Prognostic modeling to evaluate the in-hospital and long-term mortality of intensive care patients
Brinkman, Sylvia

Citation for published version (APA):
Brinkman, S. (2013). Prognostic modeling to evaluate the in-hospital and long-term mortality of intensive care patients
Chapter 2

A comparison of the performance of a model based on administrative data and a model based on clinical data: Effect of severity of illness on SMRs of ICUs

Sylvia Brinkman, Ameen Abu-Hanna, André van der Veen, Evert de Jonge, Nicolette F. de Keizer

Critical Care Medicine 2012; 40: 373-378
Abstract

Objectives: It has been postulated that prognostic models based on administrative data can provide valid adjusted mortality rates in specific patient populations. In this study we compared the performance and robustness of a model based on administrative data (customized hospital standardized mortality ratio) and a model based on clinical data (customized Simplified Acute Physiology Score II) in the Dutch intensive care unit population.

Design: Cohort study of intensive care unit records from a national intensive care unit quality registry linked to administrative records from the Dutch National Medical Registration. The hospital standardized mortality ratio and Simplified Acute Physiology Score II models were first-level customized on the intensive care unit population.

Setting: Fifty-five Dutch intensive care units.

Patients: A total of 66,564 intensive care unit patients admitted from 2005 to 2008.

Interventions: None.

Measurements and Main Results: Performance expressed by measures of discrimination, accuracy, and calibration (area under the receiver operating characteristic curve, Brier score, Hosmer-Lemeshow Č-statistic, and calibration plots). Additionally, the robustness of the models was assessed by simulating changes in the population’s severity of illness and analyzing the effect on the intensive care units’ standardized mortality ratios. The area under the receiver operating characteristic curve and Brier score of the customized Simplified Acute Physiology Score II were significantly superior to that of the customized hospital standardized mortality ratio (0.85 and 0.11 vs. 0.77 and 0.13, respectively). Calibration plots showed good agreement between observed and predicted mortality for low-risk patients in both models, with more discrepancy in the high-risk patients when using the customized hospital standardized mortality ratio. Severity of illness had influence on the intensive care units’ standardized mortality ratios in both models, but the customized Simplified Acute Physiology Score II showed more robustness.

Conclusions: The customized Simplified Acute Physiology Score II outperforms the customized hospital standardized mortality ratio in the Dutch intensive care unit population. Comparing institutions based on standardized mortality ratios can be unfavorable for those with a more severely ill intensive care unit population, especially when using the customized hospital standardized mortality ratio.
Introduction

The interest in the publication of hospital mortality rates has increased in the last few years mainly due to the perceived need for transparency to the public and better quality of care in hospitals. Insurance companies and patients may choose their preferred hospital based on performance measures reported publicly such as mortality rates. Hospitals will be inclined to benchmark their performance to that of other institutions and to improve their quality of care if needed. However, fair comparison of the performance of hospitals based on mortality requires adequate case-mix adjustment. Case-mix adjustment is commonly obtained by using prognostic models. The data used for case-mix adjustment may differ between models, some models use clinical data and others use administrative data.

The Simplified Acute Physiology Score (SAPS) II prognostic model is specifically developed to quantitatively assess severity of illness in intensive care unit (ICU) patients using clinical ICU data (1). The hospital standardized mortality (HSMR) model developed by Jarman et al. is developed for all patients admitted to the hospital with a primary diagnosis within the diagnostic groups that nationally account for 80% of all in-hospital mortality and uses routinely collected administrative data (2-4). Currently, a HSMR model is used in the UK, USA, Canada, Sweden, Australia, Denmark, and the Netherlands (2-6). In the UK the HSMR of all hospitals is already publicly available for benchmarking purposes (7). The Dutch government decided that a HSMR of the Dutch hospitals will be published in 2011 for the same purpose.

Aylin et al. stated that routinely collected administrative data can be used to predict mortality risks in specific subgroups with similar discrimination to clinical databases (8). A major advantage of using administrative data for case-mix correction is that it is easier and less expensive to collect than clinical data.

The aim of our study was to analyze whether a model based on routinely collected administrative data, can be used to predict mortality in ICU patients. We compared performance of an on ICU data customized HSMR (cHSMR) model with the customized SAPS (cSAPS) II model. Furthermore, we studied the robustness of the cHSMR and cSAPS II models by investigating the relationship between severity of illness of the ICU population and standard mortality ratio (SMR) derived from both models. It should be noted that as the original HSMR model was not developed for ICU patients solely, the results of the cHSMR model are not generalizable to the original HSMR model.
Chapter 2

Materials and methods

Prognostic models
The SAPS II model is a logistic regression model that uses clinical ICU data of the first 24 hours after ICU admission. It includes 17 variables: 12 physiology variables, age, type of admission (scheduled surgical, unscheduled surgical, or medical), and three underlying disease variables (acquired immunodeficiency syndrome, metastatic cancer, and hematologic malignancy) (1).

The HSMR model uses routinely collected administrative data that are recorded at hospital admission for case-mix correction. The model includes the following factors: primary diagnosis (50 diagnostic groups based on the ICD-9, that are responsible for 80% of the hospital mortality), age (stratified in 20 age groups), sex, admission urgency (urgent/not-urgent, equivalent to emergency/elective), length of stay (stratified in 5 LOS groups), co-morbidity (measured by the Charlson Index), social deprivation (from the Dutch Central Office of Statistics), month and year of admission, and the type of organization that made the referral (7 options). These factors and their coefficients vary among each of the 50 diagnostic groups (2-4).

Both models are customized on the included ICU data using first level customization in which a new logistic regression model was fitted with the in-hospital mortality as the dependent variable and the logit-transformed original probability as the sole independent variable. This does not change the influence of individual covariates included in the model but only modifies their joint influence on the observed mortality in the external dataset (9-11). The customized HSMR and SAPS II models will be referred to as the cSAPS II and cHSMR model in the sequel.

Data
This study used a linked dataset from the Dutch National Intensive Care Evaluation (NICE) registry and the National Medical Registration (LMR) for patients admitted to the ICU between January 1, 2005 and January 1, 2008. The NICE registry contains all clinical data required to calculate mortality risk predictions according to the cSAPS II model for all consecutive ICU patients admitted to the participating ICUs (12). In 2007 approximately 60% of all Dutch ICUs recorded the data for all their admissions in the NICE registry. The physiological and clinical variables are collected using the patient’s status and charts manually or by using a Patient Data Management System (PDMS). The LMR contains all administrative data required to calculate the cHSMR for all patients admitted to Dutch hospitals (13). The LMR does not contain information to distinguish ICU patients from non ICU patients. In 2007 approximately 70% of all Dutch hospitals submitted their data to the LMR (14). The NICE and LMR data has been encrypted in a way that all patient identifying information, such as name and patient identification number, has been removed. In the
Netherlands, there is no need to obtain consent to make use of registries without patient identifying information. Both registries are officially registered according to the Dutch Personal Data Protection Act.

The linked dataset has been created by a deterministic linkage algorithm based on hospital of admission, gender, date of birth, hospital admission date, and hospital discharge date. Case-mix characteristics of the linked and non-linked records of the NICE registry were compared to evaluate potential bias due to incomplete linkage. The LMR and NICE data were linked by using SAS 9.1, and the statistical analyses were performed using the statistical environment R version 2.9.2.

**Performance assessment**

To assess the performance of both customized models, measures of discrimination, calibration, and accuracy were used. The discrimination indicates how well a model is able to distinguish survivors from non-survivors and is expressed as the Area Under the receiver operating characteristic Curve (AUC) (15). Good calibration (i.e. agreement between observed and predicted mortality for patient groups) of a model is essential for using the model for benchmarking. To evaluate models’ calibration, calibration plots were inspected and the Hosmer-Lemeshow Č-statistic was used in which observations are grouped based on deciles of predicted probability and compared to the proportions of the actual outcomes. Models’ accuracy is assessed by calculating the Brier score (10), which measures the average squared difference between the observed outcome and its predicted probability at the patient’s level. Estimates of the AUC, Brier score and the associated 95% confidence intervals of the models are obtained by bootstrapping with 1000 samples (16-18). A difference between the models was considered statistically significant when the confidence intervals did not overlap.

The hospital reason for admission is equivalent to the ICU reason for admission for patients directly admitted to the ICU. This is in contrast to patients admitted to the ICU after hospitalization to other wards, where deterioration of the patient’s condition necessitates ICU admission. We hence hypothesized that the cHSMR model performs better for the patients directly admitted to the ICU than for those admitted after hospitalization. Therefore, the performance of the models is assessed in the total ICU population as well as in the patients directly admitted to the ICU, and in the patients admitted to the ICU after hospitalization.
Chapter 2

Model’s robustness assessment

The Standardized Mortality Ratio (SMR) is the ratio of the observed in-hospital mortality and the expected mortality as calculated by either the cSAPS or cHSMR model. The overall SMR of an ICU should ideally not depend on the case-mix of the admitted patients but solely on the quality of the delivered care. A robust prognostic model should not lead to changes in SMR solely due to a different distribution of case-mix. To evaluate the robustness of the models we investigated the effect of increasing or decreasing the severity of illness of the population on the SMR of an ICU. The median SMR and the associated 95% confidence intervals for each ICU were obtained using bootstrapping with 1000 samples of the original dataset (16-18).

Simulation

For each ICU the influence of varying the severity of illness of the population on the SMR is determined. In order not to have one model at disadvantage, we expressed the severity of illness of the ICU population as the mean of the predicted cHSMR and the predicted cSAPS II mortality risk. For each ICU we calculated the severity of illness (as we defined it) and subsequently increased or decreased severity using a series of weighted bootstrap samples (19). In the weighted bootstrap samples patients with a higher severity of illness received a higher or lower probability to be selected for the bootstrap sample leading to a simulated ICU population with a higher or lower mean severity of illness. The weight in the weighted bootstrap was set as the severity of illness to the power ω. The parameter ω was increased or decreased from -0.6 to 0.6 with steps of 0.2. For each step of decrease or increase in the parameter ω, we simulated 1000 weighted bootstrap samples. The probability of selecting patients for the simulated bootstrap samples was either negatively (parameter ω from -0.6 to 0) or positively (parameter ω from 0 to 0.6) proportional to their severity of illness. This results in under or over-representation of patients with high mortality risk in the 1000 weighted bootstrap samples of the respective ICU. Accordingly, each ICU has 1000 weighted bootstrap samples per step of decrease or increase in the severity of illness. For each step of decrease or increase in the severity of illness, the cHSMR and cSAPS II SMR was calculated in the 1000 weighted bootstrap samples. Subsequently, the median SMR and the associated 95% confidence interval for the simulated ICU were obtained. The SMR of a robust and well calibrated model should not be influenced by changes in the severity of illness. Therefore SMRs should be stable over the simulated changes in severity of illness, which are induced by changing the weights in the bootstrap samples. To inspect the influence of severity of illness on SMR we calculated the slope (beta coefficient of the model) of the linear regression line between SMR and severity of illness for each ICU. The more the coefficient deviates from 0, the more the SMR is dependent on the severity of illness. Significant difference between the beta
coefficients of the two models was tested using the Wilcoxon rank sum test, P-values below 0.05 were considered statistically significant.

Results

Data

From June 1, 2005 to January 1, 2008, 79,585 patients fulfilling the cSAPS II inclusion criteria were admitted to the 55 ICUs included in the study. The included ICUs are mixed medical-surgical units located in university hospitals (n=6), teaching hospitals (n=22) or non-teaching hospitals (n=27). In total 67,135 (84.4 %) records could be deterministically linked with the LMR data. In total 571 (0.9%) records with a discrepancy between in-hospital mortality were removed. The remaining 66,564 records were used for the performance assessment. For the model’s robustness analysis only data of ICUs with >50 records were included, which resulted in 66,527 records from 53 ICUs. In the study period the number of admissions per ICU ranged from 141 to 3602. In figure 2.1 a flow diagram of the included data is given. Table 2.1 shows the demographics of the included patients and of the patients in the non-linked dataset. The demographic data shows significant differences between the groups such as higher severity of illness and mortality among the non-linked patients.

Table 2.1: Demographics of the patients in the linked and non-linked data

<table>
<thead>
<tr>
<th></th>
<th>Linked NICE data</th>
<th>Non-linked NICE data</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of admissions</td>
<td>66,564</td>
<td>12,450</td>
<td></td>
</tr>
<tr>
<td>Hospital mortality (%)</td>
<td>19.5</td>
<td>23.9</td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>ICU mortality (%)</td>
<td>13.1</td>
<td>15.8</td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>ICU length of stay in days (median (25-75%))</td>
<td>1.14 (0.79-3.71)</td>
<td>1.41 (0.78-4.29)</td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>Male (%)</td>
<td>57.7</td>
<td>56.5</td>
<td>0.023</td>
</tr>
<tr>
<td>Admission type (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical</td>
<td>47.2</td>
<td>52.5</td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>Urgent surgery</td>
<td>17.8</td>
<td>18.6</td>
<td>0.032</td>
</tr>
<tr>
<td>Elective surgery</td>
<td>35.0</td>
<td>28.9</td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>Age (mean (sd))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survivors</td>
<td>60.4 (16.8)</td>
<td>60.5 (17.3)</td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>Non-survivors</td>
<td>68.9 (14.0)</td>
<td>68.6 (14.5)</td>
<td>0.037</td>
</tr>
<tr>
<td>SAPS score (median (25-75%))</td>
<td>31.0 (20.0-45.0)</td>
<td>33.0 (21.0-48.0)</td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>SAPS probability (median (25-75%))</td>
<td>0.12 (0.04-0.35)</td>
<td>0.14 (0.04-0.41)</td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>HSMR probability (median (25-75%))</td>
<td>0.02 (0.01-0.07)</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

ICU: Intensive Care Unit, SAPS: Simplified Acute Physiology Score, HSMR: Hospital Standardized Mortality Ratio, NA: not applicable, * Significant difference based on a p-value<0.05
Chapter 2

Figure 2.1: Flow diagram of the included data

*Deterministic linked on: hospital, gender, date of birth, hospital admission date, and hospital discharged

Figure 2.2: AUC curves of the cHSMR and cSAPS II model
Performance assessment

Both models are first level customized on the included ICU data in which the joint influence of the included covariates is modified. The alpha and beta coefficients of the customized HSMR model were both 0.62, the alpha and beta coefficients of the customized SAPS II model were -0.44 and 0.81 respectively. In the total ICU population the AUC and Brier score of the cSAPS II were 0.85 (0.85-0.85) and 0.11 (0.11-0.11), both were significantly better than that of the cHSMR (0.77 (0.76-0.77) and 0.13 (0.13-0.13), resp.). The AUC curves of both models are shown in figure 2.2. The Hosmer-Lemeshow statistic of the cHSMR model (374.6, 340.5-458.3) was significantly lower than that of the cSAPS II model (881.2, 844.5-942.3). However direct comparison between these statistics is not warranted and the calibration plots in figure 2.3 show that the cHSMR model has calibration problems in the high-risk groups. In patients directly admitted to the ICU as well as in patients admitted to the ICU after hospitalization, the cSAPS II model outperforms the cHSMR model. The AUC and Brier score of the cSAPS II were 0.85 (0.84-0.85) and 0.12 (0.12-0.12) for patients directly admitted to the ICU and were 0.85 (0.84-0.85) and 0.10 (0.10-0.10) for patients admitted to the ICU after hospitalization. The AUC and Brier score of the cHSMR were 0.79 (0.78-0.80) and 0.14 (0.14-0.14) for patients directly admitted to the ICU and was 0.74 (0.74-0.75) and 0.13 (0.13-0.13) for patients admitted to the ICU after hospitalization.

![Figure 2.2: AUC curves of both models.](image)

Figure 2.3: Calibration plots of the cHSMR and cSAPS II model. The histograms represent the number of patients per risk group.
Models’ robustness assessment

The original SMR values of the included ICUs based on the cHSMR model ranged from 0.58 (0.49-0.66) to 1.84 (1.62-2.06) and ranged from 0.59 (0.52-0.66) to 1.56 (1.41-1.73) when using the cSAPS II model. In the appendix the SMRs of all ICUs according to both models are given, showing that the SMR of ICUs is influenced by the choice of model that is used. Increasing or decreasing the severity of illness, by means of the series of weighted bootstrap samples, has influence on the SMR in both models. For each of the 53 included ICUs the slope (beta coefficient), which describes the average change of SMR on a unit change in severity of illness, was obtained. A slope of 0 would mean that the severity of illness has no influence on the SMR. In figure 2.4 the beta coefficients of the two models are shown. Over the 53 ICUs the median cSAPS II beta coefficient (0.98 (0.14-2.44)) was significantly lower than the median cHSMR beta coefficient (2.47 (1.68-3.73)) (p<0.001) indicating that the cHSMR model is more influenced by the severity of illness than the cSAPS II model. There was only 1 ICU in which the beta coefficient of the cHSMR model was lower than the beta coefficient of the cSAPS II model.

Figure 2.4: Beta coefficients of the 53 included ICUs based on the two customized models. The beta coefficients represents the mean change of SMR per unit of change of severity of illness. A robust and well calibrated model should show beta coefficients of 0, independent of severity of illness.
Discussion

This study shows that the cHSMR model, originally developed for the general hospital population, performs fair for the Dutch ICU population, though there are certain calibration problems when applied to the more severe ICU patients. The cSAPS II model, specifically developed for ICU patients, performed better than the cHSMR model in the Dutch ICU population. Furthermore, the cSAPS II model is more robust than the cHSMR model as the SMR of ICUs is less influenced by the mean severity of illness of the ICU population. The Hosmer-Lemeshow Ĉ-statistic of the cHSMR model (374.6, 340.5-458.3) was lower than that of the cSAPS II model (881.2, 844.5-942.3) but these statistics cannot be compared directly because the deciles created by the two models differ markedly. When the cSAPS II model was given the deciles of the cHSMR model then its Hosmer-Lemeshow Ĉ-statistic was 1516.5 (1382.3-1653.3) in comparison to 7457.0 (7143.7-7751.9) of the cHSMR with the cSAPS II bins. This indicates that the cSAPS II model behaves better in terms of calibration. When comparing different ICUs based on the SMR we advise to use a clinical model, especially when there are large variations in the severity of illness across the different ICUs as the SMR based on the administrative model is less robust.

We have shown that the SMR of ICUs is influenced by the choice of model that is used. This result was also found in other studies (9,20,21) and might be explained by the different aspects of the case-mix that are emphasized by the two models. Importantly, we showed that the SMR is influenced by the mean severity of illness in a population. This means that hospitals admitting more severely ill patients might be disadvantaged compared to hospitals admitting less severe patients when using SMRs to compare quality of care. The cHSMR model is more sensitive to this phenomenon than the cSAPS II model. For this reason the SMRs should only be used to signal when performance might be poor, triggering further investigations, and not as an absolute indicator of quality of care. For measurement of the quality of care it has been suggested to combine process and outcome indicators (22).

In a study by Aylin et al. the performance of three predictive models based on clinical data was compared to a general prognostic model based on administrative data of the Hospital Episode Statistic (HES) in England (8). The AUC of the administrative model was similar to that of models based on clinical data in different specific subgroups. Our study shows that for the Dutch ICU population the performance of the clinical cSAPS II model is significantly better than that of the administrative cHSMR model. The cHSMR model uses primary diagnoses at hospital admission instead of primary reason for ICU admission. These two diagnoses can differ widely, especially for patients hospitalized before ICU admission, and can result in different predicted in-hospital death rates. The diagnosis at ICU admission is likely to be more closely associated with outcome
than the initial reason for hospital admission. This could partly explain the lower overall performance of the cHSMR model compared to the cSAPS II model. On the other hand, case-mix adjustment including the initial reason for hospital admission allows evaluating quality of care of the entire process of care in the hospital. Previous studies have shown that clinical models need periodical customization to maintain their validity (23,24). It can be anticipated that the cHSMR model is less sensitive to changes over time than the cSAPS II model because the latter is a clinical based model. This can be considered as another theoretical advantage of the cHSMR model compared to the cSAPS II model. However, when using a split-sample validation method on our data in which the models are customized on the data of 2005 and validated on the data of 2006 and 2007 there was no difference in the performance of both models than when using bootstrapping.

The predictive performance of the models is partially a reflection of the quality of the database and the type of patients it covers (8). Aside from the fact that the cHSMR model is not specifically developed for ICU patients, the lower predictive performance could be explained by the lower quality of the LMR data. Several studies already have addressed the issue of accuracy of administrative databases such as the LMR database (25-27). To assure quality of the data in the NICE registry it is compulsory for all participants of the NICE registry to attend training in collecting the data accurately, according to the stated data definitions reported in the NICE data dictionary. Furthermore, an onsite data quality audit is in place to ensure the validity of the data (12,28).

**Limitations**

In this study 15.6% of the NICE registry records could not be linked. This could lead to a biased selection of included patients (Table 1). The patients in the excluded non-linked dataset are more severely ill than the patients in the included linked dataset (SAPS II predicted mortality was 0.14 and 0.12 respectively). Although it is hard to extrapolate our findings due to the 15.6% excluded records, it is likely that including the more severely ill patients of the non-linked dataset would have strengthened our conclusion because the cHSMR was found to perform worse in the higher risk patients as shown in the calibration plot (Figure 2.3). Therefore we believe it is unlikely that the non-linked population influenced the conclusion that the cSAPS II model outperforms the cHSMR in ICU patients. In this study we used data from 2005 till 2008, though we believe that the conclusions of this study would not change if more current data were available. A limitation of the study is that we used the older SAPS II model which has the risk of becoming obsolete. During the study period, the more recently developed prognostic models such as the SAPS 3 and Acute Physiology and Chronic Health Evaluation (APACHE) IV models were not available in the used NICE registry. However, previous studies showed that the SAPS II model fit well to the Dutch ICU population Peek et al (29) and Brinkman et al (30).
Conclusions

This study shows that the cSAPS II model outperforms the cHSMR model on Dutch ICU patients. Both the cSAPS II model and the cHSMR model are influenced by the severity of illness of the admitted patients, though the cHSMR model is more sensitive to this phenomenon. As a consequence, using a SMR for comparing outcome of ICU patients might give biased results and therefore should be used with caution, especially when using the cHSMR.

Acknowledgements

We express our gratitude to the foundation Dutch hospital Data for providing us the LMR data that was required to accomplish this study.

References

**Appendix:** The median SMR and the associated confidence interval (CI) for each ICU according to both customized models. The right y axis of each figure represents the number of ICU admissions, and the left y axis represents the hospital ID.