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.
Chapter 4

Frequent multiple organ failure patterns in the Intensive Care: prevalence and prognostic value over days of stay

Toma T. et al. *In submission*
Abstract

Background: Organ failure (OF) is associated with increased mortality risks in the intensive care unit (ICU). However, little is known about the exact prevalence and risks of death associated with single and multiple OF during ICU stay. The objective of this study was to identify frequent single and multiple OF occurring at the ICU, and investigate their utility to predict risks of death.

Materials and Methods: Data were collected between July 1998 and February 2007 in an 18-bed mixed general-surgical Intensive Care Unit of a teaching hospital. The Simplified Acute Physiology Score II (SAPS-II) exclusion criteria were applied. Based on daily Sequential Organ Failure Assessment (SOFA) scores, frequently-occurring OF-patterns were discovered. These patterns were subsequently used as candidate covariates in mortality prediction models for the first 7 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.

Results: From 9,103 admissions, 2,928 adhered to the SAPS-II inclusion criteria and were used in the analysis. From these, 2,678 (91.5%) experienced OF at least once during their ICU stay. Single OF was more common (n=1,413 cases, 48.3%) than multiple OF (n=1,265 cases, 43.2%), and the large majority of patients (n=1,291 cases, 91.4%) with single OF had only respiratory failure. From all occurrences of multiple OF, n=2,581 (79.8%) involved exactly two organs. The most frequent combination of two or more failing organs includes the respiratory and central nervous systems (21.7% of all admissions and 48.8% of all occurrences of MOF). 56.4% of patients having multiple OF including a failing nervous system did not survive. Replacing the OF count in prognostic models by patterns of frequently-occurring OF yields a small and statistically significant increase in predictive performance. AUC-values of the pattern-based models ranged from 0.814 at day 1 to 0.722 at day 6. Single OFs were never selected in the final models. Respiratory and neurological failure appeared both in various combinations in six out of seven models.

Conclusions: Most occurrences of multiple OF in the ICU pertain to two organ systems. Mortality predictions benefit from including OF patterns instead of OF counts in logistic regression models. Patterns of multiple OF have a larger predictive value than single OF patterns.
4.1 Introduction

Prognostic models have various applications in medicine including benchmarking; stratification and triage; clinical decision support; and for informing patients and their families [1, 2].

Most prognostic models currently used in the intensive care unit (ICU) predict the risk of hospital mortality based on a quantification of the severity of illness of a patient at ICU admission. This family of models includes the Simplified Acute Physiology Score II (SAPS-II) model [3] and the Acute Physiology and Chronic Health Evaluation II and III (Apache-II, Apache-III) model [4, 5] and they are mainly used for benchmarking the quality of care in different ICUs, using mortality as a quality indicator. The Sequential Organ Failure Assessment score (SOFA) was introduced in the mid 1990s and measures organ function derangements not only on the day of admission but also on each other day of ICU stay [6]. The SOFA score quantifies the degree of organ dysfunction (from 0 to 4) for six major organ systems: respiratory (Resp), coagulation (Coag), hepatic (Hepa), circulatory (Circ), neurological (Neuro), and renal (Ren). The value 0 represents normal function, values 1 and 2 indicate increasing levels of organ dysfunction, and values above 2 indicate organ failure.

Although the SOFA score was mainly designed to communicate about and report on individual and patient groups’ organ dysfunction, many studies investigated the prognostic value of this score for predicting ICU and hospital mortality as shown in the systematic review by Minne et al [7]. The reviewed papers differed in the way the SOFA scores were handled. In some studies the number of organs failing on a day of stay was used as predictor in the model, in others SOFA-based abstractions were used such as the mean SOFA score, maximum SOFA, or the difference between the third and first days [8, 9, 10]. Several studies were concerned with the individual components of the SOFA score [11, 12, 13, 14]. Some earlier studies of ours [11, 15] used patterns in the SOFA scores over several consecutive days (e.g. showing increase in the scores) that were discovered from
the data. To our knowledge, however, no studies have investigated the prognostic value of multiple organ failure (MOF) patterns. In particular, it is unclear whether the specific combination of co-occurring organ failure is more prognostic than simply counting the number of failing organs. Studying the prognostic value of MOF patterns is complicated by the fact that the number of possible organ failure combinations grows exponentially in the number of organ systems (six). A modeling approach based on exhaustive testing of all possible MOF patterns is therefore not feasible.

The objectives of this study were (i) to identify frequently-occurring patterns of single and multiple organ failure (OF); (ii) to describe their incidence and association with hospital mortality; and (iii) to investigate the added prognostic value of frequently-occurring OF patterns over mere specification of the number of failing organs per day.

4.2 Materials and Methods

4.2.1 Data

The data set originated from the OLVG Intensive Care Unit in Amsterdam, an 18-bed mixed general-surgical Intensive Care Unit of a teaching hospital. We retrospectively collected data of all consecutive ICU admissions between 1998 and 2007 resulting in 9,103 admissions. Because the great majority of cardiac surgery patients leave the ICU within two days and because we wanted to adjust for the SAPS-II severity-of-illness score in our models we applied the SAPS-II exclusion criteria. Admissions excluded consisted of cardiac surgery patients (n=5,291), readmissions (n=496), admissions lasting for less than 8 hours (n=337), admissions for burns (n=8), and patients under the age of 18 years (n=33). A total of 2,928 admissions remained for analysis.

From a total of 13,507 patient-days, 226 (1.7%) had missing SOFA data. These were imputed as the maximum adjacent day’s score, and if more than 2 consecutive values were missing the admission was excluded (10 admissions).
4.2.2 Analysis

The first step in our analysis was the discovery, in the entire dataset, of frequent patterns of organ failure (both single and multiple OF) and the assessment of their relation to hospital death. We define MOF as any combination of organ systems failing (i.e. with organ-specific SOFA subscore greater than 2) at the same day of stay in the ICU. These length-1 (covering only 1 day) patterns are an example of vertical temporal abstraction in terms of the framework of Y. Shahar [16]. Any OF pattern occurring in at least 1% of all included admissions was considered to be frequent. For brevity the term MOF in the sequel will refer to a frequent MOF. The MOFs are denoted by a combination of shorthand names of the failing organ systems. For example \(<\text{RespCirc}^*\rangle\) represents failure of both the respiratory system and circulatory system at the same day. The asterisk means that other failure may also (but do not have to) be present at that day. The absence of an asterisk means that no other organ is failing at the same day, for example \(<\text{Resp}\rangle\) means that only the respiratory system is failing. Patterns explicitly indicating only one organ type, such as \(<\text{Resp}^*\rangle\), cover occurrences of both single and multiple organ system. In addition, we use \(<\text{Resp}+\rangle\) to denote the co-occurrence of respiratory failure and at least one other type of organ failure at the same day, thus effectively referring to multiple organ failure only.

The second part of our investigation is the prognostic analysis. Here we focused on the data of first 7 days of ICU stay of each admission in the dataset. For each day, a multivariable logistic regression model [17] was build to predict hospital death, using organ failure assessments up to the day of prediction. Patients who were discharged earlier or did not survive until that day were excluded from that analysis. OF patterns were represented as dummy (binary) variables in each model; the other covariates considered were the SAPS-II score, and the SOFA score and the number of failing organs at the day of prediction.

For assessing the added predictive value of patterns, in comparison to just counting the
number of failing organs, we also developed a second category of daily models trained on the same dataset as the OF models and including as covariates the SAPS-II score, the SOFA score and number of failure at the day of prediction.

A bootstrap procedure was used for selecting the OF patterns in the daily prognostic models and to obtain confidence intervals around parameter estimates. In algorithm 5 we formalize the approach used for model fitting.

Algorithm 5 The inductive method used to generate daily models, using OF patterns, for hospital mortality.

- PAT - original ICU patient dataset
- B - number of required bootstrap samples
- D - set of prediction days
- Discover(S) - returns a set of frequent patterns of organ failure discovered from dataset S
- MethodStepAIC(C, S) - returns the model that was fit on dataset S using candidate variables from set C with Akaike Information Criterion (AIC) balancing model simplicity and the goodness of fit
- getCov(M) - returns the set of independent variables from model M
- updateCount(CS, VS) - increments the counter values in set CS for the variables in VS set
- sortByCount(AC, CC, sortOrder) - sorts the set AC by correspondent values from CC in the order indicated by sortOrder: DESC = descending or ASC = ascending
- BS Sample(S) - returns a bootstrap sample of set S

1: for i = 1 to B do
2:  \( b_i \leftarrow BS\text{Sample}(PAT) \) /*Obtain a bootstrap sample from original dataset*/
3:  \( OF\text{patterns}_i \leftarrow Discover(b_i) \) /*Discover frequent OF patterns from sample \( b_i \)*/
4:  for all \( d \in D \) do
5:    candidates_{i,d} \leftarrow \{SAPS-II_{i,d},SOFA_{i,d},\#OF_{d},OF\text{patterns}_i\} /*set of variables candidates for day \( d \) model on sample \( b_i \)*/
6:    \( Mod_{i,d} = \text{MethodStepAIC}(\text{candidates}_{i,d}, b_i) \) /*fit day \( d \) model on bootstrap sample \( b_i \)*/
7:  end for
8: end for
9: for all \( d \in D \) do
10:  allCov_d \leftarrow \{\} /*initialize the set of all \( d \) day bootstrap covariates*/
11:  covCount_d \leftarrow \{\} /*initialize the variable selection count set for day \( d \)*/
12:  for i = 1 to B do
13:    allCov_d \leftarrow allCov_d \cup getCov(\text{Mod}_{i,d}) /*add model covariates to all covariates set*/
14:    covCount_d \leftarrow updateCount(covCount_d, getCov(\text{Mod}_{i,d})) /*increment or add counters for model \( \text{Mod}_{i,d} \) covariates*/
15:  end for
16:  candidates_d \leftarrow sortByCount(allCov_d, covCount_d, DESC)[0 : 9] /*return the top 10 variables candidates for day \( d \) model inclusion*/
17:  \( Mod_d = \text{MethodStepAIC}(\text{candidates}_d, PAT) \) /*generate final model for day \( d \) of prediction*/
18: end for

At any prediction day, the top 10 variables, ranked by the number of times they were retained in the bootstrap models (1 model for each of the 1000 bootstrap samples) were considered candidates for inclusion in the final model. Our procedure requires that
frequent OF patterns are discovered from each bootstrap sample and used for developing a prognostic model from the sample. Initially, all frequent patterns are included in the model, a variable elimination step based on the Akaike Information Criterion (AIC) being subsequently applied to remove the less predictive ones. The AIC is a measure favoring the most predictive covariates but penalizing model complexity [19]. The 10 final candidate variables are undergoing a last model inclusion process using the original dataset.

Predictive model performance was measured by the AUC, which provides a measure of the model’s ability to discriminate between survivors and non-survivors, and the Brier score, which is an accuracy measure combining both aspects of discrimination and calibration. A higher value of the AUC and a lower value of the Brier score indicate better model performance. We tested the statistical significance of the differences in the AUC and the Brier scores between the two kinds of models. We applied bootstrapping for obtaining bias corrected prognostic performance estimates for the two types of daily models as well as confidence intervals for the accuracy measures.

In order to better explain the models, an additional multicollinearity investigation was carried out using the Variance Inflation Factor (VIF) [20] which measures how much the variance of an estimated model coefficient is increased because of collinearity. According to the common rule of thumb, a model variable’s VIF value above a threshold of 10 or even 4 indicates (very) high collinearity associated with that variable. The Variance Inflation Factor for the $k^{th}$ predictor can be expressed as:

$$VIF_k = \frac{1}{1 - R^2_k}$$

where $R^2_k$ is the $R^2$ value obtained by regressing the $k^{th}$ predictor on the remaining predictors.
4.3 Results

Table 4.1 lists admission characteristics of all patients in the dataset, and separately for survivors and non-survivors. From the 2,928 patients included the dataset, 728 (24.9\%) died in the hospital. The hospital discharge mortality rate consistently increased as patients stayed longer in the ICU: from the 1,830 patients who stayed at least two days at the ICU, 551 (30.1\%) died, and for longer lengths of ICU stay these were 446 out of 1,360 (32.8\%; 3 days); 266 out of 792 (33.6\%, 5 days); and 183 out of 499 (36.7\%, 7 days) patients. The SAPS-II severity-of-illness score, the APACHE-II severity-of-illness score, the SOFA score at the day of admission and its individual organ system scores at the day of admission were all greater in non-survivors than survivors. The respiratory component of the SOFA score had the highest mean value at the day of admission in both survivors and non-survivors of 3.67 and 3.85, respectively. The hepatic system had the lowest mean admission score of 0.30 in survivors and 0.48 in non-survivors. Non-survivors had a longer length of ICU stay than survivors, the median length of stay was 3 days for non-survivors and 2 days for survivors. Also, the percentage of medical admissions in the non-survivor group was much larger that in survivors (i.e. 76.8\% versus 55.1\%). Moreover in the first 24 hours of admission the non-survivors had a more sever brain injury than survivors as indicated both by the SOFA score for the central nervous system (i.e. “SOFA Neuro adm” - higher value means more severe condition) and the Glasgow Coma Scale (GCS) values (lower values represent less severe brain injury).

4.3.1 Patterns of organ failure

Table 4.2 lists the most frequent patterns of organ failure discovered from all their days of stay of all patients. We distinguish between patterns describing single failure, multiple failure, and mixed patterns. Mixed patterns had the highest support (i.e. incidence) in the dataset.
Table 4.1: General characteristics of the patient population. The mean SOFA values are computed at patient admission day and the p-value indicates the statistical significance, at the level 0.05, for the differences between survivors and non-survivors.

Respiratory failure is the most frequently occurring organ failure and appears in the majority of patients, 2,542 (86.8%). The mortality risk is much larger in the group of patients developing respiratory and other failure(s), 531 (43.3%), compared to single
Table 4.2: Patterns of organ failure occurring in at least 1% of patient population. The support over the total number of patient days and total number of patients is depicted both in percentage and absolute values. The mortality in percentage and in absolute value for patient groups defined by each pattern occurrence is also presented. The symbol “**” represents a placeholder for the other organ systems that fail but are not explicitly shown in a pattern. The notation “+” indicates that a pattern, explicitly describing only one organ system, is a MOF pattern.

respiratory failure, 337 (17.7%). Risk associated with single respiratory failure is smaller than the average mortality risk of 24.9%. Single organ failure, other than respiratory failure, are rare. Except for the coagulation and renal systems, they occur in less than 1%
of patients. The highest mortality risk for a single failure is 32.0% (within only 9 patients) for the central nervous system.

The most frequent MOF patterns with respiratory failure also include the central nervous system (647 patients / 21.7% of patient population), the circulatory (344 patients / 11.7%), the coagulation (324 patients / 11.1%), or the renal system (372 patients / 12.7%). Among these, the highest mortality percentage (56.3% / 357 patients) is recorded in the group of patients developing the pattern <RespNeuro*> (respiratory and central nervous system failure but other systems can fail too) and the lowest (32.3% / 111 patients) for the respiratory and circulatory system failure (represented by pattern <RespCirc*>). The mortality for <RespCoag*> and <RespRen*> was 45.7% (148 patients) and 49.2% (183 patients) of the population developing the patterns.

The MOF patterns including the failure of the following combinations of organs show the highest mortality risk (between 70.7% and 75.0% of patients developing them): nervous and coagulation systems; nervous and renal systems; nervous and circulatory systems.

The number of organ failure, as expected, was associated with mortality. From Table 4.3 we noticed a direct relation between the number of organ failure and mortality. Patients without organ failure (but perhaps one or more dysfunctional organs) during their ICU stay amount to 8.50% (250 patients) of the sample and have mortality of 10.0% (25 patients). Patients with at most one organ failing at each day of their ICU stay, 48.34% (1413 patients), had a mortality of 11.7% (166 patients) and with at most two organs failing amounted to 932 patients (31.83%) with a mortality of 34.8% (324 patients). Maximum three or more organs failing are less common, there are 263 patients (9.0%) with at most 3 organs failing per day during ICU stay (64.6% mortality) and 2.33% (70 patients) with more daily failures (with an average mortality of 61.4%). Another finding was the association between the number of failures and the length-of-stay (LOS). The groups of patients with a higher number of failures had a longer length-of-stay, depicted in Table 4.3 by its median values, indicating a positive relationship between ICU stay and severity of organ failure.
### Table 4.3: The number of daily organ failure distribution in patient population expressed in percentage and absolute value. Relative mortality in the groups defined by the number of failure are also indicated. The SOFA median interquartile ranges (IQR) are also shown.

<table>
<thead>
<tr>
<th>Number of failure</th>
<th>Occurrence % / #patients (n = 2,928 patients)</th>
<th>Group mortality % / #patients</th>
<th>Group LOS median (IQR)</th>
<th>Group mean SOFA median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8.5% / 250</td>
<td>10% / 25</td>
<td>1 (1)</td>
<td>4 (2)</td>
</tr>
<tr>
<td>1</td>
<td>48.34% / 1,413</td>
<td>11.7% / 166</td>
<td>2 (2)</td>
<td>6.3 (1.6)</td>
</tr>
<tr>
<td>2</td>
<td>31.83% / 932</td>
<td>34.8% / 324</td>
<td>3 (4)</td>
<td>9 (2.5)</td>
</tr>
<tr>
<td>3</td>
<td>9% / 263</td>
<td>64.6% / 170</td>
<td>5 (6)</td>
<td>11.6 (2.5)</td>
</tr>
<tr>
<td>4</td>
<td>2% / 60</td>
<td>56.7% / 34</td>
<td>7 (10.5)</td>
<td>13.9 (3)</td>
</tr>
<tr>
<td>5</td>
<td>0.3% / 9</td>
<td>88.9% / 8</td>
<td>7 (2)</td>
<td>15.6 (2.8)</td>
</tr>
<tr>
<td>6</td>
<td>0.03% / 1</td>
<td>100% / 1</td>
<td>6 (0)</td>
<td>17.3 (0)</td>
</tr>
</tbody>
</table>

4.3.2 Prognostic models

Table 4.4 describes all the logistic regression models, in terms of their linear predictor (which calculates the predicted log odds of hospital mortality). The table also presents their (bias corrected) predictive performance quantified by the Brier score and AUC. The models using patterns, referred to as OF-models, always had a slightly better accuracy and discrimination than those using only the number of failure, sometimes by a statistically significant margin with the Brier score between 0.138 and 0.200 (significantly better at days 2 to 5 and 7) and AUC between 0.722 and 0.814 (significantly better at days 2 to 4 and 6).

OF-patterns combining respiratory failure with other organ failure are the most selected patterns in the models. The respiratory single OF pattern was never selected in the daily models despite its large relative frequency. Patterns explicitly including the central nervous system failure were associated with large odds ratios, between 1.6 and 3.0 at all days of analysis, except for the pattern <Coag.Neurom> at day 3 which has a negative model coefficient.

Besides the respiratory system, the most selected MOF patterns pertain to the central
Table 4.4: The daily prognostic models using admission information and patterns of failure in prediction. Their performance is expressed in terms of the Brier score and the AUC; in bold font the best performance values, marked with ‘*’ their statistically significant differences between the two models. For a simplified notation, the SAPS-II score is identified by SAPS in the prognostic models’ logit.

nervous and coagulation systems. MOF patterns including circulatory failure were the least selected while hepatic failure patterns did not appear in any model. MOFs explicitly indicating four or more organs failing are very rare and did not qualify as candidates for inclusion in the models.
Models for days 3, 4, and 7 exhibited negative coefficients for some patterns of failure they include. This might not seem intuitive in a logistic regression model as its coefficients can be translated in odds ratios and a negative coefficient might indicate that some organ failure helps survival. Our investigation in the phenomena using VIF showed its values to be low for the patterns at days 3, 4, and 7 (VIF < 1.7) well under the accepted threshold for high collinearity. The number of failure selected in day 1 model had a VIF of 5.9 which indicates strong collinearity. Higher VIF values were noticed in general for the SOFA score, reaching a maximum of 3.6 in day 7 model.

4.4 Discussion and Conclusions

4.4.1 Main findings

Multi organ failure (MOF) occurring in the ICU mostly includes combinations of 2 organs failing at the same day. Single organ failure remains dominant, however, respiratory failure (<Resp>) represents 92.9% of all cases of daily single failure suggesting that the majority of patients were on respiratory support (and hence considered to have respiratory failure). The majority of MOF patterns involve combinations with respiratory failure, a sign this failure occurs in most patients regardless of their survival outcome.

The patterns presented in this work addressed only the organ failure at a single ICU day. Our choice is motivated by several aspects we intended to investigate: we wanted to learn about the types of organ failure frequently occurring at any day during ICU stay and, using a pattern representation, to compare them with the daily number of failure as predictors in logistic regression for hospital mortality. Moreover, using simpler patterns (i.e. length equal 1) we reduced the complexity of representation. For example, this avoids the need to address complex temporal dependencies that may characterize longer multivariable patterns such as temporal ordering of failure from different organ systems.

Our analysis encloses a procedure that assigns the patterns that are candidates for model inclusion and which can be seen as a sort of prognostic value driven pattern dis-
covery. The candidate patterns for the final model are the result of a bootstrap strategy involving pattern discovery and model induction from each of the 1000 bootstrap samples. The procedure returns the top 10 variables (including patterns) most selected in the model for the day of analysis and to be used as candidates for inclusion in the final prognostic model.

The central nervous system followed by the coagulation system, were the most selected organ systems in the daily prognostic models, often appearing in combination with respiratory failure. Patterns including central nervous system failure show, by means of model coefficients, a direct relation to mortality (odds ratio in the range 1.63 – 3) in line with our expectation. The coagulation system, on the other hand, appears in patterns receiving positive (indicating a covariate bad for survival) as well as negative coefficients (indicating a covariate helping survival). For example \(<\text{CoagNeuro}\times\text{Neuro}>\) representing the combined coagulation and central nervous system failure received a negative coefficient in the day 3 model. The respiratory failure in itself is not directly associated with mortality also because providing patients with mechanical ventilation alone does not reflect the severity of a patient’s current health condition. Due to its dominance, the respiratory system failure is often part of various multiple failure patterns. Patterns such as \(<\text{Neuro}\times\text{RespNeuro}\times\text{Neuro}>\) and \(<\text{RespNeuro}\times\text{RespNeuro}>\) will hence largely overlap and may correspond to similar mortality risks.

In 3 out of 7 models there were patterns receiving a negative coefficient. This is usually the result of high correlation between model variables \([19]\) such as inter-pattern multicollinearity, correlation between patterns and the number of failure, and between patterns and the daily SOFA score, respectively. While this phenomenon does not have an impact on the accuracy of predictions, it can hinder the interpretation of the model coefficients. However, our investigation using the Variance Inflation Factor index did not reveal the presence of high multicollinearity for the patterns. On the other hand the VIF is known to be not very informative if some variables are algebraically connected to each other \([21]\). The negative coefficients can also indicate a “correction” applied
by the logistic regression for the average mortality associated with other model variables (e.g. SOFA score) that can be too low or too high. For example, at day 7, a too high average mortality predicted by the combined SAPS-II, SOFA, and number of failure was tempered by the OF patterns. Finally, it worth noting that all OF-patterns influence each other and, by definition, correlation between some of them is implicit as, for example, the occurrence of MOF patterns correlates with that of SOF patterns and other MOF patterns (partially) involving the same organ systems. For example, the patterns selected in day 3 model, \(<\text{RespNeuro}^*>\) and \(<\text{CoagNeuro}^*>\), have a large overlap of the patient groups they define (94% of the patients with pattern \(<\text{CoagNeuro}^*>\) have also pattern \(<\text{RespNeuro}^*>\)).

The accuracy evaluation showed that the OF-models (using MOF patterns) had a better calibration and discrimination at all days than organ failure models. The gain in accuracy was small but often statistically significant. Moreover, although the daily number of organ failure is a predictor for mortality, when competing for model inclusion against the patterns of failure, it is often eliminated from the final model by the selection strategy. We allowed the number of failure and OF-patterns to be candidates for selection in the same model as we intend to pinpoint the added value OF-patterns have in prediction. Our model variable selection relies on a stepwise backwards elimination guided by the search for the smallest Akaike’s Information Criterion (AIC). This measure favors the goodness of fit but also model simplicity by penalizing the number of model covariates.

The least improvement in accuracy, when using patterns, was recorded on day 1. One reason for this is the characteristic of the patient population at admission day consisting of many cases that are discharged within the first 24 hours of ICU stay. The more difficult cases are usually patients staying more than just several hours in the ICU where their mortality outcome is dependent not only on the severity of illness at arrival in the unit but also on the developments in their health during the ICU stay as a response to therapy. The gain in prediction, for the OF-models over models based on number of failure, after
2 or more days of stay, was often statistically significant for $\Delta$Brier at 5 days, and for $\Delta$AUC at 4 days, respectively.

### 4.4.2 Strengths and shortcomings

The strength of this work resides in the new perspective on multiple daily organ failure we introduce. We discover frequent patterns of organ failure that can occur at any day of patient stay in ICU and which account for the type of failing organs. The pattern discovery is data-driven which, unlike predefined patterns, has the benefit of resulting in knowledge representative for the patient population. It also provides a more complete perspective of data and can identify unexpected phenomena or issues from data.

Our analysis shows the effect of multiple organ systems, failing at the same day of stay, on mortality. We also address the importance of the type of failure for mortality predictions.

We identified several shortcomings. Our study is based on a single centre which can have influence on the generality of some findings due to data particularities (determined by the severity of admissions, medical staff, etc...).

The use of a template-like pattern representation (e.g. < RespNeuro* >) has the limitation of not guaranteeing a precise assessment of which combination(s) of organ failure is the most predictive for mortality. The current representation allows the overlap between patient groups defined by different patterns, a result to that being collinearity. Collinearity directly affects the interpretability of the models and trying to attach meaning to them is hard, if not even misleading. Another possible limitation is the development of models specialized only for a single day of stay which is less convenient than having a single model applicable to any day of prediction.

### 4.4.3 Implications

This work provides physicians extra information on organ failure, such as type and influence on mortality, that can raise awareness and enrich the decision making on interventions...
or treatment. For example one knows that failures of the central nervous system and coagulation system at the same day associate with very high mortality rate (see Table 4.2) and that the MOF-pattern defined by them at day 3 of stay is a significant predictor for hospital mortality. Another implication can be related to ICU auditing as the performance of an ICU can be stratified on patient groups defined by the most important organ failure patterns. Moreover the temporal developments in organ failure (on daily basis) can be compared between different ICUs to assess for example the recovery rates in different units.

4.4.4 Future directions

One interesting development of this research, addressing multivariate patterns of organ function and their use in prediction, would be to devise a pattern representation describing complex temporal relations between the organ function scores. This would require the use of historical patient data and could make use of concepts as trend, order, time windows, etc. One possible approach is the use of a representation similar to that proposed in [22] where a pattern is a multivariate set of attributes. Moreover this can be paired with the use of patterns discovered from the sequences of daily number of organ failure in order to address a possible follow up of the research question tackled in the current work.

4.4.5 Conclusions

The results of this study provide insight in the importance of the type of failure in predicting patient survival. Our findings indicate the MOF-patterns’ added predictive value relative to the use of the daily number of organ failure. Moreover, discovering patterns facilitates the identification of the most dominant types of organs involved in multiple organ failure and represents a good modality to include them in prediction. To the best of our knowledge this is the first study describing frequent single and multiple organ failure in Intensive Care, their prevalence and their prognostic value over days of stay. In particular it is the
first study to address the added value of MOFs in comparison with models that only use the number of organ failure.

**Acknowledgements**

This work was supported by the Netherlands Organization for Scientific Research (NWO) under the I-Catcher project number 634.000.020.

**Bibliography**


