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

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 8

General Discussion
8.1 Introduction

It is important to monitor quality of care at Intensive Care Units (ICUs) as this can initiate projects to improve it and therewith reduce the mortality among ICU patients. Mortality is a frequently used indicator of quality care, but because the observed mortality is strongly related to the type of admitted patients the observed mortality should be adjusted for case-mix. Currently different prognostic models are used to predict case-mix adjusted mortality. The standardized mortality ratio (SMR, i.e. the observed mortality divided by the predicted case-mix adjusted mortality) is used as quality indicator to monitor and compare quality of care among ICUs (i.e. benchmarking) or with a reference value. With a well customized prognostic model an SMR of 1 serves as a reference point for “average” quality among the ICUs, an SMR below 1 implies less observed deaths than expected by the prognostic model (and hence better quality of care than average) and an SMR above 1 implies more observed deaths than expected by the prognostic model.

The data needed to calculate the case-mix adjusted mortality is often available in quality registries. In this thesis we used several registries, namely the clinical quality registry National Intensive Care Evaluation (NICE) (1), the administrative National Medical Registration (LMR) (2), and the administrative insurance claims database of Vektis (3). These data registries have all been set-up for different purposes, respectively: to monitor quality of ICU care, to support the policies of hospitals, and to enable decision making and implementation within the healthcare market. However, due to overlapping variables between these registries it is possible to link the different registries. Linking these different registries enabled us to compare prognostic models of which the required variables were recorded in different registries and enabled us to predict not only the ubiquitously used in-hospital mortality but also the long-term mortality of ICU patients.

Before prognostic models can be used for mortality prediction and quality of care assessment, the performance of the prognostic models should be sufficiently validated. Furthermore, the prognostic models should be applied correctly with the right interpretation corresponding to their merits and limitations. In this thesis we have validated, compared, and applied different prognostic models to the Dutch ICU population in several settings. This chapter provides an overall discussion of the research described in this thesis by addressing the following research questions:

1. How do prognostic models perform in the Dutch ICU population?
2. Is there an association between ICU admission time and the case-mix adjusted mortality?
3. What is the long-term mortality of ICU patients and does the choice of follow-up end-point (i.e. in-hospital mortality or long-term mortality) affect the quality indicator SMR?
For each research question the main findings are summarized and discussed. Furthermore, the strengths and limitations of the studies are described. In conclusion we discussed the merits and limitations of using the SMR to assess the quality of ICU care.

### 8.2 Performance of prognostic models

The first research question was “How do prognostic models perform in the Dutch ICU population?” This research question was addressed in Chapter 2 and Chapter 3 in which the performance of different prognostic models was assessed using different statistics and performance measures.

Chapter 2 focused on the difference in the performance of a clinical and an administrative prognostic model (i.e. the customized Simplified Acute Physiology Score (SAPS) II model (4) and the customized Hospital Standardized Mortality Ratio model (HSMR) (5)). The SAPS II model used clinical data of ICU patients to adjust for case-mix differences which were more time consuming to collect than the administrative data which were used in the HSMR model. This study showed that the customized SAPS II model, specifically developed for ICU patients, outperformed the customized HSMR model, originally developed for the general hospital population. Furthermore, the performance of the customized SAPS II model was less influenced by the mean severity of illness of the ICU population than the customized HSMR model. In conclusion of this study we advised to use a clinical prognostic model for the assessment of the quality of care of ICUs especially when the quality of care is used for ICU benchmarking purposes. In this study we were not able to compare the more recently developed clinical APACHE IV model (6) with the administrative HSMR model as the clinical data used in the APACHE IV prognostic model was not yet collected in the period of this study. In Chapter 3, however, we evaluated the prognostic reliability of the APACHE IV model in the Dutch ICU population and compared its performance to the performance of the older APACHE II (7) and SAPS II prognostic models in different subgroups of the ICU population. The overall discrimination and accuracy of the customized APACHE IV model were statistically significantly better and the overall calibration was inferior compared to the customized APACHE II and SAPS II models, although the found differences were small and probably not very relevant in clinical practice. As the performance of the APACHE IV was equivalent to the SAPS II model, the APACHE IV model will most likely also outperform the HSMR model. According to the conclusions of Chapter 3 the customized APACHE II, SAPS II and APACHE IV model could equally be used for quality assessment purposes in view of prognostic performance. However, the APACHE IV model may be preferred as this model incorporates a large
number of reasons for ICU admission which enables analysis in specific diagnostic subgroups.

8.3 Variation in the quality of care

The second research question was: “Is there an association between ICU admission time and the case-mix adjusted mortality?” This research question was addressed in Chapter 4 in which we assessed whether there is variation in the in-hospital mortality of patients admitted during office hours and off hours while adjusting for the severity of illness.

It has been postulated that the quality of care varies during the day and that patients admitted to the ICU outside office hours are more likely to die. In Chapter 4 we used the APACHE II prognostic model to calculate the physiological dysfunction (APACHE II score) for case-mix correction. We analyzed the possible relationship between ICU admission time and in-hospital mortality in the Dutch ICU population. We calculated the relative risk for in-hospital mortality of patients admitted outside office hours while adjusting for the severity of illness (i.e. age, gender, APACHE II score, admission type and reason for ICU admission). We showed that the in-hospital mortality varied with time but was consistently higher outside office hours and lower during office hours. If further investigation would show that the increased in-hospital mortality outside office hours was caused by organizational factors then this could have great implication for health care institutions as organizational rules might be sharpened by the Dutch Health Care Inspectorate.

It is likely that the patients admitted during off hours are on average more severely ill than the patients admitted during office hours. Although we corrected for several case-mix factors it is still possible that there were some important differences between the patients admitted during office hours and off hours not accounted for in our analyses. This also means that it is possible that the prognostic model had a lower performance in the patients admitted during off hours.

In general it can be stated that if prognostic models are sufficiently validated and showed evidently good performance, then they can be reliably used for obtaining the quality indicator SMR (i.e. the observed mortality divided by the predicted case-mix adjusted mortality). However, reporting the SMR of different ICUs may still lead to misinterpretations as the SMR of an ICU is significantly influenced by the prognostic model that is used due to the existing differences in the in- and exclusion criteria of the models. For example the APACHE IV model excludes patients that are admitted from another ICU while the SAPS II model excludes patients that are transferred to an ICU in another hospital. This means that the SMR of an ICU can vary when different prognostic models are used. This has also consequences for the so called SMR ranking lists which are often
consulted by payers and consumers of health care to gain information on quality of care. The best performer in these SMR ranking lists depends among others on the prognostic model that is used. The absence of a gold standard prognostic model makes it impossible to identify the ICU with best or worst quality of care. Furthermore, prognostic models only adjust for the included covariates and therewith miss factors that might have a considerable influence on mortality (for instance Down syndrome and low vitality etc.). Therefore it has been suggested that 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 (8).

8.4 Long-term mortality of ICU patients

The third research question was: “What is the long-term mortality of ICU patients and does the choice of follow-up end-point (i.e. in-hospital mortality or long-term mortality) affect the quality indicator SMR?” This research question was addressed in Chapters 5, 6 and 7.

It can be argued that assessing the long-term mortality is more important than the in-hospital mortality as little has been achieved when patients die soon after hospital discharge. In Chapter 5 we first performed a literature review on the long-term outcome of ICU patients in which we extracted the determinants which were used for case-mix adjustment of the long-term mortality. This study showed that the long-term mortality found in the existing literature was difficult to compare among the studies due to differences in study design, case-mix, and case-mix adjustment. In Chapter 6 we assessed the unadjusted and adjusted long-term mortality of the Dutch ICU population. Case-mix adjustment was done with the determinants of the long-term mortality of the Dutch ICU population identified in Chapter 5. We showed that the comparison of the crude mortality among different diagnostic subgroups may lead to wrong conclusions. Some diagnostic subgroups have a high crude mortality while after case-mix adjustment this increase in mortality was not statistically significant and probably caused by pre-existing co-morbidities or higher age. Furthermore, we showed that the mortality in the first months after hospital discharge was substantial (5.4%). To which extent the additional mortality after hospital discharge was attributable to the preceding ICU admission should be further investigated. However, presumably this additional mortality could partly be explained by existing discharge policies. For instance if patients were discharged to a hospice for palliative care the long-term mortality will by definition increase considerably compared to the in-hospital mortality. In Chapter 7 we investigated the influence of using the in-hospital mortality versus long-term mortality on SMR and SMR rank position. This study showed that benchmarking on the in-hospital mortality or the mortality at 3, 6, or 12
months after ICU admission has influence on the SMR and SMR rank position of ICUs. Which SMR should be used when comparing or monitoring the quality of care depends on the setting. If an ICU wishes to monitor its own performance to set-up quality improvement projects, the in-hospital mortality might be sufficient. Small ICUs are often required (e.g. if the expected duration of treatment is longer than 72 hours) (9) to transfer very severely ill patients to larger, more equipped ICUs. Hence they are not able to improve their quality of care regarding these patients. Therefore information on the long-term mortality of the transferred patients is not needed for intern quality improvement projects. However, if an ICU has a low in-hospital mortality due to relatively many patients discharged to a hospice for palliative care or to another better equipped ICU, benchmarking on the in-hospital mortality might be unfair. The transferred patients can have different influences on the SMR based on the mortality 3 months after hospital discharge. For instance if an ICU transfers patients who die in the receiving ICU despite the good quality of care, the SMR based on the mortality 3 months after hospital discharge increases, and justly so. However, if the transferred patients die due to inadequate quality of care in the receiving ICU this increase of the long-term SMR is unjustified. Ideally, a combination of the in-hospital SMR and the long-term SMR would be used. If the in-hospital SMR is low and for instance the 3 month mortality is high then this might indicate that a relatively high percentage of the patients are discharged to another ICU or to another hospice for palliative care, or that the patients were discharged too early from the hospital.

8.5 Strengths and limitations

In this thesis the performance of different prognostic models was described using measures of discrimination, accuracy, and calibration. We used the area under the Receiver Operating Characteristic curve (AUC) to describe the discrimination, the Brier score to describe the accuracy, and the Hosmer-Lemeshow Č-statistic along with calibration plots to describe the calibration of the models. The values of these measures should be interpreted with caution as they all have some limitations. The AUC as well as the Brier score are dependent on the prevalence of mortality. If the observed mortality of the ICU population is very high or very low, the AUC and the Brier score values tend to improve (10). The Hosmer-Lemeshow Č-statistic is very sensitive to the sample size of the ICU population, if the sample size is very large the calibration according to the Hosmer-Lemeshow Č-statistic tends to be worse (11). These limitations should be considered when comparing the performance of prognostic models across different ICU populations. This means that comparing performance measures across different studies is difficult as the differences found can be caused by differences in case-mix and sample size of the population and not necessarily by the difference in the performance of the used models.
In this thesis, we therefore only compared the performance of prognostic models when applied to the same ICU population included according to the same inclusion criteria. Another reason why comparison of performance measures between different studies is complex is the difference in the time of model development and model validation. Several studies showed that prognostic models are not stable over time (12-14), causing a performance decrease in the years after model development. This means that direct comparison of newly developed models to older ones is not feasible. A solution for comparison of models with a different year of development is model customization on the data in which the model is validated, which is applied in Chapters 2 and 3.

In Chapter 5 we performed a literature review in which we extracted determinants of the long-term outcome. Subsequently we assessed the importance of these determinants on the long-term outcome of Dutch ICU patients by using the clinical database of the NICE registry. This implies that we could only use the identified determinants that were available in the NICE registration and currently used to predict the in-hospital mortality. This is a limitation as it is possible that there were other important determinants influencing the long-term mortality but that were not registered in the NICE registry. However, the most commonly used determinants in the literature corresponded to the determinants registered in the NICE. To assess the long-term outcome of the Dutch ICU patients the NICE registry was linked to the administrative insurance claims database of Vektis. In Chapters 5 and 6 we used NICE data of 2007 to 2010 of which approximately 29% could not be linked with the insurance claims database and in Chapter 7 we used NICE data of 2008 to 2011 of which 4% could not be linked with the insurance claims database. This could lead to some bias in the selection of included patients. We used deterministic linkage which overall produces a low number of false positive links (15) meaning that the linked data is reliable. Vice versa, it also explains the rather high percentage non-linked records as the deterministic linkage approach that we used can miss matches due to errors in the linking variables (false negative links). In the future it should be investigated how this linkage can be improved for instance by using probabilistic linkage (15) or by extending quality registries with a social security code. In Chapter 7 we only included the ICUs of which at least 80% of the admissions could be linked to the assurance claims database to minimize selection bias.

In this thesis we focused on the mortality of the ICU patients to assess the quality indicator SMR. However, besides mortality there are also other important quality indicators not presented in this thesis (e.g. length of ICU or hospital stay, the number of readmissions, and the quality of life of ICU patients after hospital discharge). It is questionable whether a low mortality at the expense of the quality of life is desired. However the quality indicator SMR gives important insights and can be used to further investigate the reason for mortality and thereby facilitate quality improvements projects. An important strength of this thesis is that in each chapter the used sample size is very
large and representative for the total Dutch ICU population. Furthermore, we were able to adjust the outcome of ICU patients and perform analyses in several ICU subgroups.

8.6 Implications for using mortality to benchmark ICUs

In the Netherlands hospitals are increasingly forced by the Dutch government and insurance companies to report their mortality figures publicly due to the increased interest in the transparency of hospital mortality. In other countries such as the UK the hospital mortality figures are already publicly available and the chance that this will be implemented in the Netherlands is increasing although there is much discussion about this approach. An important benefit of making the mortality numbers publicly available is that hospitals are legally required to actively monitor their quality of care and if necessary improve their performance. In general it can be said that the public has the right to know what happens in the Dutch hospitals. However, the public should be informed correctly which is hard due to the complexity of the quality indicator SMR. Furthermore, in case of the intensive care, most patients require an emergency admission and the nearest hospital will be designated. This implies that the patient has no influence on the choice of hospital and this might diminish the importance of making the mortality rates of Dutch ICUs publicly available.

We showed that the quality indicator SMR is influenced by the data that are used for case-mix adjustment (i.e. administrative or clinical data), the choice of prognostic model used for case-mix adjustment, and the choice of end-point of follow-up. Furthermore, the SMR of a hospital is also influenced by the quality of the data in the used quality registries. This means that the mortality ranking lists based on the SMR can identify different best performers depending on the method used for quality assessment. As long as the results of quality assessment is interpreted with caution and understanding of possible biases this is no problem. However, if the results are interpreted without insight in the used methods and the merits and limitations of each different method, this might be a large problem especially when there are political consequences and unjustified reputation damage. Publically presenting the raw mortality of hospitals does not hold meaningful information as these mortality rates are strongly related to the case-mix of the admitted patients. However, with the absence of a gold standard of the best prognostic model it is difficult to choose which prognostic model should be used for this purpose, and also the choice of follow-up end-point for this purpose is difficult. Even if the SMR is publically available under the right circumstances there are still some drawbacks. Hospitals can manipulate their data to create a better SMR, patients might develop a strong preference for a specific hospital of admission which could lead to waiting lists (in case of elective admissions) and in the most extreme case hospitals may refuse very severely ill patients.
with a low survival chance. For all the above mentioned reasons the mortality data of hospitals can only be made publically available very cautiously.

8.7 Conclusions

In this thesis we showed that clinical prognostic models outperform administrative prognostic models. The clinical APACHE IV model might be preferable for benchmarking purposes as it enables analysis in specific diagnostic subgroups. We showed that the post hospital mortality of ICU patients is substantial which should be considered when benchmarking ICUs. Taking the long-term mortality instead of the in-hospital mortality as the outcome measure can achieve this. We have shown that not only the choice of prognostic model but also the end-point of follow-up influence the SMR. This renders the interpretation of the SMR as a quality indicator a very complex task with its merits and limitations.

In conclusion of my thesis I would say that quality assessment and monitoring are of great importance, but the results of the quality assessment is only worthwhile when interpreted correctly.

References

