Discovery and Integration of Organ-Failure Episodes in Mortality Prediction

Tudor Toma[1], Ameen Abu-Hanna[1], and Robert-Jan Bosman[2]

[1] Academic Medical Center, Universiteit van Amsterdam, Department of Medical Informatics, P.O. Box 22700, 1100 DE Amsterdam, The Netherlands
[2] Department of Intensive Care, Onze Lieve Vrouwe Gasthuis, 1e Oosterparkstraat 279, P.O. box 10550, 1090 HM Amsterdam, The Netherlands

Abstract. Current predictive models in the intensive care rely on summaries of data collected at patient admission. It has been shown recently that temporal patterns of the daily Sequential Organ Failure Assessment (SOFA) scores can improve predictions. However, the derangement of the six individual organ systems underlying the calculation of a SOFA score were not taken into account, thus impeding the understanding of their prognostic merits. In this paper we propose a method for model induction that integrates in a novel way the individual organ failure scores with SOFA scores. The integration of these two correlated components is achieved by summarizing the historic SOFA information and at the same time by capturing the evolution of individual organ system failure status. The method also explicitly avoids the collinearity problem among organ failure episodes. We report on the application of our method to a large dataset and demonstrate its added value. The ubiquity of severity scores and sub-scores in medicine renders our approach relevant to a wide range of medical domains.

Keywords: Prognostic models, temporal patterns, Intensive Care, organ failure scores.

1 Introduction

Probabilistic predictions of patient outcomes such as mortality and length of stay in the intensive care unit (ICU) are useful for supporting decisions at the level of individuals and groups [1]. Current models for predicting hospital mortality, after admission to the ICU, use summaries of patient information collected within the first 24 hours of admission. These summaries, which take the form of severity-of-illness-scores such as the APACHE-II [2] and SAPS-II [3], are used as covariates in a logistic regression model (see appendix).

Since a decade ago, some ICUs started collecting Sequential Organ Failure Assessment (SOFA) scores [4] *on each day* of ICU stay. A SOFA score is an integer ranging from 0 to 24 that quantifies the derangement of *all* organs of a patient on a given day, the higher the score the greater the derangement. A SOFA score is calculated as the sum of 6 individual organ system failure (IOSF) scores, each ranging between 0 and 4.

R. Bellazzi, A. Abu-Hanna, and J. Hunter (Eds.): AIME 2007, LNAI 4594, pp. 97–106, 2007.

Although not specifically targeted towards prediction of mortality, the relationship between SOFA scores and mortality has been investigated. In previous work [5] we devised a new method for integrating the SOFA temporal information in the existing logistic regression models. The method, more elaborated on in the next section, is based on the idea of using frequent temporal patterns, called episodes, as covariates in the model. Although the SOFA episodes improved predictions, the use of only SOFA scores has two disadvantages. First, no insight is obtained into the qualitative contribution of the individual organ systems to mortality. Second, it is unclear whether the IOSF scores would further improve the quality of predictions because these scores are *correlated* with the SOFA scores and it is unclear how to combine the two.

In this paper we propose a method for model induction that incorporates IOSF scores alongside the SOFA scores. The method deals with the overlap between the two types of scores by summarizing the historic SOFA information in one summary statistic, and by capturing the evolution of individual organ system failure status in frequent temporal patterns. The summary statistic and the organ failure (OF) episodes are used as covariates in the familiar logistic regression framework. For a given day d, the application of the proposed method results in a model predicting, for patients staying at least d days, the probability of their eventual survival status at discharge (regardless of when this happens). We report on the application of our method to a large real-world dataset and demonstrate the added value in interpreting the models and in their improved predictive performance. In the sequel we will refer to a model using only the SAPS-II (in short SAPS) as the *static model*; to a model using SAPS and SOFA episodes as a *SOFA-model* (as described in [5]); and to a model using SAPS, a summary of SOFA and failure episodes as an organ-failure model (*OF-model*). The resulting OF-models will be subject to comparison with the other models.

The rest of the paper is organized as follows. Section 2 describes the proposed method to induce OF-models and the data types it operates on. Section 3 and Section 4 describe the case study used for demonstrating the method and the obtained results. We discuss our method in Section 5 and put it in perspective by relating it to other work.

2 Data and Methods

Data. We consider two categories of data: the static data, represented by the SAPS score (collected at admission) and temporal data consisting of the daily SOFA score along with its 6 components (IOSF scores) corresponding to the following systems: respiratory (Resp), coagulation (Coag), hepatic (Hepa), cardiovascular (Cardio), central nervous system (CNS), and renal (Ren) systems. Table 1 shows an example of data for a patient who stayed for 4 days in the ICU before dying on the fifth day.

Method. In previous work [5] we showed how to induce SOFA-models. In a nutshell, this is done by the following process. First the SOFA scores, ranging from 0 to 24, are categorized into three qualitative states: Low (L), Medium (M)

Table 1. Example of available temporal data for an ICU patient admitted for 4 days. The SOFA scores indicate a constant health status deterioration.

Day	SOFA	Resp	Coag	Hepa	Cardio	CNS	Ren	Outcome
1	10	4	2	0	0	1	3	
2	12	4	1	2	1	2	2	
3	14	4	2	2	0	4	2	
4	15	4	1	2	1	4	3	
5	–	–	–	–	–	–	–	died

and High (H). For each day d on which hospital mortality is to be predicted the subsample of patients that stayed at least d days is selected. Next, frequent episodes of consecutive SOFA states that are aligned with the day of prediction (later clarified in this paper) are discovered in these patients. The SAPS and a set of binary variables representing the occurrence of the SOFA episodes in patients are then considered as possible covariates (input variables) in a logistic regression model to predict the hospital mortality for day d. For example if the linear predictor LP (see appendix) of the model for day 5 is: $-2 + 0.02SAPS - 1.5LL + 0.7H$ then for a patient with SAPS of 40 having the episode $\{L, L\}$ at days 4 and 5 will be $-2 + 0.02 * 40 - 1.5 = -2.7$ which corresponds to a probability of dying of 0.063 while a patient with SAPS of 40 but having the episode $\{H\}$ on day 5 will have an LP of $-2 + 0.02 * 40 + 0.7 = -0.5$ which corresponds to a probability of death of 0.38. In this paper we adapt and extend our approach described above to induce OF-models. This process is described below followed by a description of the main differences between the new and previous approach.

Categorization. An IOSF score ranging between 0 to 4 is categorized based on clinical definitions into two categories: *failure* (F), for values $\in \{3, 4\}$ and *non–failure* (NF) otherwise. For example, the renal scores during 3 days of 1–4–2 become NF, F, NF. Aside from clinical interpretability, limiting the number of categories allows the emergence of episodes with higher support in the data.

Frequent episode discovery. We rely on the A-priori-like algorithm [6] described in [7] for frequent pattern discovery. This is based on the *downward closure* property which implies that a subsequence of a frequent episode must be frequent too. The discovery procedure is an iterative process. We adapted the algorithm to search for a special type of episodes: their occurrence in a patient's sequence of values is *consecutive* and also *aligned* to the day of prediction d. For example, given the patient's sequence F, F, NF, F, NF starting at admission day, then for d=2 the episode $\{F, F\}$ occurs in the patient data because aligning the episode at day 2 (i.e. positioning the last F in the episode at the second element in the patient's sequence) results in a match with the subsequence F, F in the patient sequence. However, for d=4 the episode is not aligned to the patient's sequence. The decision to use aligned episodes is motivated by the belief that the last days before prediction are more relevant than information at earlier days.

In each iteration, the algorithm extends frequent episodes from the previous iteration with elementary episodes (F and NF for the organ failure data) and assesses the frequency of the resulting episodes. For example given the frequent episode $\{F, F, NF\}$, the extended aligned candidate episodes are $\{F, F, F, NF\}$ and $\{NF, F, F, NF\}$. In general, an episode aligned to day d is said to be frequent when its frequency rate in the subset of patients staying at least d days exceeds a pre-specified threshold (e.g. 5% of cases) referred to as minimum support rate. The discovery process continues until no more frequent episodes are encountered.

Model fitting strategy. Not all the frequent OF episodes are relevant for prediction and their excessive use can lead to overfitting. Our feature selection strategy is based on an information-theoretic measure, the Akaike's Information Criterion (AIC) [8] used in an iterative backward variable elimination selection process. In every iteration, the current model with N variables is used to produce N models, each having one subset of $N-1$ distinct variables. From the produced models only the one that further reduces, by largest margin, the AIC of the model with N variables is considered for the next iteration. The AIC, defined as $-2logL(\theta) + 2k$, where L(θ) is the maximized likelihood [9] of the model and k the number of parameters, strikes a balance between likelihood and parsimony. Use of an information-theoretic criterion mitigates the problems associated with approaches based on significance testing [10]. Finally, we use background medical knowledge to eliminate model coefficients not compliant to clinical expectations. In particular we: (1) eliminate any episode with "failure" at the day of prediction and a negative β coefficient in the model (e.g. $\beta = -0.7$ for $\{NF, F, F\}$) and (2) eliminate any episode with "non-failure" at the day of prediction and a positive β coefficient in the model (e.g. $\beta = 1.1$ for $\{NF, F, NF\}$). Keep in mind that a negative coefficient reduces the probability of mortality, and a positive one increases it. A similar idea was introduced in [11], under the name "sign OK", defining a variable selection strategy based on the plausibility of the sign of the regression coefficient.

Another thorny issue requiring attention is the phenomenon of *collinearity*, a situation in which at least one of the covariates can be predicted well from the other covariates [10]. This leads to instability of the model and jeopardizes the interpretability of the β coefficients in the logistic model (see appendix) since it is based on the idea of studying a change of a covariate while fixing the others. However, holding down the values fixed of the collinear covariates is unattainable because, by definition, they will be affected. One strong type of collinearity which is ubiquitous in our domain when dealing with aligned episodes, is the occurrence of the *logically entailed* episodes [5]. For example we say that episode $\{NF, F, F\}$ logically entails episode $\{F, F\}$ since the occurrence of the first in a patient implies the occurrence of the second episode. To eliminate logical entailment we included a ranking step in the modeling stage (this procedure is more stringent but simpler than the one suggested in [5]). For each one of the six organ systems, all its discovered frequent OF episodes are ranked, from those with smallest (best) AIC value to the largest, based on a univariate analysis between mortality and the episode. For each organ system we retain only its highest ranked episode.

This eliminates logically entailed episodes and provides simple models. This risks eliminating other possibly useful episodes, but with only 2 categories (F and NF) any two aligned episodes are at least partially correlated. The episodes obtained in this manner are then fed into the AIC-based feature elimination strategy described above.

Evaluation. For each day of prediction d a separate training and testing set are created. An important performance aspect of a probabilistic model is its calibration ability. We applied the commonly used Brier score, $\frac{1}{N}\sum_{i=1}^{N}(P(Y_i = 1 \mid \mathbf{x}_i) - y_i)^2$, where N denotes the number of patients, and y_i denotes the actual outcome for patient i. The vector $\mathbf{x}_i$ represents the covariate values for patient i. The Brier score is a *strictly proper scoring rule* [12] which, unlike measures like the area under the ROC curve, means it is optimal only when the true probability of the event is provided. The performance of each of the OF-models is compared to its corresponding SOFA and static models. To test for statistical significance in performance difference we advocate the use of the non-parametric bootstrap method [13] with 1000 bootstrap samples of differences.

3 Case Study

The ICU patient dataset is available from the OLVG, a teaching hospital in Amsterdam and was collected during July 1998 and October 2006 including all 2785 patients (25% mortality) eligible for analysis [5]. Both SAPS and SOFA scores values were larger in the non-survivors (averages are: SAPS 61±15.3 vs. 39±18.4 for survivors, SOFA: 9.7±3.2 vs. 7.3±2.6). The mean number of failures (IOSF scores values $\in \{3, 4\}$) per patient, give a clear indication of the high association between organ failure and survival outcome (9.8 organ failures in non-survivors versus 4.4 organ failures in survivors).

4 Results

Based on the method described above, four OF-models corresponding to the ICU days 2–5 (day 1 cannot show temporal evolution), were created for predicting the hospital mortality. In episode discovery, a threshold of 5% was used for minimum support rate. Each OF-model includes the SAPS covariates (SAPS, log(SAPS+1)) and, potentially, after variable selection the average SOFA and frequent OF episodes. For comparative purposes the same training set was used to induce the static and SOFA-models for the given day. Table 2 shows the resulting models described by their's linear predictor (LP). logSAPS represent $log(SAPS + 1)$ (used in compliance with the original SAPS model). The organ failure episodes are labeled to identify their type of organ system. For example Resp$\{F, NF\}$ represents a failure followed by a non-failure in the respiratory system. The SOFA-models use the elements $\{L, M, H\}$ to describe frequent SOFA episodes e.g. $\{HM\}$. Table 3 exemplifies the interpretation in terms of odds-ratios (equal to $exp(\beta)$) of the OF-model coefficients for day 2 and 5. For

Table 2. Temporal models (OF and SOFA) and static models for days 2–5 of ICU stay described by their linear predictors (LPs)

Day	OF–model LP	SOFA–model LP	Static model LP
2	-9.3 +0.005SAPS + 1.9logSAPS+0.065meanSOFA -1.85Resp{F,NF} +1.1CNS{F,F}	-5.9 +0.03SAPS + 0.87logSAPS +0.6H -0.7L	-7.7 +0.036SAPS +1.26logSAPS
3	-10.8 -0.01SAPS +2.2logSAPS +0.13meanSOFA +0.4Resp{F,F} +1.1CNS{F}	-7.26 +0.01SAPS +1.4logSAPS+ 1.1H -0.66L	-10.35 +0.02SAPS +2.2logSAPS
4	-6.7 -0.006SAPS +1.9logSAPS +0.45Resp{F,F,F,F} +1.56CNS{F} -0.62Hepa{NF,NF,NF,NF} -0.8Cardio{NF} -0.8Ren{NF,NF,NF}	-5.5 +0.014SAPS +1.18logSAPS -1.95L -0.83MM -0.65HM	-7.88 +0.027SAPS +1.42logSAPS
5	-6 -0.006SAPS +1.5logSAPS +0.5Resp{F,F,F,F,F} -0.9Coag{NF,NF} +1.4CNS{F} -0.5Ren{NF,NF,NF,NF}	-2.5 +0.02SAPS +0.12logSAPS -1.04L +0.65H	-5.5 +0.02SAPS +0.85logSAPS

Table 3. Model covariates, their coefficients and odds-ratios ($exp(\beta_i)$) in the OF-models for day 2 and 5 of ICU stay

Day	Covariate	β	e^{β}	Day	Covariate	β	e^{β}
2	SAPS	0.005	1.005	5	SAPS	-0.006	0.99
	logSAPS	1.9	6.68		logSAPS	1.5	4.48
	meanSOFA	0.065	1.07		Resp{F,F,F,F,F}	0.5	1.64
	Resp{F,NF}	-1.85	0.16		Coag{NF,NF}	-0.9	0.4
	CNS{F,F}	1.1	3		CNS{F}	1.4	4.05
					Ren{NF,NF,NF,NF}	-0.5	0.6

Table 4. Performance evaluation – Brier score

Day	Brier score			OF–model win		SOFA–model win
	OF	SOFA	SAPS	vs. SAPS	vs. SOFA	vs. SAPS
2	0.157	0.158	0.163	Yes	Yes	Yes
3	0.179	0.185	0.197	Yes*	Yes	Yes*
4	0.199	0.209	0.212	Yes	Yes	Yes
5	0.186	0.189	0.207	Yes*	Yes	Yes*

example, for day 2, after adjusting for the other variables, the odds of dying for patients with the episode CNS$\{F, F\}$ is three times the odds for those without it. The OF-models shown in Table 2 where evaluated on an independent test set for day 2 till day 5 of ICU stay and compared to the static and SOFA-models. The Brier scores (the lower the better) are shown in Table 4. An * indicates a statistically significantly better Brier score than the static model.

5 Discussion and Related Work

In this section we discuss the results, our approach in relation to others, delineate further work, and conclude the paper.

Results. Table 4 ascertains that the OF episodes can improve predictions: the OF-models performed better than the SOFA-models on all days. Also, both kinds of temporal models (SOFA and OF) were consistently better (sometimes with statistically significant differences) than the traditional static model (SAPS model). This evidence needs of course corroboration by a more stringent cross-validation design that we plan to do in the future. The results also show the usefulness of the coefficient qualitative and quantitative interpretations of organ derangement (see Table 2). In all days the central nervous and respiratory systems were present in the models. The renal organ system was the next best predictor included in two models. In related work, when the central nervous system was considered for analysis [14,15,16] it was indeed a good discriminator of mortality, otherwise, as in [17,18,19], the cardiovascular system emerged as a strong predictor. When considering the frequent episodes selected we note that those denoting constant organ conditions (failure e.g. $\{F, F\}$ or non-failure $\{NF, NF, NF, NF\}$) were dominant. Similar findings about the "constant patterns" have been reported by [20]. We hypothesize that their dominance is rooted in the high support they enjoy in the data: individual scores are not likely to change often between the only two categories discerned (Non-failure, Failure).

Table 3 can be used to provide quantitative interpretations. For example, in the model for day 5 the central nervous system episode $\{F\}$ is associated with the odds-ratio of about 4: the odds of dying given a failure of the central nervous system failing on day 5 (day of prediction) is about 4 times the odds of dying if the central nervous system did not fail on that day. By the same token, in case of non-failure at the day of prediction an odds-ratio < 1 (corresponding to a negative coefficient in the LP) indicates a beneficial effect on the prognosis. This holds for example for the renal episode $\{NF, NF, NF, NF\}$.

Another finding is that, starting from day 4 of prediction, the models do not include the mean SOFA score anymore. A possible reason is that the importance of the latest days is diluted in the *unweighted* SOFA score mean by the early days which might be less important. Future work consists of using a weighted summary of the SOFA scores where later days enjoy more weight.

Approach, related and future work. The main differences between the OF-model induction approach and the one described in [5] to induce SOFA-models, aside from applying it to other data types, are the following. First, for the induction of OF-models, SOFA scores are aggregated in one summary statistic in order to avoid major overlap with the IOSF scores. Second, each of the 6 organ systems is categorized using clinical knowledge corresponding to organ failure. Third, to avoid strong collinearity between the organ failure episodes we follow a new stringent feature selection strategy based on ranking. Fourth, clinical knowledge is infused in the covariate selection process by using the "sign OK" rule.

The *categorization* we adopted of the IOSF scores in 2 categories can be characterized as vertical (contemporaneous) data abstraction [21] or State Temporal Abstractions [22]. The resulting categories are not necessarily the most useful ones in prediction. Future work will include the use of the outcome (mortality),

e.g. by using entropy-based categorization [23], or additional medical knowledge [18] to generate more predictive categories.

Our frequent episodes are discovered in a *data-driven* manner. In using these episodes we also apply various *non-stationarity* assumptions: the episodes are *aligned* to the prediction day, and can be of *different* lengths. These two properties distinguish our work from the work appearing in the intensive care literature in which pre-specified summaries (e.g. maximum value, average value during ICU stay) or qualitative trends of IOSF scores are used [16,18]. This is also in contrast to the methodological work described in [24] which assumes a strict form of stationarity where the mean value of adverse events (highly abnormal values of medical measurements in a patient) is used to summarize the temporal process.

Valuable future work consists of investigating more expressive episodes like those described in [25] where a pattern includes a multivariate set of attributes. This will allow one to capture complex interactions between the IOSF scores in one pattern instead of having various univariate episodes as describes in this paper.

In [20] patterns similar in nature to ours are discovered not based on frequency but on their discriminative capabilities (Area Under the Curve) and forms an interesting future work. The most predictive ones are then used in a Naive Bayes Classifier method. Given the nature of our episodes, the Naive Bayes approach combined with an assessment geared towards discriminatory model capabilities does not provide incentive to predict reliable probabilities. This is because our particular episodes are correlated and because of the overlap between SOFA and IOSF information, clearly violating the conditional independence assumption used in the Naive Bayes approach.

Making use of logistic regression allows a fair comparison between our method and that currently used in the ICU. It also generates coefficients with meaningful interpretation as demonstrated above. These coefficients' interpretability is enhanced by the application of a logical entailment elimination step by means of the AIC criterion avoiding drawbacks related to the p-value based variable selection approaches. In [24] a comparison of various methods showed that in a given setting, different than ours, the logistic regression model was slightly outperformed only by the neural network model in terms of accuracy. Whether neural networks will lead to better calibrated results in our case is unclear, but if one is also interested in the interpretability of the models the logistic regression is a good choice. Future work could investigate how hybrid methods can be employed. For example in [26] a classification tree based on baseline information was used to stratify patients into homogeneous subgroups, then on each of these subgroups a logistic regression model was fit to predict mortality. A similar idea could be applied for temporal data.

In sum, this paper proposed a method for inducing predictive models based on the integration of the information residing in baseline data with the temporal information concerning organ system functioning into the logistic regression framework. The results are encouraging as the temporal organ failure episodes improve predictions' quality and interpretability. The ubiquity of scoring systems in

medicine for baseline and repeated measurements suggests the wider applicability of our approach to other domains.

Acknowledgments. This work was supported by the Netherlands Organization for Scientific Research (NWO) under the I-Catcher project number 634.000.020. Thanks for useful feedback are also due to Arno Siebes from the University of Utrecht and to the other members of the Methodology in Medical Informatics group in our department.

References

1. Abu-Hanna, A., Lucas, P.J.F.: Editorial: Prognostic models in medicine - AI and statistical approaches. Methods of Information in Medicine 40, 1–5 (2001)
2. Knaus, W., Draper, E., Wagner, D., Zimmerman, J.: APACHE II: a Severity of Disease Classification System. Crit. Care Med. 13, 818–829 (1985)
3. Le Gall, J., Lemeshow, S., Saulnier, F.: A new Simplified Acute Physiology Score (SAPS-II) based on a European/North American multicenter study. The Journal of the American Medical Association 270, 2957–2963 (1993)
4. Vincent, J.-L., Ferreira, F.L.: Evaluation of organ failure: we are making progress. Intensive Care Med. 26, 1023–1024 (2000)
5. Toma, T., Abu-Hanna, A., Bosman, R.J.: Discovery and Inclusion of SOFA Score Episodes in Mortality Prediction. Journal of Biomedical Informatics (to appear)
6. Agrawal, R., Srikant, S.: Fast algorithms for mining association rules. In: Proc. of the 20th Very Large Data Bases Conf., pp. 487–499 (1994)
7. Mannila, H., Toivonen, H., Verkamo, A.I.: Discovering frequent episodes in sequences. Data Mining Knowledge Discovery 1(3), 259–289 (1997)
8. Burnham, K.P., Anderson, D.R.: Model Selection and Multimodel Inference: A Practical-Theoretic Approach, 2nd edn. Springer, New York (2002)
9. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning Data Mining, Inference, and Prediction. Springer Series in Statistics (2001)
10. Harrell Jr., F.E.: Regression Modeling Strategies. 1st edn. Springer Series in Statistics (2001)
11. Steyerberg, E.W., Eijkemans, M.J., Harrell Jr., F.E., Habbema, J.D.: Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat. Med. 19(8), 1059–1079 (2000)
12. Hand, J.D.: Construction and Assessment of Classification Rules. John Wiley & Sons, Chichester (1997)
13. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Chapman & Hall, New York (1993)
14. Peres Bota, D., Melot, C., Lopes Ferreira, F., Nguyen Ba, V., Vincent, J.L.: The Multiple Organ Dysfunction Score (MODS) versus the Sequential Organ Failure Assessment (SOFA) score in outcome prediction. Intensive Care Med. 28(11), 1619–1624 (2002)
15. Pettila, V., Pettila, M., Sarna, S., Voutilainen, P., Takkunen, O.: Comparison of multiple organ dysfunction scores in the prediction of hospital mortality in the critically ill. Crit. Care Med. 30(8), 1705–1711 (2002)
16. Russell, J.A., et al.: Changing pattern of organ dysfunction in early human sepsis is related to mortality. Crit. Care Med. 28, 3405–3411 (2000)

17. Zygun, D.A., Kortbeek, J.B., Fick, G.H., Laupland, K.B., Doig, C.J.: Non-neurologic organ dysfunction in severe traumatic brain injury. Crit. Care Med. 33(3), 654–660 (2005)
18. Levy, M.M., Macias, W.L., Vincent, J.L., Russell, J.A., Silva, E., Trzaskoma, B., Williams, M.D.: Early changes in organ function predict eventual survival in severe sepsis. Crit. Care Med. 33(10), 2194–2201 (2005)
19. Nfor, T.K., Walsh, T.S., Prescott, R.J.: The impact of organ failures and their relationship with outcome in intensive care: analysis of a prospective multicenter database of adult admissions. Anaesthesia 61, 731–738 (2006)
20. Kayaalp, M., Cooper, G., Clermont, G.: Predicting with variables constructed from temporal sequences. In: Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics. pp. 220–225 (2001)
21. Shahar, Y.: A framework for knowledge–based temporal abstraction. Artificial Intelligence 90, 79–133 (1997)
22. Bellazzi, R., Larizza, C., Magni, P., Bellazzi, R.: Temporal data mining for the quality assessment of hemodialysis services. Artificial Intelligence in Medicine 34, 25–39 (2005)
23. Kohavi, R., Sahami, M.: Error-Based and entropy-based discretization of continuous features. In: Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, pp. 114–119 (1996)
24. Silva, A., Cortez, P., Santos, M.F., Gomes, L., Neves, J.: Mortality assessment in intensive care units via adverse events using artificial neural networks. Artificial Intelligence in Medicine 36, 223–234 (2006)
25. Srikant, R., Agrawal, R.: Mining Sequential Patterns: Generalizations And Performance Improvements. In: Proc. 5th Int. Conf. Extending Database Technology, vol. 1057, pp. 3–17 (1996)
26. Abu-Hanna, A., Keizer, N.F.: Integrating Classification Trees with Local Logistic Regression in Intensive Care Prognosis. Artificial Intelligence in Medicine 29(1–2), 5–23 (2003)
27. Hosmer, D.W., Lemeshow, S.: Applied logistic regression. John Wiley & Sons, Inc., New York (1989)

Appendix: Logistic Regression

A logistic regression model [27] specifies the conditional probability of a binary outcome variable Y, given the values of the covariate vector $\mathbf{x} = (x_1, ..., x_m)$: $p(Y = 1 \mid \mathbf{x}) = \frac{e^{LP(\mathbf{x})}}{1+e^{LP(\mathbf{x})}}$. For m covariates the natural logarithm of the odds (*logit*) is equal to the *linear predictor* $LP(\mathbf{x})$: $log(\frac{p(Y=1 \mid \mathbf{x})}{1-p(Y=1 \mid \mathbf{x})}) = LP(\mathbf{x}) = \beta_0 + \sum_{i=1}^{m} \beta_i \cdot x_i$ where β_i, $i = 1, ..., m$, denote the coefficients of the m covariates. A coefficient (β_i) can be interpreted in terms of an *odds-ratio*. Suppose the linear predictor is $\beta_0 + \beta_1 \cdot SAPS + \beta_2 \cdot Ep$ where $Ep = 1$ for patients having some specific episode and 0 for patients not having the episode. The odds of dying for those having the episode, $odds(Ep = 1)$ is $P(Y = 1|Ep = 1)/P(Y = 0|Ep = 1)$ and for those not having the episode, $odds(Ep = 0)$, is $P(Y = 1|Ep = 0)/P(Y = 0|Ep = 0)$. The quantity e^{β_2} is equal to the odds-ratio $odds(Ep = 1)/odds(Ep = 0)$. A positive coefficient corresponds to an odds-ratio > 1 and indicates that, when adjusting for all other variables (here SAPS), the odds of the event is higher for those with the episode compared to those without it.