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
Chapter 1

General introduction
Context
The Intensive Care Unit (ICU) is a department of the hospital with highly trained medical staff and special equipment for the treatment of severely ill patients who need close monitoring and life-support. Intensive care is very expensive and ICUs repeatedly ask the questions whether their care is effective (are we saving lives?) and efficient (are we using our resources wisely?). To help answer both types of questions it is important to be able to predict the probability that a patient will survive his or her hospital stay. In particular mortality prediction is essential in benchmarking and in informing decisions about individual patients. In benchmarking, one compares the quality of care among different centres (ICUs) by using a predictive model as reference. Specifically, one compares the observed to the predicted mortality in each centre by calculating Standardized Mortality Ratios (SMR; i.e. observed mortality divided by predicted mortality). The severity of illness of each patient at his or her admission must be considered in this prediction (this is called case-mix-adjusted prediction). When the observed mortality exceeds the case-mix adjusted predicted mortality in a centre (SMR>1) then the centre is said to be performing worse than expected; when the observed and predicted mortality are equal (SMR=1) then quality of care is as expected; and when the observed mortality is less than predicted (SMR<1) then quality of care is said to exceed what is expected. Benchmarking is hence related to the question whether ICUs are effective. Individual predictions of mortality can also be used to inform decisions about treatments. An important treatment decision is whether to continue or withdraw intensive care treatment for a patient: Although intensive life sustaining treatments may save lives, critically ill patients do not always prefer to undergo these treatments. Patient preferences highly depend on their chances of survival; in fact, if these chances are very low, patients often prefer palliative care aiming at comfort and relief of pain [1,2]. Supporting individual decisions is hence related to the efficiency of the ICU and to whether the provided care meets patients’ wishes.

Objective versus subjective predictions of mortality chances
Several prognostic models have been devised to predict mortality in the ICU. For benchmarking, we are interested in the severity of illness at admission, so we can measure the effect of ICU treatment on patient outcomes by comparing case-mix adjusted predicted mortality to observed outcomes. Commonly known admission models are the Acute Physiology and Chronic Health Evaluation (APACHE) I-IV [3], the Simplified Acute Physiology Assessment (SAPS) I-III [4] and the Mortality Prediction Model (MPM) I-II [5] which rely on physiologic data from the first 24 hours of ICU admission.

For the support of informed individual decisions it is important to have models that provide mortality prediction at any day of ICU stay. Such prediction models, referred to as temporal models, can capitalize on all the information available during ICU stay, up and until the moment the prediction is made. Recently, our research group introduced a method to build such models that incorporate specific patterns of sequential data (e.g. the transition from normal renal function to renal failure) as predictors. These models were shown to have better predictive performance than the present admission models [6].

While clinicians’ perceptions of mortality risks may be inexact or even inconsistent [7,8], objective mathematical models are extremely consistent and able to optimally combine information into a global judgment [9]. Clinicians, however, have important implicit knowledge outside the scope of these models and can react more quickly to changing sit-
General introduction

Valuations [9]. Previous studies on expert opinion report good discriminative ability [10-21], superior to the current admission models [22]. They also report reasonable calibration, although objective models may still be better calibrated [12,13]. However, these studies did not compare daily predictions of physicians and temporal models.

Problem

In this PhD dissertation, we tackle three main problems related to prognostic models and their role in benchmarking and individual decision-making.

First, there is little in literature about the validity of prognostic models of mortality for use in clinical practice. Specifically, there are no reviews describing these models for the elderly (ICU) population. Elderly patients form a particularly interesting target group in which concerns regarding withholding therapy and end-of-life decisions are common, and comfort-focused end-of-life care may be preferred when chances of survival are low.

Second, the validity of benchmarking when using a model over time is not well understood. We hypothesized that model performance decreases over time due to changes in patient mix (e.g. different prevalence of diseases) and the arrival of new technologies in healthcare. This will outdate the model, which will no longer fit the data well and, importantly, jeopardizes the quality of care assessments that are made based on the model. Therefore there is a need to understand model behaviour over time, its effect on quality of care assessment, and if continuous updating of models can alleviate this problem.

Third, although prognostic models may have the potential to improve individual decision-making, there is still considerable resistance against their use in clinical practice [23,24]. The focus is too much on the development of models rather than their application and impact on clinical decisions and patient outcomes. Clinicians often do not trust the models due to lack of evidence about: their clinical credibility (e.g. the use of clinically relevant variables and refraining from adding unnecessary complexity); external validity (i.e. accuracy and generality) especially for their use in individual patients [25]; and clinical effectiveness (i.e. their impact on patient outcomes). In addition, for modelling at the individual patient level it is important to incorporate daily patient information (e.g. degree of organ failure) rather than admission data only, because the patient’s condition may change during ICU stay for either better or worse. There are few models that provide a prediction on each day of stay [6] and not much is known about how their predictions compare to those of physicians. There is a need for a better understanding of clinicians’ subjective predictions, (the need for) individual mathematical predictions, their validity on prospective data and their (possible) impact on the decision-making process and patient outcomes.

Goals

Given the problems stated above, the goals of this PhD dissertation is to better understand characteristics of current prognostic models in the ICU, the implications for their use over time in benchmarking, and the role they can play in supporting decision-making for individual patients. Specifically, we:

1) Identify and characterize prognostic models of mortality or survival for elderly (ICU) patients, their validity and their use in clinical practice (Chapters 2 and 3);
2) Investigate the behaviour of prognostic model performance over time and the effect of this behaviour on quality of care assessment (Chapters 4, 5, and 6);

3) Investigate clinicians’ subjective predictions, (the external validity of) individual mathematical predictions and how these compare to each other (Chapters 7-10).

**General approach and outline**
This PhD dissertation consists of three parts corresponding to the three goals described in the previous section.

**Part 1:**
Due to the increasing relevance of the elderly population to the ICU, the first part of this PhD dissertation contains two systematic literature reviews on prognostic models of mortality or survival for elderly patients. Chapter 2 presents all available prognostic models for elderly patients in different domains of healthcare, their characteristics, reported potential use, and the actual level of their (external) validity. Chapter 3 describes in-depth the characteristics of the ICU prediction models for the elderly patients and their validity for clinical use. In order to extract and assess relevant descriptors, we resorted to literature on barriers to the implementation of prognostic models in practice to identify the relevant descriptor types.

**Part 2:**
The second part contains a set of prospective temporal validation studies in which we assessed the validity of prognostic models over the course of time using techniques based on Statistical Process Control (SPC). SPC techniques combine rigorous time series analysis and graphical data presentation to identify structural changes in a process [26]. In various SPC charts, a statistic, such as mortality risk or SMR, is plotted over time for (patient) groups, and rules are used to judge whether the deviations of the statistic over time is due to chance or otherwise are structural. We used, among others, predictive performance measures as the statistics to be monitored with SPC, which is an innovation in validation studies. Specifically, we conducted validation studies for various statistics pertaining to a logistic regression model, which is a parametric model in which the form of the relationship between predictors and outcome is pre-specified, as well as a binary classification tree model, which is a non-parametric model with little assumptions about this relationship. A binary classification tree repeatedly partitions the sample data into two groups (hence “binary”) based on values of the given predictors. For example, the variable “sex” separates instances for men and women. End-nodes or “leaves” are obtained when a subgroup is not partitioned any more, based on some stopping criterion.

The validation of the logistic regression model is described in chapter 4. Specifically, we assessed the effect of changes in mortality prediction performance of an existing SAPS-II model for very elderly ICU patients on quality of care assessment. The validation of the classification tree model is described in chapter 5. Here we used a similar technique as described in chapter 4 to assess the validity over time of the tree. We assessed changes in case-mix in the population by estimating the percentage of patients falling in each end-node of the classification tree over time. In chapter 6, the parametric SAPS-II model and
the non-parametric classification tree are compared based on their prognostic performance over time in terms of discrimination and calibration.

**Part 3:**

The third part contains a set of studies related to mortality prediction for individual decision-making. As we need models that provide mortality prediction at any day of ICU stay, we first wanted to understand the predictive value of daily scores of the degree of organ failure. An organ failure score that is commonly used in the ICU is the Sequential Organ Failure Assessment (SOFA) score. We conducted a systematic review of the literature on SOFA-based prognostic models of mortality or survival in the ICU. In chapter 7, we describe the characteristics and the validity of these models.

Second, we wanted to understand the reliability of subjective predictions of survival probabilities by clinicians. We conducted a prospective study in the ICU of the Academic Medical Center in which we collected daily predictions of survival by nurses and physicians. In the prospective cohort study described in chapter 8, we assessed their general characteristics and predictive value in terms of discrimination and precision. Chapter 9 describes the consistency of nurses’ daily predictions of survival in terms of inter-observer variance and variance of observers’ assessments over time.

Finally, we built temporal prognostic models which provide a prediction of mortality at each day of ICU stay. We use two strategies described in literature (chapter 7) to incorporate SOFA-based predictors in the logistic regression model: A) SOFA derivatives (e.g. maximum or mean SOFA score) and B) frequent sequential patterns of SOFA (e.g. a low score on day 2 followed by a high score on day 3). We study in a real world setting how these models perform compared to physicians, who are exposed to additional information than the models. The findings of this study are described in chapter 10.

Chapter 11 describes a general discussion of this PhD dissertation.

**References**