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 9

Consistency of nurses’ daily predictions of survival in the Intensive Care

Minne L, Dongelmans DA, de Jonge E, Abu-Hanna A.
Consistency of nurses’ daily predictions of survival in the Intensive Care.
The original publication is available at http://booksonline.iospress.nl/Content/View.aspx?piid=31451
Abstract
In the Intensive Care Unit, clinicians are continuously faced with the difficult task of prognosis, but their predictions of patient survival status may not always be consistent. Specifically very little is known about consistency of predictions over time. The aim of this paper is to assess the consistency of nurses’ daily predictions of survival in terms of inter-observer variance and variance of observers over time. We found a low consistency of these predictions between observers and over time, even though changes in the patients’ condition are considered. Our findings have implications to the process of end-of-life decision-making, which pertains to withholding or withdrawing intensive care treatment.

Introduction
In the Intensive Care Unit (ICU), clinicians are continuously faced with the difficult task of prognosis. The relevant prognostic question in this setting is to predict hospital survival chances of their patients, which is the probability that a particular patient will eventually be discharged alive from the hospital. Although prognosis of survival status strongly influences clinical actions such as provision of life support [1], it is the part of medical practice clinicians feel most insecure about [2,3].

Previous research shows that, on the first day of ICU admission, both nurses and physicians’ predictions have good discriminative ability (i.e. ability to distinguish between survivors and non-survivors) [4]. There could however be substantial disagreement between different clinicians, even among very experienced ones and their predictions are influenced by several external factors, such as ego bias (i.e. more trust in self than in others), reverse ego bias (i.e. less trust in self than in others) and level of experience [5-9]. Therefore, it is often hypothesized [2,3] that their predictions, or those of humans in general [10], are not consistent.

The term consistency in this context refers to either 1) consistency between different observers assessing the same phenomenon (inter-observer reliability) or 2) consistency of the same observer assessing the same phenomenon more than once from one time to another (test-retest reliability) [11]. Knowledge in this area is however restricted to the inter-observer reliability of physicians with different levels of experience and little is known on the consistency of nurses’ predictions of survival in the ICU. More important, there are no published studies that assess the consistency of daily predictions of survival over time, which is closely related to the concept of test-retest reliability, although the predictions over time should be expected to change due to change in the condition of the patient. The aim of this paper is to assess the consistency of nurses’ daily predictions of survival in terms of 1) inter-observer variance and 2) variance of observers over time.

Methods
In this prospective cohort study we included all consecutive patients admitted between March 2009 and June 2010 to a 28-bed multidisciplinary mixed adult ICU of a 1002-bed university hospital in Amsterdam. We extracted demographics, patient outcomes and all data necessary to calculate severity of illness scores from the Dutch National Intensive Care Evaluation (NICE) registry [12], which is officially registered according to the Dutch Personal Data Protection Act.
In addition, we developed a software module to elicit nurses’ estimates of the likelihood of patient survival until hospital discharge. This module was integrated into the Patient Data Management System (PDMS) that is used in the ICU. After entry of daily routine clinical patient data, a self-explaining question popped up automatically in which nurses could choose between 10 probability categories (from 1 to 10) or state that they had no clue regarding the survival chance of a particular patient. Categories were indicated with red (low survival probabilities) and green (high survival probabilities) color-scales. Eliciting these survival probability categories took typically place between 9.00 and 12.00, but estimates could be changed until 0.00. We estimated that this set up would result in three nurses’ estimates of a patient’s hospital survival every 24 hours. Our ICU employs 120 nurses whom were not trained to estimate survival probabilities and did not receive feedback. They were notified of this study by email and a single announcement (EdJ) during a staff meeting and they are unaware of their colleagues’ assessments. Ethical consent for this study was waived by the Institutional Review Board (IRB).

Data were analysed using the statistical environment R version 2.10.1 [13]. Missing predictions were imputed by taking the mean value of the predictions of the previous and next day. When these were not available, the complete patient record was excluded. The consistency of the predictions was assessed in the following ways. First, inter-rater variance was measured by calculating 1) the variance $V_i^j$ per patient $i$ per day $j$, 2) taking the mean value per patient $\text{Mean}V^j_i$ of all prediction days, and 3) taking the mean value for all patients $\text{Mean}V$. Second, variance over time was obtained by 1) taking for each patient $i$ the mean prediction $\text{MeanPr}_i$ per day, e.g. $<0.2, 0.6, 0.4>$, 2) calculating for each observer $k$ the distances between prediction $\text{MeanPr}_{ik}$ and $\text{MeanPr}_i$, e.g. when the predictions of observer 1 are $<0.3, 0.6, 0.2>$ then the distances between this observer and $\text{MeanPr}_i$ are $<0.1, 0.0, -0.2>$, 3) taking the variance $V_i$ of these distances, and 4) taking the mean variance $\text{Mean}V$ for all patients. Third, the inter-rater variance was calculated for each of the first 7 days, to assess if there is a correlation between inter-rater variance and the day of prediction. These analyses were conducted for the entire patient cohort as well as patients with an ICU stay of at least 7 days. The strategy described above is robust for changes in the patient’s condition, because the mean of the predictions per day reflect these changes.

### Results

During our study period, 2,629 patients were admitted. The database contained 29,018 records with nurses’ predictions of survival. 16 (0.0006%) values were missing and in 1,481 (0.05%) of the cases, nurses recorded that they had no idea regarding a patient’s chance of survival. Data of 12 patients (0.005%) could not be completed by imputation resulting in the removal of 25 records (0.001%) of these patients. The remaining 28,993 records of 2,617 patients were included in the analyses. Patient characteristics are shown in Table 9.1. Overall severity of illness by the mean Acute Physiology And Chronic Health Evaluation-II (APACHE-II) score was 17.3 and 45.9 by the mean Simplified Acute Physiology Score-II (SAPS-II) score. The median length of stay was 2 days and mortality was 16%.

The mean inter-rater variance between different nurses was 1.95 for the entire patient cohort (which also includes very short stays) and 3.16 for patients with an ICU stay of at least 7 days. This means that the spread around the mean prediction of a particular patient on a particular day is 1.95 categories and even 3.16 categories for patients with a longer
ICU stay. The mean variance over time was 1.52 in the entire cohort and 2.11 in patients with an ICU stay of at least 7 days. For the entire cohort, this refers to a mean spread of 1.52 categories of each observer around the mean predictions of all observers over time.

<table>
<thead>
<tr>
<th>Table 9.1. Patient characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (median [range])</td>
</tr>
<tr>
<td>Male (%)</td>
</tr>
<tr>
<td>Died (%)</td>
</tr>
<tr>
<td>Length of stay (median [inter-quartile range])</td>
</tr>
<tr>
<td>APACHE II score (mean (standard deviation))</td>
</tr>
<tr>
<td>SAPS II score (mean (standard deviation))</td>
</tr>
<tr>
<td>Admission type (%)</td>
</tr>
<tr>
<td>Medical</td>
</tr>
<tr>
<td>Unplanned surgery</td>
</tr>
<tr>
<td>Planned surgery</td>
</tr>
</tbody>
</table>

Figure 9.1 shows the inter-rater variance for each of the first 7 days of prediction in the entire patient cohort. It can be observed that during the first 7 days of ICU stay, inter-rater variance is closely correlated to the day of prediction (correlation coefficient = 0.93, p<0.01). For patients with a minimum stay of 7 days this relation is less clear as it remains high at a similar level as the last observation of day 7 (data not shown).

Discussion

Nurses have a low level of consistency in predicting patient survival status in the ICU. The variance between predictions of different nurses (inter-rater variance = 1.95) was higher than the variance within observers over time (variance over time = 1.52), especially in long term patients (inter-rater variance = 3.16 vs. variance over time = 2.11). Patients with a longer ICU stay may be more difficult to assess, resulting in more divergent opinions of different nurses. Indeed, while inter-rater variance was closely related to the day of admission in the entire cohort, it was already at about its highest value at the first day of admission in patients with a stay of at least 7 days.
To the best of our knowledge, this is the first study assessing the consistency of nurses’ daily predictions of patient survival status in the ICU. We included a large patient cohort for whom three nurse predictions were available at each day of ICU stay with very few missing values (0.0006%). The main limitation of this study is that the predictions are not always provided by the same nurses, which may bias estimates of inter-observer variance and variance of observers over time. We mitigated this bias, however, by first calculating variance at patient level before taking the mean of all patients. Note also that, although survival status is important, other outcomes (such as functional status) should also be considered in the end-of-life decision-making process.

Previous studies that assessed inter-rater variance in the ICU between physicians and nurses or physicians with different levels of experience generally found the same levels of intra- and inter-rater variance [5-9,14]. Those studies did not assess inter-rater agreement of different nurses, however, and most of them were based solely on the day of admission.

Important clinical decisions regarding the withdrawal or continuation of intensive medical treatment are for a major part based on subjective interpretations of patient survival status. This study shows however that these subjective predictions of survival status by humans are not always consistent. Therefore, one should carefully think about how to improve the end-of-life decision-making process and communication of survival probabilities in the ICU. Future research should focus on this improvement and the possible (supportive) role of objective mathematical models.

References


