Mortality prediction in the intensive care: Role of mathematical models in benchmarking and decision-making
Minne, L.

Citation for published version (APA):
Minne, L. (2013). Mortality prediction in the intensive care: Role of mathematical models in benchmarking and decision-making

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 11
General discussion
The aim of this PhD dissertation was to better understand 1) characteristics of current prognostic models of mortality and/or survival in elderly (ICU; intensive care unit) patients, 2) the implications for their use over time in benchmarking, and 3) the role they can play in supporting decision-making for individual patients. In this final chapter, we will discuss the main findings of each of these three parts.

Part 1: Characteristics of prognostic models for elderly (ICU) patients

Main findings
We systematically reviewed literature on prognostic models of mortality and/or survival in elderly (ICU) patients. The majority of the 103 included studies in chapter 2 were concerned with the development of a new model (68%). Model validation was less often attempted. The models are rarely validated by others in an elderly subgroup, even the more common models developed some 20 years ago (e.g. the Charlson Index, Deyo Score, Simplified Acute Physiology Score [SAPS-II] and Acute Physiology And Chronic Health Evaluation [APACHE-II]). The clinical value of the models was not considered in most of the studies, while their reported (potential) use most frequently included support of individual clinical decisions (the actual intended use of the model was often not reported). Finally, we did not encounter any studies evaluating the model’s impact in practice, e.g. on individual patient decisions. Although the 7 studies described in chapter 3 scored relatively well on methodological quality and model performance can be considered reasonable (Area Under the Curve [AUC] range = 0.71-0.88; positive predictive value [PPV] range = 0.68-0.92, using cut-off points ranging from 0.5 to 0.8; p<0.05 in 4 of 6 cases by Hosmer-Lemeshow statistics), most studies did not use a set of performance measures specifically corresponding to the model’s intended use and multiple aspects of model performance (e.g. discrimination and calibration). We found caveats for the 17 described ICU models in the following elements that boost clinical credibility: the inclusion of the known clinically relevant variables, avoiding use of arbitrary thresholds for categorizing continuous variables, avoiding subjective data collection and reporting the range of probabilities provided by models.

General discussion
We believe that objective information regarding survival probabilities may have additional value in supporting individual treatment decisions. Concerns regarding withholding therapy and end-of-life decisions are especially relevant in elderly patients who may prefer comfort-focused end-of-life care when chances of survival are low. There are ongoing (national) discussions regarding the question whether “a patient is still allowed to die”. Physicians are trained to save lives and with the current technologies available in healthcare an increasing number of lives can be saved or extended. The population is ageing and the costs of healthcare are rising, but many elderly patients just wish to “die in peace”. Objective probabilities of survival may support physicians, patients and their families in making informed decisions about the continuation of intensive care treatment. Likewise, by providing reliable prognostic information, models may also aid in deciding to continue treatment in patients with relatively high likelihood of survival.

Our findings indicate however that the use of prognostic models in clinical practice is still very premature, as their clinical credibility is low and evidence for their external validi-
ty is scarce, which are prerequisites for building trust and acceptance [1,2]. The focus is too much on model development rather than their applicability in clinical practice. We believe that clinical credibility should be considered from the first stage of model development by: 1) including clinically relevant variables, such as data on functional and cognitive status for elderly patients [3] and information about patients during their entire stay, 2) avoiding the use of arbitrary thresholds and subjective data collection, and 3) developing models which provide a wide range of probabilities. Moreover, the model’s intended use should be considered and the appropriate performance measures should be used. Finally, there is a need for external validation studies, preferably by independent researchers in a different research setting.

Part 2: Role of prognostic models in benchmarking

Main findings
We monitored the performance of a parametric (logistic regression) model and a non-parametric (classification tree) model in a prospective dataset of 12,143 consecutive admissions of 21 Dutch ICUs of patients aged 80 years and older. We found that observed mortality was stable over time in this population, while there was an increasing trend in the patients’ severity of illness. The logistic regression model (chapter 4) was based on the SAPS-II score. Model coefficients were recalibrated on earlier data of our study population. While its performance was stable in terms of discriminative ability (AUC) and accuracy (Brier score), it was not stable in terms of calibration. This was revealed by a gradual decrease in the Standardized Mortality Ratio (SMR), which is the observed number of deaths divided by the predicted number of deaths. Repeated recalibration rendered the SMR stable again. The classification tree (chapter 5), which was developed on the same data set as used for recalibration of the SAPS-II model, had a stable AUC and an unstable Brier score. The observed mortality was lower than predicted, but the tree-model was still able to identify subgroups of patients with high risk of mortality. The proportions of patients assigned to the tree’s end-nodes (which contain subgroups of patients with similar characteristics) were not stable. The tree-model had a stable SMR (chapter 6).

General discussion
The increasing trend in severity of illness and the unstable proportions of patients assigned to the tree’s end-nodes are both indications for changes in the patient population (i.e. case-mix). The decrease of the overall SMR indicates that the SAPS-II model is outdated and overestimates mortality. This generates an overly optimistic view of the delivered quality of care when benchmarking across adult ICUs, which is a common use of the SAPS-II model. The SMR of the tree-model was stable, suggesting the tree-model has a longer shelf-life [4] than the logistic regression model. We believe however that instability in the global SMR of a model is not necessarily bad. Whereas the logistic regression model is sensitive to change and requires recalibration, the tree-model is less sensitive. Due to its binary nature, probabilities will only alter when changes are large enough for patients to fall in the “other” category. We believe that it is better to use a more sensitive model that needs to be updated frequently, than using a more stable model that takes a longer time to adjust its expected mortality in a changing environment.
We advise for the continuous monitoring of risk models and subsequent recalibration when an overall recalibration is observed to be worsening. Statistical process control is a valuable tool that can visualize performance behaviour over time. Although the simplest form of recalibration rendered our model stable again, sometimes more rigorous techniques may be required [5, 6]. In relation to the previous part, this continuous monitoring might also enhance clinical credibility and evidence for temporal (and external) validity. One may have more trust in a model when there is evidence that the model is up to date and still valid.

The increasing trend in severity of illness together with stable mortality suggests an overall improvement in the provided quality of care of the ICUs as a whole. This coincides with our earlier statement that healthcare technologies are improving and an increasing number of lives can be saved. We also found an increasing proportion of elderly patients (>80 years) in the ICU population which can either be a consequence of the ageing of the Dutch population or a change in admission policy due to an increased number of available beds. The SMR might be biased by changes in admission policy [7]. For example, under-resourced ICUs might choose to only admit patients that fail to respond to treatment (i.e. lead time bias). Although non-responders have greater mortality risk compared to responders, the underlying physiological parameters may be equal in both groups. This means that under-resourced ICUs may be disadvantaged when compared to over-resourced ones, while this does not necessarily imply poor care.

Part 3: Role of prognostic models in individual decision-making

Main findings
We systematically reviewed literature on SOFA-based prognostic models of mortality and/or survival and found that the SOFA score is competitive with admission models such as SAPS-II and APACHE-II while being based on less physiologic parameters (chapter 7). Models combining information of severity of illness at admission and information on the course of illness during treatment achieved better performance. Two strategies to incorporate SOFA in the logistic regression model were shown to be effective: 1) SOFA derivatives (e.g. maximum SOFA score) and 2) frequent sequential patterns of SOFA (e.g. a low score on day 2 followed by a high score on day 3). Temporal models based on these two strategies were competitive with physicians, although they had worse discrimination (chapter 10). The models significantly improved by including subjective information, which made their predictions superior to physicians’ predictions alone. In chapters 8 and 9, we found that, although nurses and physicians have good, comparable predictive ability, they do not always agree and their predictions are inconsistent.

General discussion
Earlier research on expert opinion on the first day of admission also found good discriminative ability [8-19], superior to the present admission models [20], and reasonable calibration [10,11]. In the comparison of mathematical models and physicians, physicians may have an unfair advantage over the models as they make decisions regarding treatment policy. A considerable concern of introducing mathematical models in clinical practice is that of the self-fulfilling prophecy. If the model predicts death and the physician operates ac-
cordingly, the patient will die for sure, unless saved by a medical wonder. This will raise difficulties for validation as the model will be correct for all non-survivors. Although this problem should not be underestimated, the self-fulfilling prophecy is already manifest in the behaviour of human actors (i.e. physicians).

In line with our findings, a combined approach led to better discrimination and calibration over either source alone [12, 15, 21-22]. The advantage of a combined approach is that (besides enhancing acceptability) it uses the strengths of both individual sources. While clinicians have important implicit knowledge outside the scope of mathematical models and can react more quickly to changing situations, mathematical models are extremely consistent and able to optimally combine information into a global judgment [23].

On the contrary, clinicians’ perceptions of mortality risks are hypothesized to be inexact or even inconsistent [24, 25], which was confirmed in chapters 8 and 9. We believe that these inconsistencies are not necessarily bad, but stress the importance of inter-collegial communication between nurses and physicians and the potential role for mathematical models in supporting informed decisions. We believe there is a need for a better understanding of (how to improve) the end-of-life decision-making process and the possible role of nurses and mathematical models. There is paucity in research on the impact and acceptance of different forms of (objective) prognostic information in clinical practice. In addition, it might be valuable to develop models that predict other clinically relevant outcomes such as quality of life after treatment.

We are aware of two studies showing that objective assessments of patient outcomes are able to alter patient management. One study reported an increased use of 13% of specified aspects of intensive care (e.g. mechanical ventilation, intubation, intracranial pressure monitoring) in patients predicted to have a good outcome, but a 39% reduction in the use of these same aspects in patients predicted to have the worst outcome [26]. However, overall outcomes, such as mortality and length of stay, were not shown to be affected. In another study, nurses provided with forecasts of decision outcomes perceived this as helpful information to support families and increase their comfort in making treatment decisions when these families where the controlling party [27]. Although this information was only used if it was consistent with the nurse’s own opinion, they were inclined to advocate for more aggressive treatments when the forecasts confirmed that a patient’s probability of dying was low and less aggressive treatment when the forecasts confirmed this probability was high.

Another use of objective information regarding a patient’s prognosis at the individual level is to inform patients and families about survival chances. To our knowledge there are no studies asking patients and their families if they desire to receive this kind of information. Although families often do not receive this information [28] and doubt the physician’s ability to predict medical futility [29-31], they do value physician-family discussions [29]. Providing objective and/or subjective prognostic information to patients and their families may improve their empowerment in end of life decision-making. It is important to note, however, that families and caregivers may wrongly estimate patient preferences towards care provision. For example, they both underestimate the desire of older patients for aggressive care [32], while caregivers making surrogate decisions based on considerations of treatment outcomes do not always effectively represent patient preferences [33].
In summary, future work should focus on the development of clinically useful models and emphasize 1) their clinical credibility (e.g. by selecting clinically relevant variables and providing predictions that are useful in clinical practice), 2) (external) validity and 3) clinical effectiveness (e.g. impact on clinical decisions and patient outcomes). Models should be continuously monitored over the course of time and updated when their calibration is seen to be worsening. Future studies should investigate the optimal strategy (e.g. number of groups, type of chart) for temporal validation. Finally, more research is needed on how to improve the end-of-life decision-making process and the possible role of objective mathematical models in this process.

References