU admission and the SOFA patterns. Type B models added to these covariates the mean, max and delta (increments) of SOFA scores. Type C models included, in addition, the mean, max and delta in physicians’ predictions of mortality. Type C models outperformed Type A models, showing that physicians had better discriminative abilities than the pattern-based models while demonstrating comparable accuracy. However combining both sources of predictions, in type C models, resulted in superior discrimination as well as accuracy. We conclude that pattern-based models and physicians draw on complementary information that can be best harnessed by combining both sources.

The answers to our research questions are summarized in Chapter 7. There we also indicate strengths and limitations of the current research and propose future work. Patterns presented in this thesis have two main benefits: (1) they provide a symbolic representation
which largely preserves the temporal information; and (2) they can be easily included in
prognostic models as indicator variables. The new methods for prediction also work in
practice, as demonstrated by their measured performance on various case-studies, relative
to existing prognostic methods and physicians’ estimates. Further work should address
more expressive patterns, new methods exploiting their prognostic value, and their impact
on making treatment decisions in the ICU.