



**Prognostic modeling to evaluate the
in-hospital and long-term mortality of
Intensive Care patients**

Sylvia Brinkman

© Sylvia Brinkman, Amsterdam, The Netherlands
ISBN: 978-94-91602-09-2
Printed by Print Service Ede

All rights reserved. No part of this publication may be reproduced, stored or transmitted in any way or by any means without prior written permission of the author.

The thesis was printed with financial support of:
Department of Medical Informatics AMC, Stichting BAZIS, Vektis, , and
.

Prognostic modeling to evaluate the in-hospital and long-term mortality of Intensive Care patients

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 3 april 2013, te 10:00 uur

door

Sylvia Brinkman
geboren te Utrecht

Promotiecommissie

Promotores: Prof. Dr. A. Abu-Hanna
Prof. Dr. E. de Jonge

Co-promotores: Dr. N.F. de Keizer

Overige leden: Prof. Dr. M.B. Vroom
Prof. Dr. P. Pickkers
Prof. Dr. E.W. Steyerberg
Prof. Dr. C.J. Kalkman
Prof. Dr. A.H. Zwinderman

Faculteit der Geneeskunde

Contents

Chapter 1: General introduction	9
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	17
Chapter 3: External validation of Acute Physiology and Chronic Health Evaluation IV in Dutch intensive care units and comparison with Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II	33
Chapter 4: Hospital mortality is associated with ICU admission time	49
Chapter 5: Determinants of mortality after hospital discharge in intensive care patients; Literature review and Dutch cohort study	61
Chapter 6: Mortality after hospital discharge in ICU patients	81
Chapter 7: Prediction of long-term mortality in ICU patients: Model validation and assessing the effect of using the in-hospital versus long-term mortality on benchmarking	95
Chapter 8: General discussion	107
Summary & Samenvatting	117
Dankwoord (Acknowledgements)	127
Curriculum Vitae & List of publications	131

Chapter 1

General introduction



1.1 Introduction

Over the last three decades interest in quality of health care has increased by health care professionals and policy makers due to the perceived need of continuous quality improvement and transparency to the public. To get insight in the quality of health care, quality indicators such as in-hospital mortality rates are frequently used. When comparing hospitals by mortality rates the high and low performing hospitals can be identified (1,2), suggesting that hospitals with lower mortality perform better than hospitals with higher mortality. However, differences in patient case-mix (e.g. diagnoses, age, and sex) influence mortality rates of hospitals. High mortality risk, for example, could be an indication for more complex and severely ill admitted patients to the hospital. Consequently, fair and meaningful comparison of the performance of hospitals based on mortality rates requires adequate case-mix adjustment of the raw mortality data. This means that patient characteristics that influence mortality should be considered. To this end, several initiatives have been undertaken to collect the relevant raw data about the patient in medical quality registries. In these quality registries data collection should be standardized to provide reliable case-mix adjustment and to enable quality monitoring, quality improvement, and comparison of different hospitals. Case-mix adjustment is commonly done by using prognostic models. To date, several prognostic models have been developed for the total hospital population and also for specific patient groups such as the Intensive Care Unit (ICU) population or the cardiac surgery population.

The prognostic models that are frequently used in the domain of ICU predict in-hospital mortality risk while adjusting for case-mix. The long-term mortality is often ignored, as it is harder to obtain the relevant data after patients leave the hospital. However, there are several reasons to consider the long-term mortality next to the in-hospital mortality. One of the reasons is that the in-hospital mortality can differ between hospitals due to differences in hospital discharge policies, the number of patients transferred to another hospital, and/or the number of patients admitted from another hospital. Ignoring these factors may when considering in-hospital mortality lead to biased results. More important is that for patients the long-term mortality is more relevant than the in-hospital mortality. This thesis addresses the validation and comparison of different prognostic models currently used in health care with a main focus on the models used for the ICU population. This first chapter introduces the domain of intensive care and describes the quality assessment in the intensive care by using prognostic models. Next, we underline the importance of assessing the long-term mortality of ICU patients. This chapter concludes with an outline of the thesis.

1.2 Intensive Care

Intensive care is defined as “a service for patients with potentially recoverable conditions who can benefit from more detailed observation and more invasive treatment than can safely be provided in general wards or high dependency areas”. This definition originates from the second half of the 20th century and since then the intensive care has been much expanded. Nowadays, care in the ICU is very complex and delivered in a highly technical and labor-intensive environment. Along with these developments also the cost of intensive care has increased substantially, resulting in a high proportion of the health care budget spent for the ICU (3). In 2011 there were 91 Dutch ICUs which were categorized by the Netherlands Society for Intensive Care (NVIC) in three levels: 46 level 1 ICUs, 22 level 2 ICUs, and 23 level 3 ICUs. Level 1 ICUs are the smaller less equipped ICUs which should contact a higher level ICU in the region to discuss the possible need for transfer after a treatment period of 3 days. Level 3 ICUs are the larger and most equipped ICUs. As almost all ICUs offer all types of care, the average volume per type of treatment per hospital is relatively low. This low volume per type of treatment is increasingly leading to discussion about centralization of the Dutch ICUs. It has been postulated that redistribution of the Dutch ICUs can lead to lower costs while improving quality and thereby reducing mortality (4). Currently, around 23% of the total Dutch hospital mortality is attributable to the death of ICU patients. The average mortality among all hospitalized patients is approximately 1.5%, while the average mortality among ICU patients is approximately 20%.

1

1.3 National Intensive Care Evaluation foundation

The National Intensive Care Evaluation (NICE) foundation (5) was established in 1996 by a group of intensivists to facilitate quality monitoring and quality improvement initiatives in Dutch ICUs. At the start of the registry, information was collected on severity of illness at ICU admission and outcome only. Over the years the registration has been extended with information on daily organ functioning, complications, quality indicators (e.g. nurse-to-patient ratio, glucose regulation, and duration of mechanical ventilation), and will be further extended with nursing workload information. The NICE registry performs analyses on the registered data, provides feedback reports, and offers the participants an online tool for monitoring quality indicators. The feedback reports and online tool can be used to compare the quality of care between ICUs but also for monitoring the quality of care of an ICU over time. This approach allows the identification of critical points in the care process and facilitates meaningful discussions about ICU treatment and organization improvements. Subsequently this can lead to actual improvement of quality of care at the ICU.

Currently, 85 ICUs are participating in the NICE registry, which amounts to about 90% of all Dutch ICUs. The participating ICUs are mixed medical-surgical units located in university hospital, teaching hospital or non-teaching hospital and are widespread over the Netherlands. The registry contains information of more than 600,000 ICU admissions providing an important basis for quality assessment and scientific research. As the ICU population is a heterogeneous group of patients with different reasons for ICU admission, varying severity of illness and different ages it is meaningless to directly compare the outcome (mortality) of different ICUs. This thesis is focused on using the registry data for case-mix adjustment to allow for comparison of ICUs.

1.4 Prognostic models used for case-mix adjustment

Prognostic models use different variables to adjust for case-mix differences, some use administrative data and others use clinical data for adjustment. Also the clinical data that are used differ between the models. The hospital standardized mortality (HSMR) model (9,10) is developed for all hospitalized patients and uses routinely collected administrative data to correct for case-mix differences. The Acute Physiology And Chronic Health Evaluation (APACHE) II and IV, and the Simplified Acute Physiology Score (SAPS) II models are developed specifically for ICU patients and use clinical data for case-mix correction (6-8). Although these last three models use clinical data, they differ in the specific clinical data that are used. As example, the APACHE II and APACHE IV models use, in contrast to the SAPS II model, the reason for ICU admission as covariate. All mentioned models have in common that they are logistic regression models providing an estimate of the in-hospital mortality risk, which is a probability, based on the characteristics of the patients (e.g. severity of illness, demographics etc.). There are diverse applications of the prognostic models in practice (11). Prognostic models can be used to identify high risk patients and as tool to correct for the severity of illness when comparing different treatment groups (i.e. high predicted mortality risks indicate high severity of illness). For instance when comparing the outcome of patients with and patient without a certain treatment the results have to be adjusted for the difference in severity of illness between the two groups. A common application of prognostic models is the use for benchmarking purposes, i.e. the process of comparing a quality indicator across different hospitals and/or with a reference value. One of the most frequently used quality indicators for benchmarking hospitals is the standardized mortality ratio (SMR). The SMR is the ratio of the observed in-hospital mortality and the expected in-hospital mortality as estimated by the prognostic models. The SMR is used by physicians and managers as the case-mix adjusted outcome to measure the quality of care, to compare their outcomes to the national average or to perform longitudinal analyses. The SMR can be calculated for the total population but also for specific subgroups for example patients with a specific reason

for ICU admission or patients admitted on a certain day of the week or in a certain hour of a day. The use of prognostic models to assess the quality of care has been the subject of discussion in many studies (2,12,13). Research has shown that different prognostic models result in different impressions about the quality of care within the same hospital (14,15). These differences might be caused by the fact that each model emphasizes different aspects of the case-mix or that the models are developed in different years for different patient populations. To correct for the latter factor it has been suggested to customize (i.e. also called recalibrate) the prognostic models before using them on a new setting (e.g. different patient population, other country, different year of admission etc.) (16,17).

1.5 Long-term mortality of ICU patients

The prognostic models commonly used for case-mix adjustment all predict the in-hospital mortality. However, it has been postulated that benchmarking of ICUs ideally should be done based on long-term mortality instead of the frequently used in-hospital mortality. In the Netherlands the observed mortality in the ICU population 3 months after hospital discharge is 5.4%. This additional mortality after hospital discharge can be partly explained by the additional mortality caused by the preceding ICU admission or by the preexisting comorbidities of the ICU patients. As the mortality in the first three months after hospital discharge is substantial, it is worthwhile to analyze the influence of assessing the SMR based on the in-hospital mortality or based on the long-term mortality. Besides the substantial mortality in the first three months after hospital discharge there are several more reasons to consider the long-term mortality instead of the in-hospital mortality. First, the in-hospital mortality can be influenced by the difference in hospital discharge policies (e.g. a hospital that discharges many patients to a hospice or that transfers many patients will have a high hospital survival, even if it has properly corrected for differences in case-mix) (18). Second, insight in the factors that influence long-term mortality could be used by clinicians and can help to identify interventions that might improve long-term mortality (19). Third, the true goal of ICUs is long-term survival with an acceptable quality of life (20). Finally, the high resource consumption and costs of ICUs above that of general hospital care (21) may be considered excessive if long-term survival is poor. One important disadvantage of using the long-term mortality of hospitalized patients is that the needed data is harder to obtain than in-hospital mortality. In the Dutch ICUs there is no systematic approach to assess and register the long-term information. Besides, when the long-term mortality of ICU patients is described in the literature, it is difficult to compare these outcomes between the studies due to differences in follow-up starting point (ICU admission, ICU discharge, or hospital discharge) and follow-up end-point. Linkage of hospital data with death registers provides a feasible method for obtaining long-term mortality after hospitalization. This linkage allows assessment of long-term

mortality and the development and validation of case-mix adjusted prognostic models of ICU long-term mortality.

1.6 Objectives of this thesis

Although studies have been published examining the performance of prognostic models and the outcome of ICU patients, still many questions remain. For instance which prognostic model has the highest performance? What happens after hospital discharge of ICU patients? The three main research questions and corresponding objectives answered in this thesis were:

1. How do prognostic models perform in the Dutch ICU population?
 - a. Assess and compare the performance of commonly used prognostic models in health care especially in the field of intensive care.
2. Is there an association between ICU admission time and the case-mix adjusted mortality?
 - a. Determine the association between ICU admission time and mortality by calculating the relative risk while adjusting for severity of illness.
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?
 - a. Examine and describe the long-term mortality of ICU patients.
 - b. Assess the performance of the APACHE IV model when applied for long-term mortality predictions of ICU patients.
 - c. Assess the effect of using the in-hospital mortality versus the long-term mortality on the SMR.

1.7 Outline of this thesis

In Chapter 2 we compared the performance and robustness of a model based on administrative data (the customized HSMR model, originally developed for all hospitalized patients) and a model based on clinical data (the customized SAPS II model, originally developed for ICU patients) in the Dutch ICU population. Furthermore, the effect of severity of illness on the calculated SMRs based on these two models is examined. The postulation in this study was that prognostic models based on administrative data can provide valid adjusted in-hospital mortality rates in specific patient populations. In Chapter 3 we performed an external validation of the APACHE IV model, developed in 2006, in the Dutch ICUs. Before newly developed models can be incorporated by quality registries, the models should be validated and compared to older existing models to assure adequate quality of the used prognostic models for case-mix adjustment. In Chapter 3 we assessed the performance of the APACHE IV model and compared this performance to that of the SAPS II and APACHE II models. All of these three models are based on clinical data and were specifically developed for the ICU patients.

Prognostic models are frequently used to calculate the case-mix adjusted SMR for benchmarking purposes. In Chapter 4 we compared the SMR of ICU patients admitted during office hours and of ICU patients admitted during off hours to determine whether the case-mix adjusted in-hospital mortality is associated with ICU admission time.

The prognostic models mentioned above all predict the in-hospital mortality, disregarding the long-term mortality. In Chapter 5 we conducted a literature review on the long-term outcome of ICU patients. In addition, we describe the mortality after hospital discharge in Dutch patients surviving ICU admission in Chapter 6. In Chapter 7 we validated the performance of the APACHE IV model when applied for the prediction of the long-term mortality of ICU patients. Furthermore, we analyzed the effect of using the in-hospital mortality versus the long-term mortality on the SMR.

Finally, in Chapter 8 we provided an overall discussion of our major results and their implications and addressed the merits and limitations of the different studies.

References

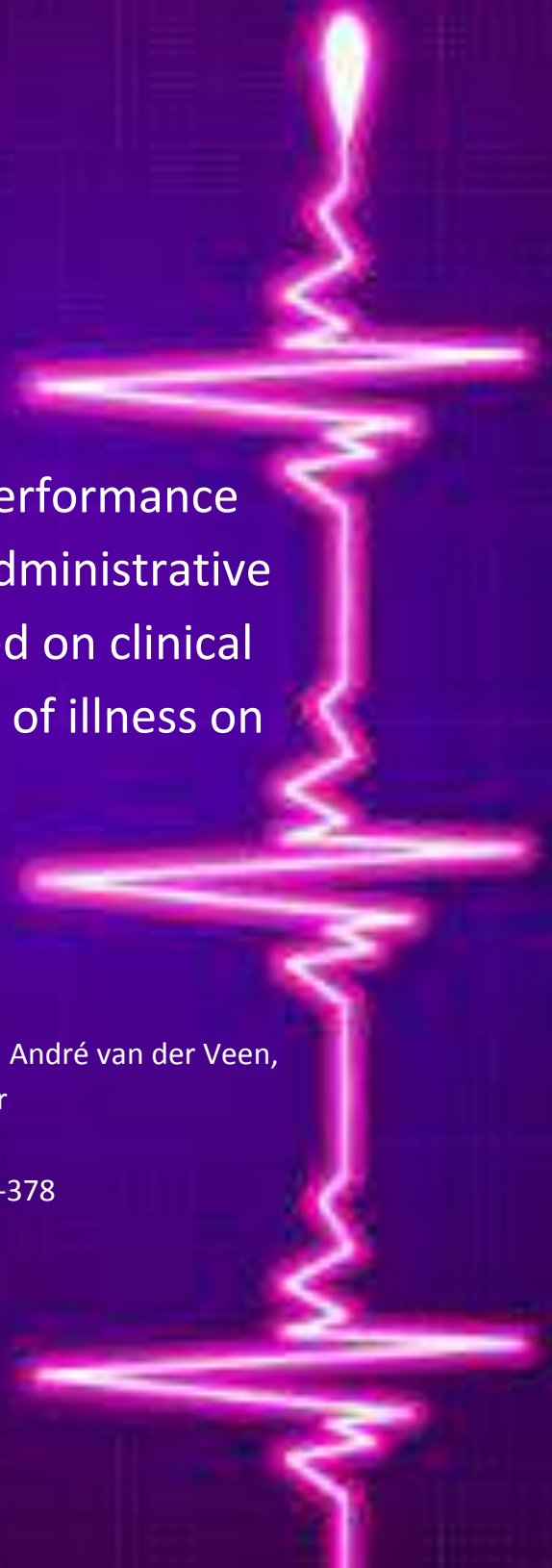
1. Lilford R, Mohammed MA, Spiegelhalter D, et al: Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004; 363:1147-1154
2. Zimmerman JE, Alzola C, Von Rueden KT: The use of benchmarking to identify top performing critical care units: a preliminary assessment of their policies and practices. *J Crit Care* 2003; 18:76-86
3. Halpern NA, Pastores SM: Critical care medicine in the United States 2000-2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. *Crit Care Med* 2010; 38:65-71
4. Rapport Concentratie specialisatie en samenwerking van ziekenhuiszorg: Available online at: http://www.regieraad.nl/fileadmin/www.regieraad.nl/publiek/Downloads/Regie_op_kwaliteit/Rapport_Concentratie_specialisatie_en_samenwerking_van_ziekenhuiszorg_120127_DEF.pdf, last access at 12-12-2012.
5. Dutch National Intensive Care Evaluation (NICE) foundation: Available online at: <http://www.stichting-nice.nl/aboutus.jsp>, last access at 12-12-2012
6. Knaus WA, Draper EA, Wagner DP, et al: APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13:818-829
7. Le Gall JR, Lemeshow S, Saulnier F: A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270:2957-2963
8. Zimmerman JE, Kramer AA, McNair DS, et al: Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34:1297-1310
9. Jarman B, Bottle A, Aylin P, et al: Monitoring changes in hospital standardised mortality ratios. *BMJ* 2005; 330:329
10. Jarman B, Pieter D, van der Veen AA, et al: The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? *Qual Saf Health Care* 2010; 19:9-13
11. Minne L, Ludikhuize J, de Jonge E, et al: Prognostic models for predicting mortality in elderly ICU patients: a systematic review. *Intensive Care Med* 2011; 37:1258-1268
12. Jacobson B, Mindell J, McKee M: Hospital mortality league tables. *BMJ* 2003; 326:777-778
13. Iezzoni LI: The risks of risk adjustment. *JAMA* 1997; 278:1600-1607
14. Bakhshi-Raiez F, Peek N, Bosman RJ, et al: The impact of different prognostic models and their customization on institutional comparison of intensive care units*. *Crit Care Med* 2007; 35:2553-2560
15. Iezzoni LI, Shwartz M, Ash AS, et al: Predicting in-hospital mortality for stroke patients: results differ across severity-measurement methods. *Med Decis Making* 1996; 16:348-356
16. Moreno R, Miranda DR, Fidler V, et al: Evaluation of two outcome prediction models on an independent database. *Crit Care Med* 1998; 26:50-61
17. Teres D, Lemeshow S: When to customize a severity model. *Intensive Care Med* 1999; 25:140-142
18. Kahn JM, Kramer AA, Rubenfeld GD: Transferring critically ill patients out of hospital improves the standardized mortality ratio: a simulation study. *Chest* 2007; 131:68-75
19. Angus DC, Carlet J: Surviving intensive care: A report from the 2002 Brussels Roundtable. *Intensive Care Medicine* 2003; 29:368-377
20. Keenan SP, Dodek P, Chan K, et al: Intensive care unit admission has minimal impact on long-term mortality. *Crit Care Med* 2002; 30:501-507
21. Williams TA, Dobb GJ, Finn JC, et al: Long-term survival from intensive care: a review. *Intensive Care Med* 2005; 31:1306-1315

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 \hat{C} -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.

Materials and methods

Prognostic models

The SAPS II model is a logistic regression model that use 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 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 \hat{C} -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.

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

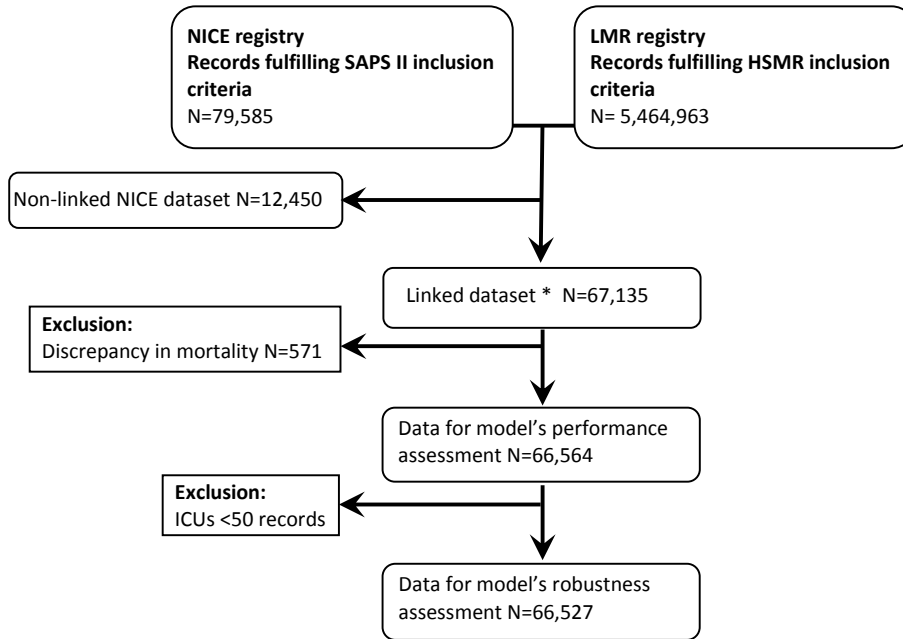
From June 1, 2005 to January 1, 2008, 79,585 patients fulfilling the cSAPS II inclusion criteria were admitted to the 55 ICUs included in the study. The included ICUs are mixed medical-surgical units located in university hospitals (n=6), teaching hospitals (n=22) or non-teaching hospitals (n=27). In 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.

2

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

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

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



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

Figure 2.1: Flow diagram of the included data

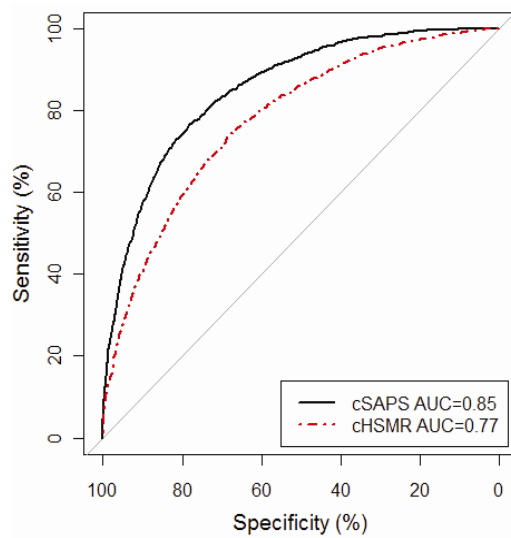


Figure 2.2: AUC curves of the cHSMR and cSAPS II model

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.

2

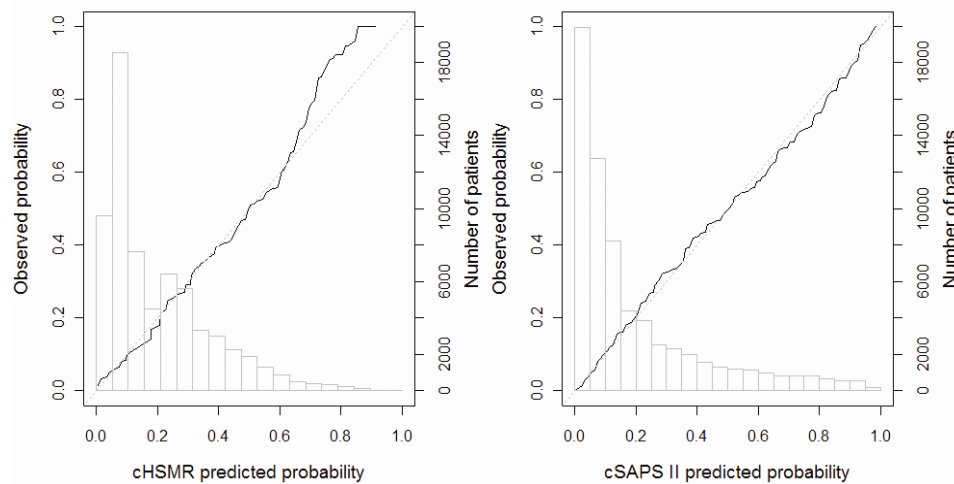


Figure 2.3: Calibration plots of the cHSMR and cSAPS II model. The histograms represent the number of patients per risk group.

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.

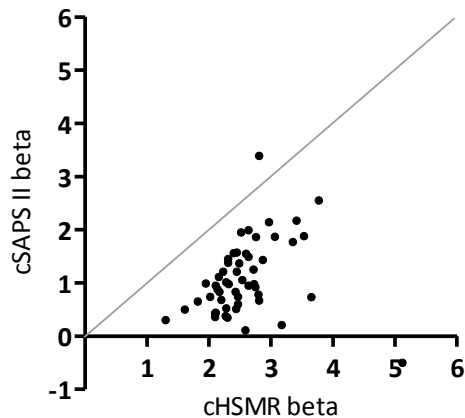


Figure 2.4: Beta coefficients of the 53 included ICUs based on the two customized models. The beta coefficients represents the mean change of SMR per unit of change of severity of illness. A robust and well calibrated model should show beta coefficients of 0, independent of severity of illness.

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 \hat{C} -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 \hat{C} -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

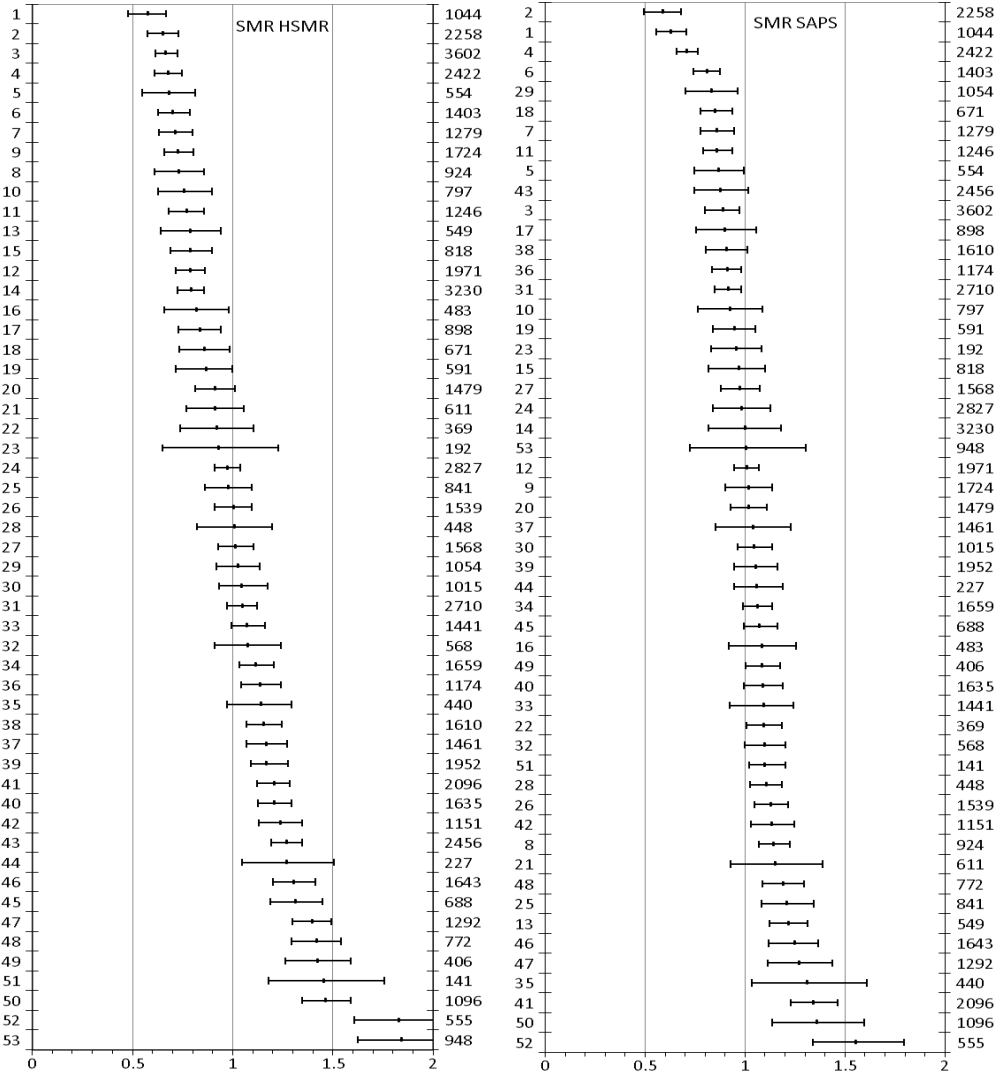
We express our gratitude to the foundation Dutch hospital Data for providing us the LMR data that was required to accomplish this study.

2

References

1. Le Gall JR, Lemeshow S, Saulnier F: A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270:2957-2963
2. Jarman B, Aylin P: Death rates in England and Wales and the United States: variation with age, sex, and race. *BMJ* 2004; 329:1367
3. Jarman B, Bottle A, Aylin P, et al: Monitoring changes in hospital standardised mortality ratios. *BMJ* 2005; 330:329
4. Jarman B, Pieter D, van der Veen AA, et al: The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? *Qual Saf Health Care* 2010; 19:9-13
5. Penfold RB, Dean S, Flemlons W, et al: Follow the Big Dots. *HealthcarePapers* 2008 2009; 8:69-75
6. HSMR: Available online at: <http://www.hsmr.nl/index.php/hsmr>.
7. Jarman B, Gault S, Alves B, et al: Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999; 318:1515-1520
8. Aylin P, Bottle A, Majeed A: Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ* 2007; 334:1044
9. Bakhshi-Raiez F, Peek N, Bosman RJ, et al: The impact of different prognostic models and their customization on institutional comparison of intensive care units*. *Crit Care Med* 2007; 35:2553-2560
10. Zhu BP, Lemeshow S, Hosmer DW, et al: Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. *Crit Care Med* 1996; 24:57-63
11. Moreno R, Apolone G: Impact of different customization strategies in the performance of a general severity score. *Crit Care Med* 1997; 25:2001-2008
12. Dutch National Intensive Care Evaluation (NICE) foundation: Available online at: <http://www.stichting-nice.nl/aboutus.jsp>.
13. National Medical Registration (LMR): Available online at: http://www.swov.nl/uk/research/kennisbank/inhoud/90_gegevensbronnen/inhoud/lmr.htm.
14. Centraal Bureau voor de Statistiek: Documentatierapport Landelijke Medische Registratie (LMR). Available online at: <http://www.cbs.nl/NR/rdonlyres/AA18B546-CA6E-40C6-8B17-1BB976F1C4E5/0/lmrmicrodata.pdf>.

15. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29-36
16. Harrell FE, Jr., Lee KL, Mark DB: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15:361-387
17. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, et al: Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001; 54:774-781
18. Steyerberg EW, Bleeker SE, Moll HA, et al: Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol* 2003; 56:441-447
19. Mason DM, Newton MA: A Rank Statistics Approach to the Consistency of a General Bootstrap. *Ann Statist* 1992; 20:1611-1624
20. Iezzoni LI, Ash AS, Coffman GA, et al: Predicting in-hospital mortality. A comparison of severity measurement approaches. *Med Care* 1992; 30:347-359
21. Iezzoni LI, Shwartz M, Ash AS, et al: Predicting in-hospital mortality for stroke patients: results differ across severity-measurement methods. *Med Decis Making* 1996; 16:348-356
22. Lilford R, Pronovost P: Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 2010; 340:c2016
23. Beck DH, Smith GB, Pappachan JV, et al: External validation of the SAPS II, APACHE II and APACHE III prognostic models in South England: a multicentre study. *Intensive Care Med* 2003; 29:249-256
24. Harrison DA, Brady AR, Parry GJ, et al: Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med* 2006; 34:1378-1388
25. Shahian DM, Silverstein T, Lovett AF, et al: Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. *Circulation* 2007; 115:1518-1527
26. Welke KF, Diggs BS, Karamlou T, et al: Comparison of pediatric cardiac surgical mortality rates from national administrative data to contemporary clinical standards. *Ann Thorac Surg* 2009; 87:216-222
27. Mack MJ, Herbert M, Prince S, et al: Does reporting of coronary artery bypass grafting from administrative databases accurately reflect actual clinical outcomes? *J Thorac Cardiovasc Surg* 2005; 129:1309-1317
28. Arts D, de Keizer NF, Scheffer GJ, et al: Quality of data collected for severity of illness scores in the Dutch National Intensive Care Evaluation (NICE) registry. *Intensive Care Med* 2002; 28:656-659
29. Peek N, Arts DG, Bosman RJ, et al: External validation of prognostic models for critically ill patients required substantial sample sizes. *J Clin Epidemiol* 2007; 60:491-501
30. Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, et al: External validation of Acute Physiology and Chronic Health Evaluation IV in Dutch intensive care units and comparison with Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II. *J Crit Care* 2011; 26:105-108



2

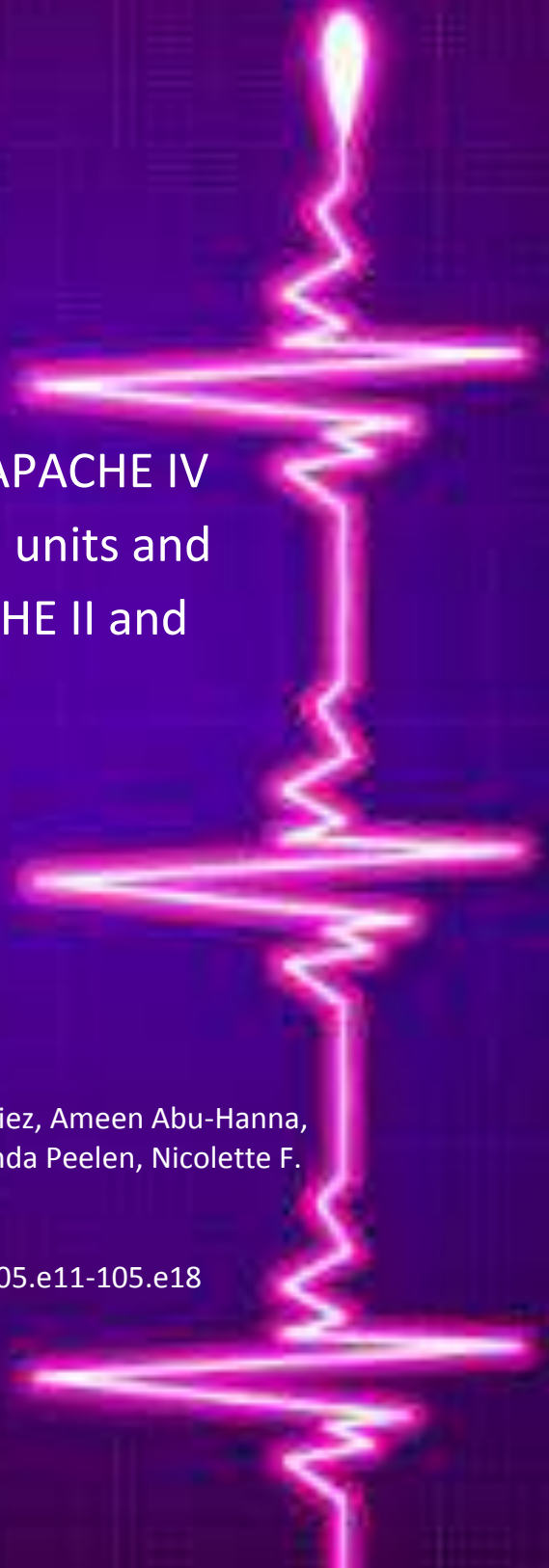
Appendix: The median SMR and the associated confidence interval (CI) for each ICU according to both customized models. The right y axis of each figure represents the number of ICU admissions, and the left y axis represents the hospital ID.

Chapter 3

External validation of APACHE IV
in Dutch intensive care units and
comparison with APACHE II and
SAPS II

Sylvia Brinkman, Ferishta Bakhshi-Raiez, Ameen Abu-Hanna,
Evert de Jonge, Robert J. Bosman, Linda Peelen, Nicolette F.
de Keizer

Journal of Critical Care 2011; 26(1):105.e11-105.e18



Abstract

Purpose: To validate and compare the performance of the Acute Physiology and Chronic Health Evaluation (APACHE) IV in the Dutch intensive care unit (ICU) population to the APACHE II and Simplified Acute Physiology Score (SAPS) II.

Materials and Methods: Prospective study based on data from a national quality registry between 2006 and 2009 from 59 Dutch ICUs. The validation set consisted of 62,737 patients, the three models were compared using 44,112 patients. Measures of discrimination, accuracy, and calibration (Area Under the receiver operating characteristic Curve (AUC), Brier score, R^2 , and \hat{C} -statistic) were calculated using bootstrapping. Additionally, the standardized mortality ratios (SMRs) were calculated.

Results: The original APACHE IV showed good discrimination and accuracy (AUC=0.87, Brier score=0.10, R^2 =0.29) but poor calibration (\hat{C} -statistic=822.67). Customization significantly improved the performance of the APACHE IV. The overall discrimination and accuracy of the customized APACHE IV were statistically better and the overall \hat{C} -statistic was inferior to those of the customized APACHE II and SAPS II, but these differences were small in perspective of clinical use.

Conclusions: The three models have comparable capabilities for benchmarking purposes after customization. Main advantage of APACHE IV is the large number of diagnoses which enable subgroup analysis. The APACHE IV coronary artery bypass grafting (CABG) model has a good performance in the Dutch ICU population and can be used to complement the three models.

Introduction

Over the last three decades, prognostic models have been developed for calculating case-mix adjusted hospital mortality probabilities for patients hospitalized in intensive care units (ICUs), based on routinely collected demographic, physiological and clinical data (1-4). By analyzing and comparing case-mix adjusted outcomes within and between ICUs (i.e. benchmarking), critical points in the care process may be identified which subsequently can lead to improvement of health care.

Old prognostic models such as the Acute Physiology and Chronic Health Evaluation (APACHE) II developed in 1985 (1) and the Simplified Acute Physiology Score (SAPS) II developed in 1993 (3) are still frequently used. But with the introduction of the APACHE IV model in 2006, it has been suggested that the older models should no longer be used because they become increasingly inaccurate (4) (The APACHE is a registered trademark of Cerner Corporation, Kansas City). In contrast to the older models, the APACHE IV model also calculates the predicted in-hospital mortality for patients admitted after coronary artery bypass grafting (CABG). It has been shown that the APACHE IV model has good discrimination and calibration in the general U.S. ICU population with varying performance in different subgroups (4,5). However, the performance of the APACHE IV model has not yet been validated outside a United States population. Such external validation is necessary if APACHE IV is to be used outside the United States (6,7).

The aim of this study was to perform an external validation of the APACHE IV by evaluating its performance in a large representative database of Dutch ICU admissions and by comparing the performance of the APACHE IV model with that of the older APACHE II and SAPS II models.

Materials and methods

Prognostic models

The APACHE IV, APACHE II and SAPS II are logistic regression models that use different predictor variables to predict mortality risk. The models incorporate different patient inclusion criteria. Importantly, only the APACHE IV model predicts mortality risks for CABG patients. Measurements that are unique for the CABG model are gender, number of grafts, diabetes, prior CABG surgery and myocardial infarction during the current hospitalization (4,8).

Data

The Dutch National Intensive Care Evaluation (NICE) registry contains data on consecutive patients admitted to participating Dutch ICUs since 1996. The registry contains all demographic, physiological and clinical variables required to calculate mortality risk predictions according to the APACHE IV, APACHE II and SAPS II models. The data for the prognostic models are collected according to the general rules and definitions for these models (1,3,4). To assure quality of the data in the NICE registry all participants of the NICE registry are obliged to attend training in collecting the data accurately, according to the stated data definitions reported in the NICE data dictionary. The data are automatically checked for range and consistency both locally and centrally. Furthermore, an onsite data quality audit is in place to ensure the validity of the data (9). This study used consecutive admissions from the NICE registry with data on patients admitted to 59 Dutch ICUs between June 1, 2006 and January 1, 2009. Performance of the APACHE IV was evaluated in the total population and in the following subgroups: CABG and non-CABG patients; elective surgery, urgent surgery, and medical (nonsurgical) patients and the 5 most common primary reasons for ICU admission in the NICE registry (bacterial pneumonia, cardiac arrest, colon/rectal cancer, abdominal aortic aneurysm, and thoracotomy for lung cancer).

To compare the performance of the three models, a shared cohort of admissions satisfying their combined inclusion criteria was used. The exclusion criteria of each model are given in table 3.1. Statistical analyses were performed using the statistical environment R version 2.6.2 and SPSS version 16.0.

Table 3.1: Number of ICU admission excluded for one or more of the prognostic models.

Exclusion criterion	APACHE IV	APACHE II	SAPS II	Admissions (%)
Age <18 yrs			x	1,073 (1.2)
Age <16 yrs	x	x		584 (0.7)
Length of stay IC < 8 hrs		x		8,914 (10.0)
Length of stay IC < 4 hrs	x			5,022 (5.6)
Length of stay hosp > 365 days	x			80 (0.1)
Missing APACHE II diagnosis		x		7,999 (8.9)
Missing APACHE IV diagnosis	x			13,308 (14.9)
Missing admission type	x	x	x	3,318 (3.7)
Missing hospital discharge date	x	x	x	566 (0.6)
Burns	x	x	x	112 (0.1)
Patients with transplants	x			56 (0.1)
Cardiac surgery		x	x	15,984 (17.8)
Readmissions	x	x	x	6,422 (7.2)
Admissions from another ICU	x			4,485 (5.0)
Transferred to ICU other hospital			x	1,232 (1.4)

* Total number of included admissions for APACHE IV validation is 62,737 and for comparison of the models is 44,112.

Customization of the models

To date, there have been no studies that have shown that prognostic models are stable over time, in a new setting, and with different case-mixes (10-13). For this reason the prognostic models should be customized on the new setting before the models can be applied for risk prediction. In this study first level customization was used in which a new logistic regression model was refitted with the in-hospital mortality as dependent variable and the logit-transformed original probability as the sole independent variable (14-16). For the validation of the APACHE IV model, the APACHE IV CABG and non-CABG models were customized in their own cohort containing all patients eligible for the APACHE IV CABG resp. non-CABG model. To compare the APACHE IV non-CABG model, APACHE II model, and SAPS II model, the three prognostic models were customized in the shared cohort. As the models are customized before comparison, it also corrects for the differences in the year of their development.

Performance assessment

To describe the discrimination of the models, the Area Under the receiver operating characteristic Curve (AUC) was used (17). The AUC is the probability that a randomly selected non-survivor will have a higher predicted probability to die than a randomly selected survivor. Perfect discrimination corresponds to AUC=1, and no discrimination to AUC=0.5. The Brier score (18) was used to assess overall accuracy. The Brier score is the mean squared difference between the observed and predicted outcome, which includes both discrimination and calibration aspects. As the Brier score is dependent on the prevalence of mortality, the sums-of-squares R^2 measure is also given (19). The R^2 measure ranges between 0 (worst performance) and 1 (best performance) and is adjusted for the mortality prevalence which makes it more suitable for comparison among different subgroups.

Good calibration of a model is essential for using the model for benchmark purposes. To evaluate the models' calibration, the Hosmer-Lemeshow \hat{C} -statistic was used in which observations are grouped based on deciles of predicted probability and compared to the proportions of the actual outcomes. The \hat{C} -statistic, which is a Chi-squared statistic, is tested with ten degrees of freedom for the original, and eight degrees of freedom for the customized APACHE IV model (20). The test is highly sensitive to sample size, in very large datasets it indicates lack of fit for even very small deviations. Therefore it should be compared only within the same sample (21). Furthermore, calibration plots and the unreliability index U originally proposed by Cox (22) of the customized models are given. The index U has a Chi-squared distribution with two degrees of freedom.

In addition, the SMR was calculated. This is the ratio of observed to expected hospital deaths. An SMR of 1 implies perfect prediction of the model, an SMR below 1 implies overprediction and an SMR above 1 implies underprediction.

External validation and comparison of APACHE IV with APACHE II and SAPS

The performance of the models were assessed in the entire sample and for the different subgroups using the ordinary bootstrap method (23) with 1000 samples. The prognostic models were customized on each of the 1000 bootstrap samples of the original dataset and their performance (Brier score, R^2 , \hat{C} -statistic, and SMR) and the associated 95% confidence intervals were obtained from the original dataset. The 95% confidence intervals are defined using the 2.5 and 97.5 percentiles of the bootstrap distribution. To compare the customized APACHE IV, APACHE II, and SAPS II models, the same procedure was followed but the “original dataset” consisted of the shared cohort. A difference between the models was considered statistically significant if the confidence intervals did not overlap. Note that first level customization on any bootstrap sample yields the same AUC when tested on the original dataset (because customization does not affect the order of the probabilities, only their absolute magnitude). This means that the reported AUCs are the AUCs of the original models.

Results

Data

From June 1, 2006 to January 1, 2009, 62,737 patients fulfilling the APACHE IV inclusion criteria were admitted to 59 Dutch ICUs included in the study. Approximately 75% of the 59 Dutch ICUs collect the physiological and clinical variables by using the patient’s status and charts manually. The other 25% of the participating hospitals made use of a Patient Data Management System (PDMS) to collect the physiological and clinical variables of the patients. It has been reported that the use of the PDMS results in a higher mortality prediction, as physiologic values used in the severity scores are registered continuously in the PDMS, thereby increasing the risk of measuring abnormal values (24).

All the participating ICUs are mixed medical-surgical units located in university hospital (n=3), teaching hospital (n=23) or non-teaching hospital (n=33). The 59 ICUs are of diverse IC-levels, are widespread over the Netherlands and are a representative sample of approximately 65% of all Dutch ICUs. Table 3.2 shows the demographics of the APACHE IV non-CABG cohort, the subgroups based on the five most common primary APACHE IV reasons for admission, the APACHE IV CABG cohort, and of the shared cohort. Missing values are not reported separately as the percentage missing values was below 0.2% for all variables used in the analyses.

Table 3.2: Demographics of ICU admissions included in the analysis. Shared cohort: patients satisfying the inclusion criteria of the APACHE IV, APACHE II, and SAPS II model

	APACHE IV non-CABG	Bact. pneumonia	Cardiac arrest	Colon/rectal cancer	Abd. aortic aneurysm	Thoracotomy lung cancer	APACHE IV CABG	Shared cohort
Number of admissions	55,661	2,662	2,290	2,380	1,897	1,435	7,076	44,112
Hospital mortality %	16.4	26.9	55.5	10.9	4.8	3.1	1.5	17.8
Length of ICU stay in days median (25-75%)	1.1 (0.8-3.2)	4.1 (1.7-8.8)	3.1 (1.7-5.9)	1.0 (0.8-1.9)	1.0 (0.9-2.0)	0.9 (0.8-1.1)	0.9 (0.7-1.0)	1.4 (0.8-3.8)
Male %	57.2	62.4	65.1	55.5	82.0	62.2	77.9	57.3
Ventilated %	46.0	65.7	92.5	28.2	34.6	10.5	95.8	42.6
One or more chronic diagnose %	29.3	53.9	25.8	32.9	22.2	29.1	27.8	30.3
Admission type %								
Medical	44.6	100.0	93.6	0.0	0.0	0.0	0.1	48.0
Urgent surgery	16.6	0.0	3.3	18.5	9.4	1.0	4.4	17.9
Elective surgery	38.8	0.0	3.1	81.5	90.6	99.0	95.5	34.1
Age, mean (SD)	63.1 (16.5)	65.6 (14.7)	65.6 (14.5)	72.5 (11.1)	69.6 (8.7)	65.0 (9.4)	66.0 (9.6)	63.2 (16.3)
Survivors	61.8 (16.7)	64.0 (14.8)	63.8 (14.5)	72.0 (11.1)	69.4 (8.7)	64.8 (9.5)	65.8 (9.6)	61.7 (16.5)
Non-survivors	70.0 (13.5)	70.1 (13.3)	67.9 (14.1)	76.9 (9.4)	73.9 (7.0)	69.5 (7.9)	71.0 (9.9)	69.8 (13.5)
SAPS score median (25-75%)	30.0 (20.0-44.0)	42.0 (33.0-53.0)	65.0 (54.0-76.0)	25.0 (20.0-34.0)	23.0 (18.0-31.0)	18.0 (14.0-23.0)	28.0 (23.0-32.0)	31.0 (20.0-45.0)
APACHE II score median (25-75%)	14.0 (10.0-20.0)	20.0 (16.0-26.0)	29.0 (24.0-34.0)	13.0 (10.0-16.0)	13.0 (10.0-16.0)	11.0 (8.0-13.0)	14.0 (11.0-16.0)	15.0 (10.0-21.0)
APACHE IV score median (25-75%)	52.0 (36.0-74.0)	71.0 (56.0-91.0)	116.0 (91.0-135.0)	49.0 (38.0-62.0)	47.0 (38.0-59.0)	39.0 (31.0-48.0)	46.0 (36.0-57.0)	54.0 (37.0-77.0)

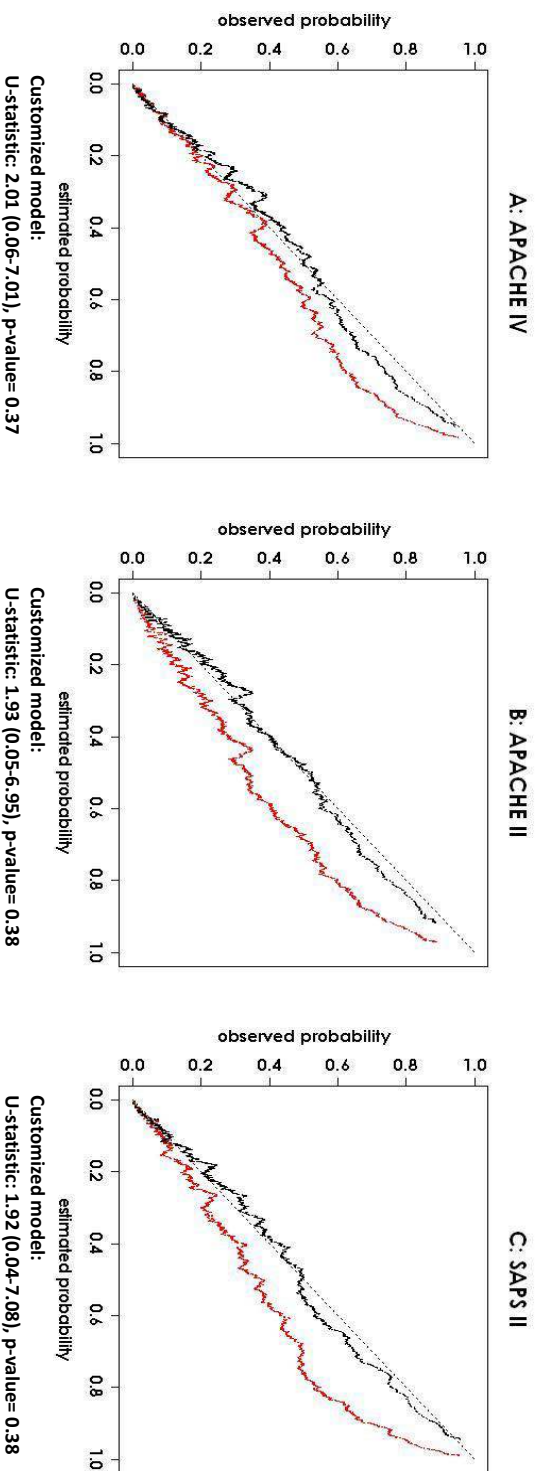


Figure 3.1: Calibration plots of the original and customized (non-CABG) APACHE IV (panel A), original and customized APACHE II (panel B) and original and customized SAPS II model (panel C). The lower line corresponds to the original model and the upper line represents the customized model. The observed probabilities (on the y-axis) are obtained by loess smoothing on the observed outcome values (0 and 1). The unreliability index U and its confidence interval are shown below the graphs.

Table 3.3: Performance of the original and customized APACHE IV model

Model	AUC (CI) ⁺	\hat{C} -statistic (CI)	Brier score (CI)	R ²	SMR (CI)
APACHE IV original non-CABG	0.87 (0.86, 0.87)	822.67 (693.31,955.75)	0.10 (0.10, 0.10)	0.29	0.87 (0.86, 0.88)
Admission type					
Medical	0.85 (0.84, 0.86)	795.62 (601.79,1002.49)	0.13 (0.13, 0.14)	0.28	0.87 (0.85,0.90)
Urgent surgery	0.83 (0.81, 0.84)	156.22 (93.32,232.98)	0.12 (0.12, 0.13)	0.23	0.87 (0.83, 0.91)
Elective surgery	0.82 (0.81, 0.84)	75.95 (40.07,120.65)	0.05 (0.04, 0.05)	0.14	0.85 (0.78, 0.91)
Subgroups*					
Bacterial pneumonia	0.73 (0.70, 0.76)	190.69 (106.34,294.89)	0.18 (0.17, 0.19)	0.09	0.82 (0.75, 0.89)
Cardiac arrest	0.71 (0.69, 0.74)	682.37 (476.19,914.48)	0.25 (0.23, 0.27)	0.00	0.78 (0.75, 0.82)
Colon/rectal cancer	0.80 (0.75, 0.83)	69.42 (36.11,1121.09)	0.09 (0.08, 0.10)	0.10	0.71 (0.60, 0.81)
Abdominal aortic aneurysm	0.77 (0.68, 0.84)	30.87 (11.14,72.77)	0.04 (0.03, 0.05)	0.08	0.83 (0.62, 1.06)
Thoracotomy for lung cancer	0.74 (0.61, 0.85)	41.40 (23.83,62.18)	0.03 (0.02, 0.04)	0.03	0.48 (0.30, 0.68)
APACHE IV original CABG	0.83 (0.78, 0.88)	32.40 (13.80,67.80)	0.01 (0.01, 0.02)	0.08	0.80 (0.65, 0.94)
APACHE IV customized non-CABG		147.73 (144.73,155.05)	0.10 (0.10, 0.10)	0.30	1.00 (0.98, 1.01)
Admission type					
Medical		109.88 (95.74,126.82)	0.13 (0.13, 0.13)	0.30	1.02 (1.00, 1.03)
Urgent surgery		29.47 (27.48,32.45)	0.12 (0.12, 0.12)	0.24	1.02 (1.00, 1.04)
Elective surgery		73.67 (61.07,87.32)	0.05 (0.05, 0.05)	0.15	0.89 (0.87, 0.92)
Subgroups*					
Bacterial pneumonia		48.05 (43.75,53.02)	0.17 (0.17, 0.17)	0.13	0.98 (0.96, 0.99)
Cardiac arrest		129.31 (109.95,150.86)	0.22 (0.22, 0.22)	0.10	0.90 (0.89, 0.91)
Colon/rectal cancer		22.45 (19.48,25.56)	0.08 (0.08, 0.08)	0.13	0.81 (0.79, 0.82)
Abdominal aortic aneurysm		12.49 (11.61,13.49)	0.04 (0.04, 0.04)	0.09	0.85 (0.83, 0.88)
Thoracotomy for lung cancer		29.06 (26.80,31.35)	0.03 (0.03, 0.03)	0.05	0.49 (0.48, 0.51)
APACHE IV customized CABG		14.81 (11.56,24.22)	0.01 (0.01, 0.01)	0.10	1.01 (0.85, 1.22)

AUC: Area under receiver operating characteristics curve; \hat{C} -statistic: Hosmer and Lemeshow goodness-of-fit \hat{C} -statistic (original model tested with 10 df and customized model tested with 8 df); SMR: Standardized Mortality Ratio; CI: 95% Confidence Interval *: subgroups based on APACHE IV reason for admission, +. Note customization does not change the AUC

External validation of APACHE IV model

The performance assessment of the original and customized APACHE IV models is shown in table 3.3 and figure 3.1 panel A. Overall, the original APACHE IV non-CABG model has a good discrimination and accuracy (AUC=0.87, Brier score=0.10, and $R^2=0.29$), but a poor calibration (\hat{C} -statistic=822.67). The overall SMR was 0.87.

There is a considerable variation in performance of the model in the subgroups. For instance, discrimination varied between an AUC of 0.85 for medical admissions and an AUC of 0.71 for patients after cardiac arrest. The Brier score ranged from 0.03 for patients admitted after thoracotomy for lung cancer to 0.25 for patients after cardiac arrest. The R^2 ranged from 0.28 for admitted medical patients to 0.00 for patients after cardiac arrest. Mortality risks were overestimated by the original APACHE IV model, especially for patients admitted after thoracotomy for lung cancer (SMR=0.48). The original APACHE IV CABG model had good performance (AUC=0.83, Brier score=0.01, $R^2=0.08$, \hat{C} -statistic=32.40), and the overall SMR was 0.80. The R^2 is low indicating that the model does not improve much on a non-informative model, because in this case with a hospital mortality of 1.5% there is not much room for improvement.

After customization, the APACHE IV non-CABG model showed statistically significant improvement in the overall sample, but not in all subgroups. The original and the customized model performed poorest in patients after cardiac arrest. Customization did not significantly improve the performance of the APACHE IV CABG model.

Comparison of APACHE IV with APACHE II and SAPS II

After applying the exclusion criteria for the APACHE IV, APACHE II, and SAPS II model, 44,112 records remained in the shared cohort that was used for the comparison of the models. Table 3.1 shows the exclusion criteria and the number of excluded patients. Figure 3.1 illustrates the calibration plots for the overall sample of the original and customized models. It shows that after customization the overall calibration of all models improved (and the unreliability index U shows no statistically significant deviations from actual outcome). However, the three prognostic models still have calibration problems in the high risk groups after customization, with APACHE IV showing the largest discrepancy between observed and expected mortality (Figure 3.1).

Table 3.4 shows the results of the performance comparison between the three customized models. In the overall sample and in some subgroups, the AUC and Brier score of the customized APACHE IV model were statistically significantly better than the customized APACHE II and customized SAPS II model. In the overall sample and in some subgroups, the \hat{C} -statistic of the customized APACHE II and SAPS II models was statistically significantly better than the APACHE IV model. In contrast, in some subgroups the \hat{C} -statistic of the customized APACHE IV model was statistically significantly better than the SAPS II model (e.g. medical admissions, bacterial pneumonia).

Table 3.4: Performance of the customized APACHE IV, APACHE II, and SAPS II model

Model	AUC (CI) ⁺	\hat{C} -statistic (CI)	Brier score (CI)	R ²	SMR (CI)
APACHE IV customized	0.86 (0.86, 0.87)	142.32 (139.93, 148.83)	0.10 (0.10, 0.10)	0.30	1.00 (0.98,1.02)
Admission type					
Medical	0.84 (0.84, 0.85)	85.99 (73.94, 100.70)	0.13 (0.13, 0.13)	0.30	1.02 (1.00,1.03)
Urgent surgery	0.82 (0.80, 0.84)	24.44 (22.27, 27.77)	0.12 (0.12, 0.12)	0.24	1.02 (1.00,1.04)
Elective surgery	0.83 (0.81, 0.85)	51.88 (40.70, 63.62)	0.05 (0.05, 0.05)	0.16	0.89 (0.86, 0.91)
Subgroups*					
Bacterial pneumonia	0.74 (0.70, 0.76)	35.48 (32.15, 39.08)	0.17 (0.17, 0.17)	0.14	0.99 (0.98,1.01)
Cardiac arrest	0.72 (0.69, 0.75)	111.10 (90.88, 131.61)	0.22 (0.22, 0.22)	0.10	0.89 (0.88, 0.90)
Colon/rectal cancer	0.80 (0.76, 0.84)	16.18 (13.70, 19.04)	0.09 (0.09, 0.09)	0.14	0.83 (0.82, 0.85)
Abdominal aortic aneurysm	0.79 (0.71, 0.86)	12.14 (10.81, 13.55)	0.04 (0.04, 0.04)	0.10	0.81 (0.78, 0.84)
Thoracotomy for lung cancer	0.75 (0.62, 0.86)	26.27(32.92, 28.58)	0.03 (0.03, 0.03)	0.06	0.51 (0.49, 0.53)
APACHE II customized	0.84 (0.83, 0.84)	91.21 (89.12, 96.65)	0.11 (0.11, 0.11)	0.25	1.00 (0.98, 1.02)
Admission type					
Medical	0.81 (0.80, 0.82)	86.88 (68.54, 107.89)	0.14 (0.14, 0.14)	0.24	1.04 (1.02, 1.06)
Urgent surgery	0.78 (0.77, 0.80)	30.78 (28.35, 34.21)	0.13 (0.13, 0.13)	0.18	0.97 (0.95, 0.98)
Elective surgery	0.80 (0.78, 0.82)	66.71 (53.73, 80.18)	0.05 (0.05, 0.05)	0.13	0.87 (0.85, 0.90)
Subgroups*					
Bacterial pneumonia	0.69 (0.66, 0.72)	43.75 (38.24, 49.63)	0.18 (0.18, 0.18)	0.08	0.98 (0.96, 0.99)
Cardiac arrest	0.69 (0.66, 0.73)	39.70 (36.11, 44.19)	0.22 (0.22, 0.22)	0.10	1.01 (0.99, 1.03)
Colon/rectal cancer	0.76 (0.72, 0.81)	16.88 (15.76, 18.06)	0.09 (0.09, 0.09)	0.12	0.96 (0.93, 0.98)
Abdominal aortic aneurysm	0.76 (0.67, 0.83)	7.74 (6.50, 9.11)	0.04 (0.04, 0.04)	0.09	0.81 (0.78, 0.84)
Thoracotomy for lung cancer	0.71 (0.58, 0.83)	12.34 (10.50, 14.23)	0.03 (0.03, 0.03)	0.04	0.62 (0.59, 0.64)
SAPS II customized	0.85 (0.84, 0.85)	112.70 (110.18, 119.11)	0.11 (0.11, 0.11)	0.29	1.00 (0.98, 1.02)
Admission type					
Medical	0.82 (0.81, 0.83)	117.07 (100.62, 135.60)	0.14 (0.14, 0.14)	0.26	1.04 (1.02, 1.05)
Urgent surgery	0.82 (0.81, 0.84)	58.38 (53.09, 64.46)	0.12 (0.12, 0.12)	0.24	0.94 (0.92, 0.95)
Elective surgery	0.81 (0.79, 0.83)	30.68 (23.01, 39.64)	0.05 (0.05, 0.05)	0.14	0.93 (0.90, 0.96)
Subgroups*					
Bacterial pneumonia	0.70 (0.67, 0.74)	74.22 (67.01, 82.08)	0.18 (0.18, 0.18)	0.09	1.08 (1.06,1.09)
Cardiac arrest	0.73 (0.70, 0.76)	56.96 (53.25, 60.86)	0.21(0.21, 0.21)	0.14	1.03 (1.01, 1.04)
Colon/rectal cancer	0.77 (0.73, 0.82)	13.12 (11.42, 15.07)	0.09 (0.09, 0.09)	0.16	1.09 (1.06, 1.11)
Abdominal aortic aneurysm	0.80 (0.71, 0.87)	30.95 (28.11, 34.08)	0.04 (0.04, 0.04)	0.07	0.58 (0.56, 0.60)
Thoracotomy for lung cancer	0.74 (0.62, 0.84)	8.65 (7.67, 9.77)	0.03 (0.03, 0.03)	0.08	0.72 (0.69, 0.75)

AUC: Area under receiver operating characteristics curve; \hat{C} -statistic: Hosmer and Lemeshow goodness-of-fit \hat{C} -statistic with 8 df; SMR: Standardized Mortality Ratio; CI: 95% Confidence Interval *: subgroups based on APACHE IV reason for admission +: Note customization does not

Discussion

This external validation study shows that the original APACHE IV model has good discrimination and accuracy in a large Dutch ICU sample, but the calibration was poor. We believe the need for customization of the APACHE IV is required due to both the length of time since the development of the model and due to changes in country setting. In our study it is difficult to determine to which degree each factor separately determines the degradation of model performance without customization. The overall AUC and Brier score of the customized APACHE IV model was statistically significantly better than the customized APACHE II and SAPS II models while the calibration of the customized APACHE IV was inferior to the customized APACHE II and SAPS II model. However, differences were small and probably not very relevant in clinical practice. In the original APACHE IV study, the authors suggested that older models probably should not be used for current performance measurement (4). Our findings support this suggestion regarding the original APACHE II and SAPS II models, though this does not apply for the customized APACHE II and SAPS II models. As all three prognostic models need customization before being applied in a new setting and the performance of the customized models is more or less similar.

The performance of the APACHE IV model in the Dutch ICU population varies in different diagnostic subgroups. This is in agreement with previous studies in which it is stated that the prognostic models may under or over predict the mortality in specific populations, which might not have been well represented in the original case-mix used to develop the model (15,25-27). The SMR of the three prognostic models was poorest in the subgroups thoracotomy for lung cancer and abdominal aortic aneurysm. It suggests that the outcome of these patients groups may be influenced importantly by variables not included in the three models, such as size and location of the aneurysm. The percentage of admissions with thoracotomy for lung cancer and abdominal aortic aneurysm at an ICU ranged from 0.0-5.0% resp. 0.0-5.9%. As these percentages are small it will probably have minimal influence on the overall SMR of an ICU. The performance of the customized APACHE IV was poorest in the group of patients after cardiac arrest (AUC=0.72, Brier score=0.22, $R^2=0.10$, SMR=0.90), whereas this subgroup had an excellent performance in the original APACHE IV study (SMR=1.00) (4). The differences in Brier scores between subgroups may be mostly explained by the different crude mortality rates within the various subgroups. Brier scores tend to be highest when mortality risk is 0.5 and decrease when this risk is either close to 0 or 1 (28). Unexpectedly, the overall calibration of APACHE IV measured by the \hat{C} -statistic appeared to be statistically significantly worse than that of the APACHE II and SAPS II seemingly due to under-estimation of risk for high-risk patients (figure 3.1). It is important to note however that the probability deciles associated with the different models may differ significantly, which render the comparison of the \hat{C} -statistics as

nontrivial. In contrast, the Brier score and the R^2 measure of the APACHE IV model appeared to be better than that of the APACHE II and SAPS II model.

The conflicting results of performance measured using different statistics can be explained by their characteristics. In our study the Brier score and \hat{C} -statistic are used to assess the accuracy and calibration of the models. The Brier score has aspects of both discrimination and calibration. It operates at the patient level as it calculates the mean of the differences between the observed and predicted mortality of each patient. The \hat{C} -statistic provides information on the calibration at a group's level as it compares the observed and predicted outcome for each risk decile.

To our knowledge, this is the first study validating the APACHE IV model on a non U.S. ICU population. In data from Californian ICUs the overall discrimination of the APACHE IV model (AUC=0.89) was, as in our study, significantly better than the discrimination of the customized SAPS II model (AUC=0.87) (5).

The older APACHE II and SAPS II models have been validated more often and also in European populations. In a study by Harrison et al. the APACHE II and SAPS II models were customized and validated on ICU admissions in the UK (11). Discrimination of those models was comparable to our results but the Brier score was lower (APACHE II: AUC=0.83 and Brier score=0.15, SAPS II: AUC=0.84 and Brier score=0.15). Disparities in case-mix and thereby mortality rates between the UK and the Netherlands are the most likely explanations for these differences. External validation of the APACHE II and SAPS II in previous studies also revealed good discrimination but poor calibration (10,13,16,29-31). In agreement with these studies, we recommend that all risk prediction models should be validated before being used to provide risk-adjusted outcomes within a new country setting. Furthermore, periodical customization is necessary to maintain the validity of the models (10,11).

A limitation of the study is that although the included hospitals spread out over the Netherlands, included both teaching and non-teaching hospitals, and had the same proportion of academic centers, we cannot exclude volunteer bias as participation to the NICE registry is not mandatory. In our study the performance of the three models was assessed in a cohort respecting their combined inclusion criteria. This might influence model performance as some patients groups might be included for one model but excluded for another model (32). The three models were first level customized as the sample size of some diagnoses was too small to perform second level customization which involves re-estimating all coefficients of the model. It has been reported that first level customization is effective in repairing structural errors (14,16). Another limitation in our study is that we performed analysis on many subgroups and we did not control for multiple comparisons, therefore the results of these analyses should be regarded as exploratory.

The overall performance of the three prognostic models indicates that they have comparable capabilities for benchmark purposes. However, for each of these models the SMRs of the various subgroups vary widely. This can lead to different conclusions about the quality of care in the ICUs when they clearly differ in their subgroup make up. For example an ICU that admits a higher than average proportion of elective surgical patients will receive undue credits as the SMR for these patients is low. Therefore higher level customization could be considered before any of the models can be considered for benchmark purposes. Even then, the SMRs should only be used to signal and identify areas for quality improvement and not as judgment for the quality of care by external health service funders (33). Some points of attention need to be considered when choosing a model. For the APACHE IV model, 11 additional variables are required to be collected, which could be time consuming especially for hospitals without an electronic interface transmitting the data. However, the increasing use of electronic interfaces will minimize the difference in data collection burden between the APACHE II and APACHE IV models. A main advantage of the APACHE IV model is that it incorporates a large number of diagnoses, which facilitates outcome analyses in specific (more homogeneous) subgroups. The APACHE IV CABG model has a good performance in the Dutch ICU population and could be used complementary to each of the three models as the APACHE II and SAPS II model exclude patients with a CABG. Comparison between cardiac surgery specific models such as the EuroSCORE and the APACHE IV CABG model should be performed.

Acknowledgments

We express our gratitude to Dr. J.E. Zimmerman, MD and Dr. A.A. Kramer, PhD for providing feedback and valuable suggestions. We also express our gratitude to the head software engineer of NICE, Mr. E. van der Zwan, MMI, for his helpful support during the data analysis.

References

1. Knaus WA, Draper EA, Wagner DP, et al: APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13:818-829
2. Knaus WA, Wagner DP, Draper EA, et al: The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; 100:1619-1636
3. Le Gall JR, Lemeshow S, Saulnier F: A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270:2957-2963
4. Zimmerman JE, Kramer AA, McNair DS, et al: Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34:1297-1310
5. Kuzniewicz MW, Vasilevskis EE, Lane R, et al: Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. *Chest* 2008; 133:1319-1327
6. Bleeker SE, Moll HA, Steyerberg EW, et al: External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003; 56:826-832
7. Terrin N, Schmid CH, Griffith JL, et al: External validity of predictive models: a comparison of logistic regression, classification trees, and neural networks. *J Clin Epidemiol* 2003; 56:721-729
8. Kramer AA, Zimmerman JE: Predicting outcomes for cardiac surgery patients after intensive care unit admission. *Semin Cardiothorac Vasc Anesth* 2008; 12:175-183
9. Arts D, de Keizer NF, Scheffer GJ, et al: Quality of data collected for severity of illness scores in the Dutch National Intensive Care Evaluation (NICE) registry. *Intensive Care Med* 2002; 28:656-659
10. Beck DH, Smith GB, Pappachan JV, et al: External validation of the SAPS II, APACHE II and APACHE III prognostic models in South England: a multicentre study. *Intensive Care Med* 2003; 29:249-256
11. Harrison DA, Brady AR, Parry GJ, et al: Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med* 2006; 34:1378-1388
12. Le Gall Jr, Neumann A, Hemery F, et al: Mortality prediction using SAPS II: an update for French intensive care units. *Crit Care* 2005; 9:R645-R652
13. Moreno R, Miranda DR, Fidler V, et al: Evaluation of two outcome prediction models on an independent database. *Crit Care Med* 1998; 26:50-61
14. Bakhshi-Raiez F, Peek N, Bosman RJ, et al: The impact of different prognostic models and their customization on institutional comparison of intensive care units*. *Crit Care Med* 2007; 35:2553-2560
15. Zhu BP, Lemeshow S, Hosmer DW, et al: Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. *Crit Care Med* 1996; 24:57-63
16. Moreno R, Apolone G: Impact of different customization strategies in the performance of a general severity score. *Crit Care Med* 1997; 25:2001-2008
17. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29-36
18. Hilden J, Habbema JD, Bjerregaard B: The measurement of performance in probabilistic diagnosis. III. Methods based on continuous functions of the diagnostic probabilities. *Methods Inf Med* 1978; 17:238-246
19. Mittlbock M, Schemper M: Explained variation for logistic regression. *Stat Med* 1996; 15:1987-1997
20. Hosmer DW, Hosmer T, Le Cessie S, et al: A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997; 16:965-980
21. Kramer AA, Zimmerman JE: Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med* 2007; 35:2052-2056
22. Cox DR: Two further applications of a model for binary regression. *Biometrika* 1985; 45:562-565
23. Efron B: Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation . Available online at: <http://www.jstor.org/stable/2288636>.
24. Bosman RJ, Oudemane van Straaten HM, Zandstra DF: The use of intensive care information systems alters outcome prediction. *Intensive Care Med* 1998; 24:953-958

25. Alberti C, Brun-Buisson C, Goodman SV, et al: Influence of systemic inflammatory response syndrome and sepsis on outcome of critically ill infected patients. *Am J Respir Crit Care Med* 2003; 168:77-84
26. Nierman DM, Schechter CB, Cannon LM, et al: Outcome prediction model for very elderly critically ill patients. *Crit Care Med* 2001; 29:1853-1859
27. Staudinger T, Stoiser B, Mullner M, et al: Outcome and prognostic factors in critically ill cancer patients admitted to the intensive care unit. *Crit Care Med* 2000; 28:1322-1328
28. Steyerberg EW: *Clinical Prediction Models A Practical Approach to Development, Validation, and Updating*. New York, Springer, 2009
29. Livingston BM, MacKirdy FN, Howie JC, et al: Assessment of the performance of five intensive care scoring models within a large Scottish database. *Crit Care Med* 2000; 28:1820-1827
30. Sakr Y, Krauss C, Amaral AC, et al: Comparison of the performance of SAPS II, SAPS 3, APACHE II, and their customized prognostic models in a surgical intensive care unit. *Br J Anaesth* 2008; 101:798-803
31. Ledoux D, Canivet JL, Preiser JC, et al: SAPS 3 admission score: an external validation in a general intensive care population. *Intensive Care Med* 2008; 34:1873-1877
32. Wunsch H, Brady AR, Rowan K: Impact of exclusion criteria on case mix, outcome, and length of stay for the severity of disease scoring methods in common use in critical care. *J Crit Care* 2004; 19:67-74
33. Lilford R, Pronovost P: Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 2010; 340:c2016

Chapter 4

Hospital mortality is associated with
ICU admission time

Hans A. J. M. Kuijsten, Sylvia Brinkman, Iwan A. Meynaar,
Peter E. Spronk, Johan I. van der Spoel, Rob J. Bosman,
Nicolette F. de Keizer, Ameen Abu-Hanna, Dylan W. de Lange

Intensive Care Medicine 2010; 36(10): 1765-1771



Abstract

Introduction: Previous studies have shown that patients admitted to the intensive care unit (ICU) after “office hours” are more likely to die. However these results have been challenged by numerous other studies. We therefore analysed this possible relationship between ICU admission time and in-hospital mortality in the Netherlands.

Methods: This article relates the time of admission to hospital mortality of all patients that were included in the Dutch national ICU registry (National Intensive Care Evaluation, NICE) from 2002-2008. We defined “office hours” as 08.00-22.00 during weekdays and 09.00-18.00 during weekend days. The weekend was defined as Saturday from 00.00 hours until Sunday 24.00 hours. We corrected hospital mortality for illness severity at admission using the Acute Physiology and Chronic Health Evaluation (APACHE) II score, reason for admission, admission type, age, and gender.

Results: A total of 149,894 patients were included in this analysis. The relative risk (RR) for mortality outside office hours (i.e. “off hours”) was 1.059 (1.031-1.088). Mortality varied with time but was consistently higher than expected during off hours and lower during office hours. There was no significant difference in mortality between different weekdays (Monday to Thursday) but mortality increased slightly on Friday (RR 1.046; 1.001-1.092). During the weekend the RR was 1.103 (1.071-1.136) in comparison to the rest of the week.

Conclusion: Hospital mortality in the Netherlands appears to be increased outside office hours and during the weekends, even if corrected for illness severity at admission. However, incomplete adjustment for certain confounders might still play an important role. Further research is needed to fully explain this difference.

Introduction

Ideally, care for critically ill patients is optimal, 24 hour per day. Unfortunately, this is not the case. During so-called off hours, staffing is often reduced, and diagnostic and therapeutic procedures might take longer or are postponed until office hours. Because treatment in the first hours after admission to the intensive care unit (ICU) is related to outcome (1,2), admission outside office hours might be associated with increased mortality (3,4). However, this increased mortality outside office hours has been questioned by several other studies (5–10), suggesting that the association was confounded by differences in the definition of office hours, differences in study population, study size, healthcare organization or insufficient correction for case-mix and illness severity. As a consequence of these differences in methodology and organizational differences between countries, it remains unclear whether there is a relationship between admission outside office hours and increased hospital mortality. We hypothesized that mortality during off hours is higher than during office hours and that previous studies that did not find this result were probably underpowered or defined office hours too broadly. We therefore analyzed hospital mortality in relation to admission time in a very large database with stringent case-mix correction. We also analyzed whether sample size could have influenced the results of previous studies. This approach enables us to analyze the relationship between admission time and hospital mortality in the most robust way thus far.

Methods

Patient data

Since 1996 the National Intensive Care Evaluation (NICE) foundation has collected data on admissions to ICUs in The Netherlands. NICE started with 6 ICUs in 1996, and in 2008 more than 70 ICUs were participating in this registry, accounting for 80% of all Dutch ICUs, and approximately 50,000 admissions were included in this registry in 2008. Details about inclusions and exclusions in the Dutch registry have been published previously (11). In short, the participating ICUs are mixed medical-surgical units located in university hospitals (n=7), teaching hospitals (n=25) or non-teaching hospitals (n=38). A data set of about 100 items is collected for each individual patient. Based upon this data several prediction models can be calculated, such as APACHE II (12), which are used to correct the crude hospital mortality for illness severity at admission. Data collection takes place in a prospective and standardized manner according to strict and uniform definitions and is subject to stringent data quality checks. Additionally, site visits are performed to ensure the quality of the collected data. This has been shown to ensure high quality of data (13). The data is aggregated at a central point (NICE Foundation, Amsterdam, The Netherlands).

All patient identifying information, such as name and patient identification number, is encrypted. Data are analyzed and stored in an anonymous way and are not traceable to any patient, and therefore informed consent was not needed. All patient records in this database between 2002 and 2008 were included in our analysis. According to the original APACHE II exclusion criteria, cardiopulmonary surgical patients, re-admissions to the ICU, patients who were discharged or died within 8 hours after admission and burn patients were excluded.

Defining office hours

Office hours were defined by the presence of a fully qualified intensivist available for patient care. In The Netherlands, three levels of intensive care are identified. The minimal demands for the three levels of intensive care are available in the ESM. Despite organizational differences, commonly intensivists are available from 08:00 to 22:00 hours. Therefore, we defined 08:00–22:00 hours during weekdays and 09:00–18:00 hours during the weekends as office hours. A weekday was defined as a day from 00:00 hours until 24:00 hours, and the weekend was defined as from Saturday 00:00 hours until Sunday 24:00 hours.

Data analysis

Hospital mortality by day of week, in or outside weekend and during office hours versus off hours were examined first using univariate analysis and then with multiple logistic regression adjusted for case-mix. Adjustment for case-mix was undertaken by using the Dutch APACHE II model. To date, there have been no studies that have shown that prognostic models are stable over time, in a new setting, and with different case-mixes (14,15). As a result of medical progress and advancement of science, it is expected that the model's performance will decline over time (16). To account for this decline in performance we recalibrated the APACHE II before using it for case-mix correction. This involved fitting a new logistic regression equation with in-hospital mortality as dependent variable and the original covariates (APACHE II score, APACHE II reason for admission and admission type) as independent variables. Accordingly, customization does not change the influence of individual covariates included in the model but modifies their joint influence on the observed mortality in the external dataset (17–19). Three different recalibrated logistic regression models were used to analyze the association between in-hospital mortality and off hours versus office hours (model 1), weekdays versus weekend (model 2) and day of week (model 3). For better risk adjustment, patients admitted during office hours versus off hours, weekend versus weekday, and day of the week were matched based on a calculated propensity score. This score expresses the probability that a patient falls in the office hours or off hours, weekday or weekend, or day of week group, and its log odds was added as a covariate in the recalibrated APACHE II model for adjustment.

When the incidence of an outcome of interest is common in the study population (10%), the adjusted odds ratio derived from the logistic regression could under- or overestimate the risk ratio. Therefore, the associations between admission time and mortality are reported as relative risk and their 95% confidence interval (95% CI). The relative risks were approximated from the adjusted odds ratio derived from the three logistic regression models (18) and considered statistically significant if the confidence interval did not contain zero. Categorizing a continuous variable (admission time) into office hours and off hours may conceal the behavior of the variables over time. Therefore, the relationship between admission time and hospital mortality was also graphically inspected using locally weighted scatterplot smoothing (LOWESS). In simple terms, at each time point, e.g. 15:00 hours, simple models are fitted to the mortality data in the “vicinity” of the point, e.g. from 13:00 to 17:00 hours, with points closer to 15:00 hours having more influence than those farther away. This procedure is performed for all data points. Instead of the scatter plot of 0 and 1 values of mortality over time we now have a smoothed curve that reflects the underlying structure of the data. To investigate whether smaller sample size could explain the results of previous articles on this subject that showed no relation between mortality and time of admission, we used 300 random sub-samples from our research dataset with sample size equal to that in two previous articles: 56,250 and 6,725 patients (6,10). Data were analyzed using the R2.6.2 statistical environment and SPSS 16.0 (Chicago, IL, USA).

Results

From January 2002 to January 2009, a total of 149,894 patients eligible according to the APACHE II inclusion criteria were included in the analyses. Baseline characteristics of these patients are shown in table 4.1.

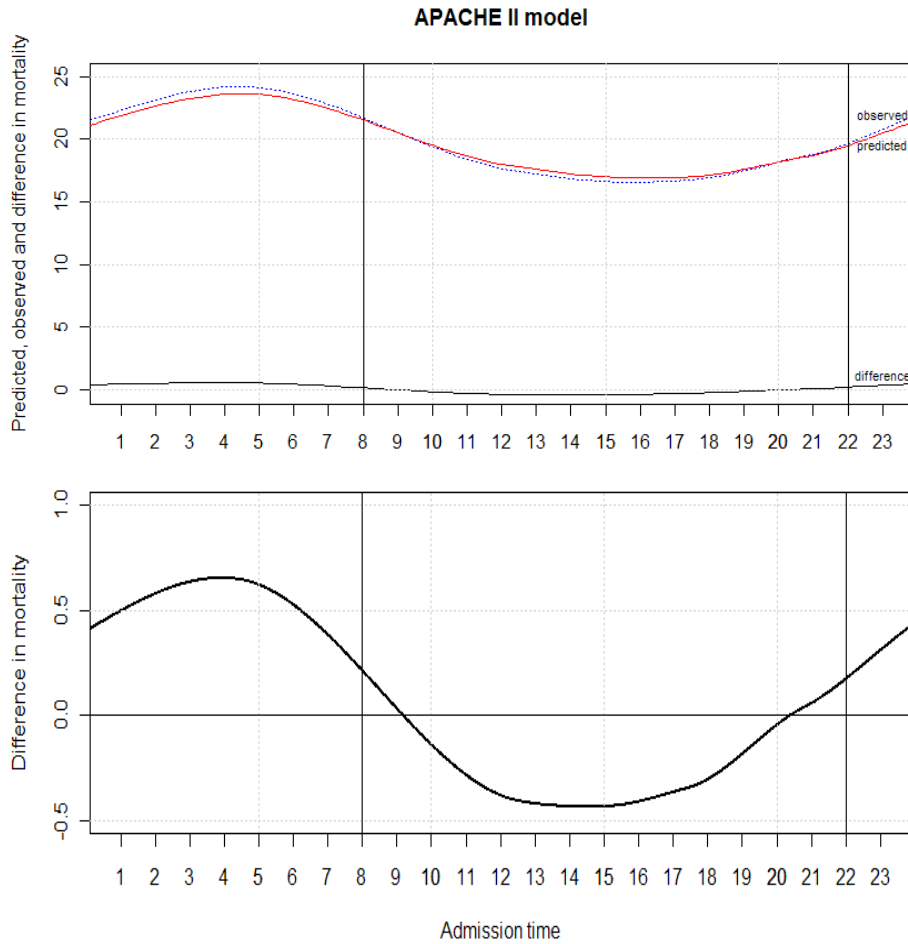
Table 4.1 Baseline characteristics of the study population (N = 149,894)

Patient characteristic	N (%)
Hospital mortality	28,135 (19)
ICU mortality	18,051 (12)
Male	86,977 (58)
Age (years), mean (SD)	61.9 (17)
Ventilated in first 24 h of admission	71,294 (48)
Median APACHE II score (25–75%)	15.0 (10-21)
Type of admission	
Non-surgical/medical	71,181 (47)
Emergency surgery	26,298 (18)
Planned surgery	52,415 (35)

Table 4.2 shows the distribution of patients according to admission day, weekend versus weekdays and off hours versus office hours. This table shows that the predicted and observed mortality of patients during the weekends and off hours is higher than during the week and office hours, respectively. Figure 4.1 shows the actual observed mortality and the predicted mortality probability of the recalibrated APACHE II model in relation to admission time. The lowest line shows the difference between the observed and predicted mortality proportion. This figure shows qualitatively that the APACHE II model predictions closely follow the observed mortality proportion over all admission times. The difference between the observed and predicted mortality is minimal (the lowest line is almost horizontal) except for the hours close to midnight. The lowest mortality is seen during office hours, especially during the afternoon. Then, mortality slowly increases again, with a peak around 05:00–06:00 hours in the morning. The figure shows that the observed proportion minus the predicted probability during office hours was lower than during off hours (17.5% versus 22.7%). In the whole off hours region the difference between observed and predicted mortality was positive (meaning that there was more mortality than predicted), which was not the case in the office hours region. This observation provides face validity for treating office hours versus off hours as a dichotomous variable in the logistic regression model. The graph also suggests that the effect of off hours on mortality would have been even stronger if off hours had been defined as, say, 20:00–08:00 hours.

Table 4.2: Characteristics of the patients admitted per time variable (N = 149,894)

Time variable	Admissions N	Male (%)	Age Mean (SD)	APACHE II score Median (25-75%)	Expected mortality Median (25-75%)	Expected mortality Mean (SD)	Observed Hospital mortality N (%)	SMR (95% CI)
Sunday	13,518	57	59.9 (18.7)	18 (12-24)	26.6 (11.2-51.5)	33.3 (26.9)	3,517 (26.0)	0.78 (0.76-0.81)
Monday	24,499	59	61.9 (16.4)	15 (10-21)	15.2 (6.2-35.9)	24.6 (24.2)	4,336 (17.7)	0.72 (0.70-0.74)
Tuesday	25,394	58	62.6 (16.0)	14 (10-20)	14.6 (6.2-34.8)	24.1 (24.0)	4,352 (17.1)	0.71 (0.69-0.73)
Wednesday	25,627	58	62.5 (16.1)	14 (10-20)	14.3 (6.1-34.6)	23.8 (23.7)	4,333 (16.9)	0.71 (0.69-0.73)
Thursday	24,222	58	62.2 (16.4)	15 (10-20)	14.8 (6.3-35.3)	24.2 (23.7)	4,066 (16.8)	0.69 (0.67-0.72)
Friday	23,259	58	62.4 (16.5)	15 (10-21)	15.9 (6.7-36.9)	25.1 (24.1)	4,255 (18.3)	0.73 (0.71-0.75)
Saturday	13,375	57	60.3 (18.4)	17 (12-24)	25.9 (11.2-49.5)	32.7 (26.0)	3,276 (24.5)	0.75 (0.72-0.78)
Weekend	26,893	57	60.1 (18.6)	18 (12-24)	26.2 (11.2-50.1)	33.0 (26.1)	6,793 (25.3)	0.77 (0.75-0.78)
Week	123,001	58	62.3 (16.3)	15 (10-21)	14.8 (6.3-35.5)	24.4 (23.9)	21,342 (17.4)	0.71 (0.70-0.72)
Office hours	107,762	58	62.8 (15.9)	14 (10-20)	14.3 (6.2-34.6)	23.9 (23.8)	18,452 (17.1)	0.72 (0.71-0.73)
Off hours	42,132	56	59.6 (18.5)	17 (11-23)	23.6 (9.6-47.1)	30.9 (25.8)	9,683 (23.0)	0.74 (0.73-0.76)



4

Figure. 4.1: Relationship between observed and predicted mortality (in percentages) in relationship to admission time. The predicted mortality is based upon the APACHE II model and observed mortality. The APACHE II model has been recalibrated to better fit the Dutch ICU population. The upper lines are overlapping, which shows that the model correctly predicts mortality in the general Dutch ICU population. The difference between the upper lines is minimal (almost horizontal lower line). To illustrate the pattern of the difference between the lines, the lower figure blows up this difference. Both the model as well as the observed mortality change with admission time. The lowest mortality is seen during office hours (08:00–22:00 hours), and the highest mortality is seen during off hours (22:00– 8:00 hours)

Table 4.3: Relative risk for hospital mortality during “off hours”, the weekend and during the week

Time variable	Relative risk	95% Confidence interval
<i>Model 1</i>		
Office hours	1.0	-
Off hours	1.059	1.031-1.088*
<i>Model 2</i>		
Week	1.0	-
Weekend	1.103	1.071-1.136*
<i>Model 3</i>		
Monday	1.006	0.962-1.051
Tuesday	0.999	0.956-1.044
Wednesday	1.0	-
Thursday	0.977	0.934-1.022
Friday	1.046	1.001-1.092*
Saturday	1.062	1.012-1.122*
Sunday	1.113	1.062-1.166*

* Significant difference

Table 4.3 shows the relative risk and 95% confidence intervals of the three different models, investigating the days of weeks, weekend versus weekdays and off hours versus office hours after case-mix correction. Admissions during off hours have a significantly higher mortality risk [RR 1.059 (1.031–1.088)] than admissions during office hours. The same holds for admissions during the weekend [RR 1.103 (1.071–1.136)] in comparison with weekdays, although the mortality risk is already increased for patients admitted on Friday [RR 1.046 (1.001–1.092)]. We also analyzed whether smaller sample size would have influenced our results. The 300 random sub-samples from our research dataset based on the largest sample size among studies reporting negative findings (N=56,250) (6) showed a significant difference between office hours versus off hours in 85% of our samples. However, the analyses on 300 random sub-samples based on the smallest sample size among studies reporting negative findings (N=6,725) (10) only showed significantly higher mortality during off hours in 15% of samples, and significantly higher mortality during office hours in 1% of samples.

Discussion

This study showed an increase in the risk of hospital mortality for patients admitted during off hours compared with patients admitted during office hours (RR 1.059), and an increase of hospital mortality risk for patients admitted during the weekend compared with patients admitted during the week (RR 1.103). Several analyses that describe the difference in mortality between patients admitted during office hours and those admitted to the ICU during off hours have been published (5–10,20–23). Unfortunately, all of these studies defined office hours differently. We defined working hours as those hours when a qualified intensivist was available for direct patient care. For most ICUs in The Netherlands this is from 08:00 to 22:00 hours. This definition is in accordance with another recent Dutch publication (10). However, our results contradict some of these more recent publications on this subject. For example, Meynaar et al. (10) analyzed the difference in mortality of 6,725 patients admitted during office hours and outside office hours. They could not detect a difference in mortality after correction for case-mix and illness severity. As their sample size of only 6,725 patients was much smaller than our research sample, we used 300 random sub-samples from our research dataset to replicate their sample size. We found a statistically significant difference in only 49 of 300 samples (15%). Although the sub-samples are overlapping, this suggests that their sample size might have been too small to detect these differences in mortality. However, they analyzed only three ICUs located in teaching hospitals, which are possibly much more homogenous in performance over time compared with our mixed set of 70 participating ICUs. On the other hand, our results also contradict the largest analysis thus far (6). In a UK database of 56,250 ICU patients, Wunsch et al. found increased mortality for the weekends (Friday–Sunday) and during the evening and night. However, after correction for case-mix, this difference disappeared. They concluded that there was no difference in outcome between office hours and off hours. However, they defined office hours differently, choosing three shifts (08:00–18:00, 18:00–24:00 and 24:00–08:00 hours) that best reflect ICU care in the UK. An analysis on 300 random sub-samples from our dataset sample with their sample size ($n = 56,250$) showed a significant difference in mortality between office hours and off hours in 256 of the 300 samples. Although the sub-samples are overlapping, this suggests that the difference in conclusion is not likely based upon limited power in the study by Wunsch et al. We corrected in a similar way for potential confounders, and therefore our analyses are comparable. This suggests that the increased mortality in our study might be based on differences in staffing or logistics between the UK and The Netherlands. Of course, ICU performance is influenced by an intricate interplay of various factors. Besides the ICU organization during off hours there is intensive interaction with other medical disciplines, and changes in their quality of care during off hours might influence ICU outcome as well, which is not reflected by illness severity at admission. Such unknown

confounders might be stronger in smaller hospitals, which often have less staff to fill the roster and/or have less sophisticated diagnostics than larger hospitals. Furthermore, the performance of health care workers (physicians and nurses) varies during the day. Although speculative, the detrimental effect of the circadian biorhythm on human performance during the night shift and especially at the end of the night shift is a known factor (24). We also found the highest (predicted and observed) mortality at the end of the night shift (05:00–06:00 hours), when both health care workers and patients perform at their worst.

This study has several limitations. Although the data were collected in prospective fashion during the first 24 h of admission to the ICU and the outcome (discharged alive or dead) was not influenced by subjective assessment, this remains a retrospective analysis. Therefore, true cause-and-effect relationships cannot be ascertained, and unmeasured confounders still might play a role. For example, differences of care beyond the ICU might be confounders in the association between admission timing and mortality. However, all surviving patients are discharged to the same wards, and it is reasonable to assume that they all experience the same quality of care on the wards.

Although APACHE II was used to correct for illness severity and patients were matched based upon a propensity score (age, gender, APACHE score, admission type, reason for admission), this still does not fully exclude the influence of case-mix differences. Surgical patients admitted in the middle of the night are different from patients admitted during office hours. For example, does waiting for emergency surgery during off hours result in higher acute physiology score and subsequently to higher APACHE II score (so-called lead-time bias)? Other, possibly very important, confounders might be differences in organizational aspects. Unfortunately, the data in this study could not be corrected for these organizational aspects, and we assumed office hours to be 08:00–22:00. Figure 1 shows that there is an increase in the difference between observed and predicted mortality from 16:00 hours onwards. The effect of off hours on mortality would have been stronger if off hours had been defined as, say, 20:00–08:00 hours. As of 2008, all ICUs have collected data on quality parameters, such as nurse-to-patient ratio, physician-to-patient ratio, availability of ICU beds, etc. Such information might explain the differences between ICUs and explain why some ICUs apparently have equal performance during the entire day while others perform worse during off hours. However, a strong feature of this analysis is its size and its power to detect these differences. This is one of the largest analyses of admission timing and survival. Additionally, this analysis is performed in a large sample of the Dutch ICUs (up to 80% of the Dutch ICUs in 2008), and therefore these results can be extrapolated to represent the level of critical care in The Netherlands. Previous studies were often performed in a smaller subset of ICUs and might represent the better-performing ICUs. Such decreased external validity might explain a lack of difference between office hours and off hours in these studies. We conclude that

admission timing is associated with differences in outcome, even when mortality is corrected for illness severity by means of recalibrated APACHE II score. Patients admitted during the night (22:00–08:00 hours) or weekend days have a decreased chance of survival in comparison with patients admitted during office hours. However, the cause of this association needs further analysis that corrects for more potential confounders. This investigation has been started in The Netherlands with the registration of various quality of care variables.

References

1. Pearse RM, Rhodes A, Grounds RM (2004) Clinical review: how to optimize management of high-risk surgical patients. *Crit Care* 8:503-507
2. Rivers E, Nguyen B, Havstad S et al (2001) Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 345:1368-1377
3. Bell CM, Redelmeier DA (2001) Mortality among patients admitted to hospitals on weekends as compared with weekdays. *N Engl J Med* 345:663-668
4. Ensminger SA, Morales JJ, Peters SG et al (2004) The hospital mortality of patients admitted to the ICU on weekends. *Chest* 126:1292-1298
5. Barnett MJ, Kaboli PJ, Sirio CA et al (2002) Day of the week of intensive care admission and patient outcomes: a multisite regional evaluation. *Med Care* 40:530-539
6. Wunsch H, Mapstone J, Brady T et al (2004) Hospital mortality associated with day and time of admission to intensive care units. *Intensive Care Med* 30:895-901
7. Arabi Y, Alshimemeri A, Taher S (2006) Weekend and weeknight admissions have the same outcome of weekday admissions to an intensive care unit with onsite intensivists coverage. *Crit Care Med* 34:605-611
8. Luyt CE, Combes A, Aegerter P et al (2007) Mortality among patients admitted to intensive care units during weekday day shifts compared with "off" hours. *Crit Care Med* 35:3-11
9. Sheu CC, Tsai JR, Hung JY et al (2007) Admission time and outcomes of patients in a medical intensive care unit. *Kaohsiung J Med Sci* 23:395-404
10. Meynaar IA, van der Spoel JI, Rommes JH et al (2009) Off hour admission to an intensivist-led ICU is not associated with increased mortality. *Crit Care*. Doi:10.1186/cc7904
11. de Lange DW, Dusseljee J, Brinkman S et al (2009) Severity of illness and outcome in ICU patients in the Netherlands: results from the NICE registry 2006-2007. *Neth J Crit Care* 13:16-22
12. Knaus WA, Draper EA, Wagner DP et al (1985) APACHE II: a severity of disease classification system. *Crit Care Med* 13:818-829
13. Arts DG, Bosman RJ, de Jonge E et al (2003) Training in data definitions improves quality of intensive care data. *Crit Care* 7:179-184
14. Harrison DA, Brady AR, Parry GJ et al (2006) Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med*, 34:1378-1388
15. Beck DH, Smith GB, Pappachan JV et al (2003) External validation of the SAPS II, APACHE II and APACHE III prognostic models in South England: a multicentre study. *Intensive Care Med* 29:249-256
16. Teres D and Lemeshow S (1999) When to customize a severity model. *Intensive Care Med* 25:140-142
17. Bakhshi-Raiez F, Peek N, Bosman RJ et al (2007) The impact of different prognostic models and their customization on institutional comparison of intensive care units. *Crit Care Med* 35:2553-2560
18. Zhu BP, Lemeshow S, Hosmer DW et al (1996) Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. *Crit Care Med* 24:57-63

Chapter 4

19. Moreno R and Apolone G (1997) Impact of different customization strategies in the performance of a general severity score. *Crit Care Med* 25:2001-2008
20. Morales U, Peters SG, Afessa B (2003) Hospital mortality rate and length of stay in patients admitted at night to the intensive care unit. *Crit Care Med* 31:858-863
21. Uusaro A, Kari A, Ruokonen E (2003) The effects of ICU admission and discharge times on mortality in Finland. *Intensive Care Med* 29:2144-2148
22. Laupland KB, Shahpori R, Kirkpatrick AW et al (2008) Hospital mortality among adults admitted to and discharged from intensive care on weekends and evenings. *J Crit Care* 23:317-324
23. Cram P, Hillis SL, Barnett M et al (2004) Effects of weekend admission and hospital teaching status on in-hospital mortality. *Am J Med* 117:151-157
24. Kuhn G (2001) Circadian rhythm, shift work, and emergency medicine. *Ann Emerg Med* 37:88-98

Chapter 5

Determinants of mortality after hospital discharge in intensive care patients; Literature review and Dutch cohort study

Sylvia Brinkman, Ferishta Bakhshi-Raiez, Ameen Abu-Hanna, Evert de Jonge, Nicolette F. de Keizer

In press for publication in Critical Care Medicine, 2013



Abstract

Objectives: First, to conduct a literature review on the long-term mortality of ICU patients and its determinants. Second, to assess the influence of the found determinants on the 3, 6, and 12 months mortality after hospital discharge in the Dutch ICU population.

Design: Combination of a literature review to evaluate determinants of long-term mortality and a Dutch cohort study in which the found determinants are applied.

Setting: Pubmed and EMBASE were searched on English written articles published between 1966 and 2011. The cohort study was conducted in ICU patients from 81 Dutch mixed ICUs.

Data: 24 articles with a main focus on describing or predicting the case-mix adjusted long-term mortality of the general ICU population were identified. The cohort study consisted of 48,107 ICU patients who were discharged alive from the hospital between January 1st 2007 and October 1st 2010.

Interventions: None

Measurements and Main Results: The included articles were summarized on patient and study characteristics, methods, results and determinants used for case-mix adjustment. Additionally, the quality of the included articles was assessed using a checklist for studies on long-term survival. The median mortality rate of the general ICU population 1 year after ICU admission was 24% (range 16-44%). The determinants used for case-mix adjustment differed widely between the studies.

In the cohort study we found that age, reason for ICU admission, and co-morbidities were associated with all long-term mortality end-points. However, the magnitude and direction of the influence by these determinants differed for the different endpoints (i.e. 3, 6, and 12 months after hospital discharge).

Conclusions: The long-term mortality found in the included articles were difficult to compare due to low quality, variation in case-mix, study design, and differences in case-mix adjustment. The most commonly used determinants in the literature were comparable to the most important determinants found in the Dutch cohort study.

Introduction

Quality indicators on care process and patient outcomes are increasingly used in the health care debate, especially in a complex and expensive environment such as intensive care (1-3). In the intensive care (ICU), the ICU and in-hospital mortality are commonly used as quality indicators while the long-term mortality is often ignored as it is harder to obtain. Yet, it can be argued that the long-term mortality is even more important than the ubiquitously used in-hospital mortality as little is achieved if patients die soon after hospital discharge. Therefore, the first aim of this study was to conduct a literature review to describe the long-term mortality of ICU patients.

The risk of mortality is highly correlated to the underlying case-mix of patients (e.g. highly severe patients have a higher mortality risk). Consequently the ICU, in-hospital, or long-term mortality of patients should be adjusted for case-mix differences before they can be used as quality indicators, for instance to compare the quality of care between different ICUs. In the literature numerous determinants are used to adjust for case-mix differences, though insight in the determinants for long-term mortality in ICU patients is lacking. Therefore, the second aim of this study was to summarize the literature on the determinants which are used for case-mix correction and to assess which of the found determinants are actually of influence on the long-term mortality. The determinants identified in the literature review will be applied to the Dutch ICU population to assess their influence on the long-term mortality of ICU patients after hospital discharge. The study will use data from a large Dutch ICU quality registry, i.e. the National Intensive Care Evaluation (NICE) registry (4).

Methods

Literature review

Figure 5.1 shows the search strategy used to identify the relevant articles in Pubmed and EMBASE published between January, 1st 1966 and January, 1st 2011. The search was divided in four parts. In part A, we applied keywords and MeSH terms to identify the ICU population. In part B, we applied keywords to identify articles reporting on the long-term mortality. In part C we searched for terms related to case-mix adjustment in long-term analyses. In the final part (part D) the results of the first three parts were combined using the Boolean operator "AND" to identify the relevant set of articles.

We used the following entry criteria: analyses of a general adult (age \geq 18 years) ICU population (no analyses of specific diagnostic subgroups), original study with main focus on describing or predicting long-term mortality, and no case report. Furthermore, the long-term mortality of ICU patients had to be corrected for at least one confounder and

had to be assessed at least 3 months after ICU admission. This inclusion criterion is based on the study by Taori et al. which showed that a minimum of 90 days follow-up is necessary to fully capture the mortality effect of certain subgroups (5). Based on title and abstract, two reviewers (SB and FBR) independently selected English language articles that met these inclusion criteria. In case of disagreement a third reviewer could be consulted (NdK). The full text of the selected articles was read to determine the final inclusion. Searching was supplemented by scanning the reference lists of the included articles. The included articles were summarized on patient characteristics, study characteristics, methods, and results, e.g. the start-point and endpoint of follow-up and the determinants which were used for case-mix adjustment. Additionally, the quality of the included articles was assessed using a checklist with quality criteria for studies on long-term survival. These quality criteria were based on the list of recommendations for reporting on long-term survival presented by Williams et al. (6) complemented with the quality criteria on prognosis studies in systematic reviews by Hayden et al (7). The combined checklist addresses the following components: study participation, study attrition, prognostic factor measurement, outcome measurement, and analysis (see table 5.2).

Cohort study: Determinants in the Dutch ICU population

Data

Data from the Dutch National Intensive Care Evaluation (NICE) foundation registry was used to assess the influence of the different determinants identified in literature on different end-points of follow-up in the Dutch ICU population. The NICE registry contains demographic, physiological and clinical data of all consecutive ICU patients admitted to participating ICUs, including the Acute Physiology and Chronic Health Evaluation (APACHE) II, III, and IV (8-10) and the Simplified Acute Physiology Score (SAPS) II (11) variables. The 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. The NICE initiative is officially registered according to the Dutch Personal Data Protection Act.

The NICE registry includes data until hospital discharge, so the long-term mortality is unavailable. The long-term mortality was obtained by linking the NICE registry to a national administrative database from health insurance companies (insurance claims database of Vektis). The databases were anonymously linked by a deterministic linkage algorithm (12) that used the hospital of admission, gender, date of birth, ICU admission date, and ICU discharge date. After the linkage the state of the patients (either alive or death) on January 1st 2011, and if applicable the date of death, was extracted from the insurance claims database. In this study the linked NICE data of all ICU patients that were

discharged alive between January 1st 2007 and October 1st 2010 was used. To avoid double inclusion of patients, we excluded the patients that were discharged to another hospital. At the end of this study period, 81 Dutch ICUs (85% of all Dutch ICUs) recorded data on all their admissions in the NICE registry.

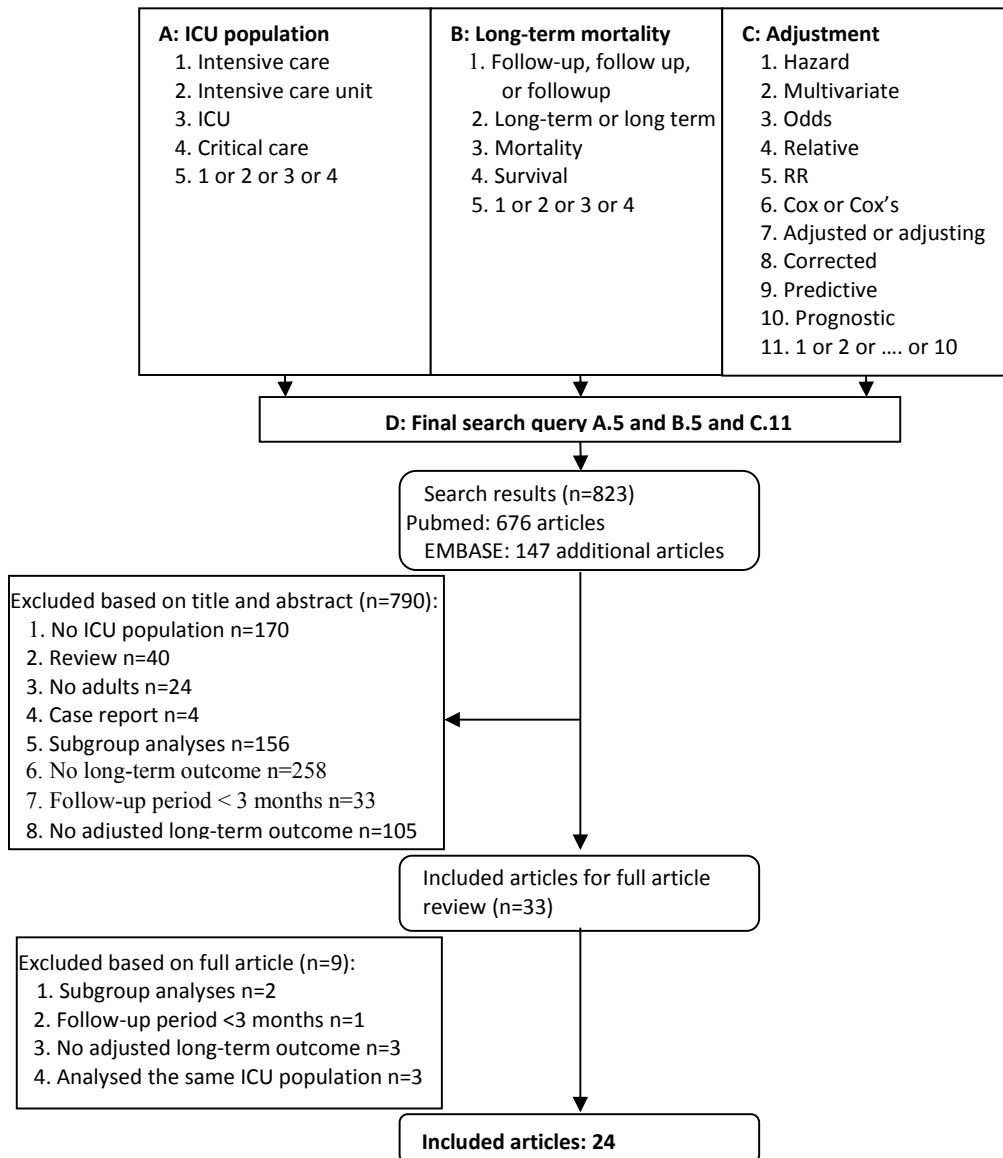


Figure 5.1: flow diagram of inclusion of the articles

Assessment of determinants

To assess the influence of the determinants found during the literature review, three multivariate Cox proportional hazard models were developed (i.e. to predict the 3, 6, and 12 months mortality after hospital discharge) including the found determinants. To assess which determinants have a statistically significant effect on the 3, 6, and 12 months mortality after hospital discharge a tenfold cross validation was used. In cross validation the dataset is randomly split in 10 equally sized subsets. In each of the 10 folds one of the subsets is used for validation and the rest 9 subsets are used to develop the Cox proportional hazard model. Model development in each fold relied on stepwise backward variable selection based on the Akaike Information Criterion (AIC) (13). The determinants that were selected in at least six of the ten developmental sets (i.e. significantly decreased the AIC of the model) were identified. Subsequently, in order to assess the hazard ratio of the identified determinants, a Cox proportional hazard model including only the identified determinants was fit in the developmental sets of a 10 fold cross validation design. The 95% confidence intervals of the hazard ratios are obtained by adding and subtracting 1.96 times the standard error of the 10 results. All statistical analyses were performed using the statistical environment R version 2.14.1 and PASW statistics 18.

Results

Literature review

The literature review identified 823 articles. Based on the title and abstract, 33 articles were selected for inclusion. In 793 (96.2%) of the cases there was an agreement in the inclusion of the articles between the two reviewers. For the remaining articles the reviewers were able to reach consensus on final inclusion of the articles. Figure 5.1 summarizes the inclusion process and provides the reasons for exclusion. After reading the full papers, 27 articles met the inclusion criteria of which four articles described the same ICU population admitted to one hospital between 1987 and 2002 (14-16). We only included the results of the most recent study (17). Finally, 24 articles were included in the literature review. No additional articles were found by scanning the reference lists of the included articles. Based on the ICU population of interest, we categorized 9 articles as general ICU population, 5 articles as general elderly ICU population (age > 64 years) and 10 articles as specific ICU population (i.e. general ICU population with a specific focus, including ICU patients with a high length of ICU stay, mechanically ventilated ICU patients, and medical or surgical ICU patients). The characteristics of the included articles are provided in table 5.1.

Table 5.1: Baseline characteristics of the included articles

Group	Author	Country	Study period	Start point	Years of follow-up	N	Age Mean \pm SD	% Male	Severity of illness	LOS ICU	Population of interest
General	Williams et al.	Australia	1987-2002	Hospital discharge	17	19921	57 ^b	68	6.8 \pm 4.6 ^e	NR	General
	Nilsson et al.	Sweden	1997-1998	ICU admission	5	92	65 \pm 17 ^b	62	18 (13-24) ^{cd}	1 (1-3) ^c	General
	Wright et al.	UK	1985-1992	ICU admission	12	2104	54 \pm 18 ^b	NR	14 \pm 7.8 ^d	4.5 \pm 7.2 ^b	General
	Keenan et al.	Canada	1994-1997	Hospital discharge	3	27103	54 \pm 24 ^b	57,1	NR	NR	General
	Niskanen et al.	Finland	1987-1988	ICU admission	6	12180	57 \pm 12 ^b	62,9	11,7 \pm 7.4 ^d	3.3 \pm 5.8 ^b	General
	Laupland et al.	Canada	1999-2002	ICU admission	2.8	4845	65 (51-74) ^f	62	24.9 \pm 8.8 ^d	NR	General
	Rockwood et al.	Canada	NR	ICU discharge	1	984	NR	63	17 \pm 9 ^d	4.1 \pm 5.8 ^b	General
	Ridley et al.	UK	1985-1987	ICU discharge	3.5	479	55 ^c	NR	13 ^d	NR	General
	Engoren et al.	Helsinki	2001-2002	ICU admission	5	2213	59 \pm 18 ^b	57	24 \pm 22 ^d	4 \pm 5 ^b	General
	Wunsch et al.	USA	2003-2006	Hospital discharge	3	35308	78 \pm 6.9 ^b	45,8	NR	1 (1-3) ^c	>65 year
	Somme et al.	France	1991-1996	ICU discharge	5.9	410	77 \pm 0 ^b	61	20.5 \pm 0.7 ^d	9.9 \pm 0.7 ^b	>74 years
	Elderly	Boumentil et al.	France	1998-2000	ICU discharge	2	233	86 \pm 4 ^b	39,9	45,1 \pm 18.9 ^f	6.3 \pm 5.8 ^b
Sacanella et al.		Helsinki	NR	ICU discharge	1.5	230	75 \pm 6 ^b	61	19,7 \pm 5.7 ^d	11,7 \pm 11.6 ^b	>64 years medical patients
Dardaine et al.		France	1989-1990	ICU discharge	1.5	110	78 \pm 1 ^b	55	18 \pm 0.7 ^f	23 \pm 4 ^b	>70 years ventilated > 1 days
Schneider et al.		Germany	1993-2005	196 days after ICU admission	2	1462	64 (16-96) ^g	69	11 (0-36) ^{gd}	9 (5-177) ^g	Surgical patients LOS > 4 days
Hartl et al.		Germany	1993-2005	ICU discharge	12	392	65 \pm 14 ^b	71,5	18,4 \pm 6.9 ^d	62,8 \pm 46.4 ^b	Surgical patients LOS ICU > 28 days
Iribarren et al.		Spain	1999-2000	Hospital discharge	1	283	60 \pm 17 ^b	65,4	14 (10-19) ^{cd}	3 (2-8) ^c	LOS ICU > 1 day
Hofnuss et al.		Netherlands	2000-2004	ICU discharge	0.5	451	71 (63-71) ^e	61,2	19 (15-23) ^{cd}	8 (5-16) ^c	LOS ICU > 2 days
Carden et al.		New Zealand	1999-2004	Hospital discharge	1	207	58 ^b	61,4	NR	14 ^b	LOS ICU > 7 days
Friedrich et al.		Toronto	2001-2004	Hospital discharge	0.5	266	61 \pm 17 ^b	62	23 \pm 8 ^d	NR	LOS ICU > 30 days
Grandeur et al.		Finland	2001-2004	ICU discharge	2	765	63 \pm 15 ^b	60,5	30 \pm 16 ^f	2 (1-4) ^c	Medical patients LOS ICU > 1 day
Specific	Babuín et al.	USA	2000-2001	30 days after ICU adm.	5	929	67 \pm 16 ^b	56,4	63,7 \pm 29.5 ^d	NR	Medical patients LOS ICU > 1 day
	Chelluri et al.	USA	1997-1999	ICU discharge	1	817	65 ^c	54,2	NR	NR	Mechanical ventilated > 2 days
	Combes et al.	USA	1995-1999	ICU discharge	3	197	63 \pm 14 ^b	66	46 \pm 13 ^f	42 \pm 25 ^b	Mechanical ventilated > 14 days

NR: not reported, a: Median (range), b: Mean \pm SD, c: Median (IQR), d: APACHE II score, e: APACHE II score, f: SAPS II score

Determinants of mortality after hospital discharge

In these articles, the information on the long-term mortality was assessed by contacting the patients or their families by phone (n=5) (18-22), by using a death register (n=10) (16,23-31), or by using a death register as well as a telephone call (n=5) (32-36). Four studies did not report on the source of the long-term mortality information (37-40). The starting point of the follow-up period differed between the included articles, i.e. ICU admission (n=5), 30 days after ICU admission (n=1), 196 days after ICU admission (n=1), ICU discharge (n=10), and hospital discharge (n=7).

In 17 studies the unadjusted long-term mortality of the population of interest was provided (16,18,19,21,22,24-26,28-31,33,35,36,38,40), mainly represented as a Kaplan-Meier curve. Figure 5.2 shows the large variation in long-term mortality between the studies.

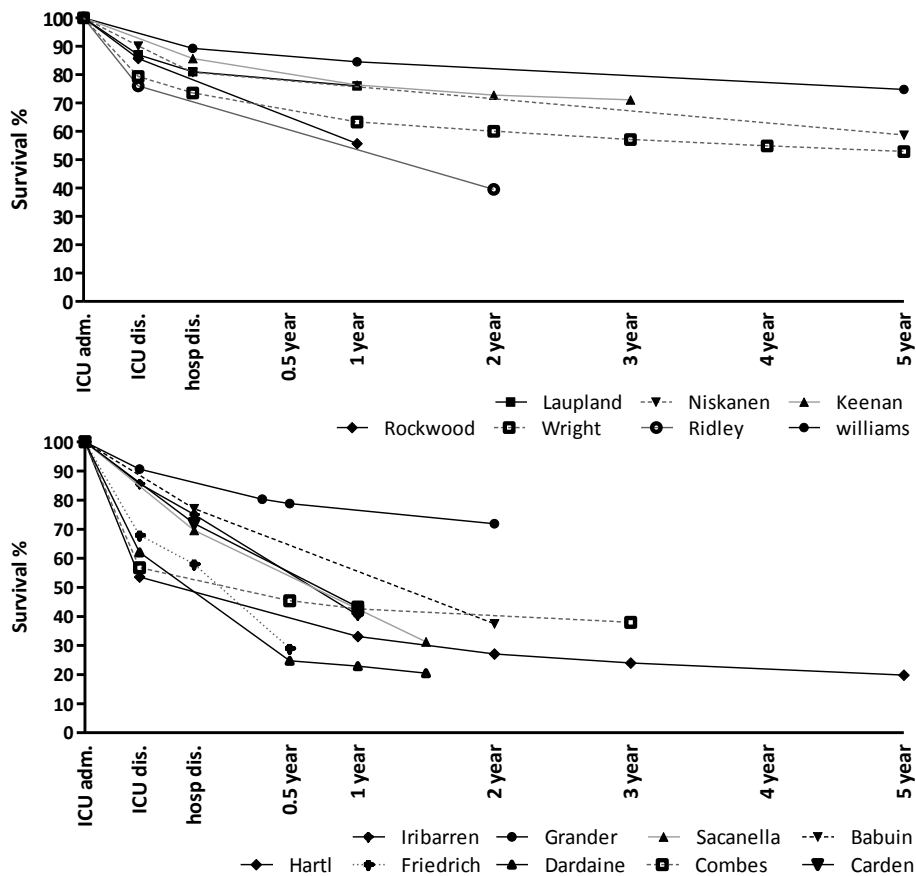


Figure 5.2: Survival of the ICU population at ICU admission (adm.), ICU discharge (dis.), hospital discharge (hosp. dis.), and subsequent years. The upper graph represents the general ICU population and the lower graph represents the older (Sacanella and Dardaine) and specific ICU populations (i.e. ICU patients with a high length of ICU stay, mechanically ventilated, or medical patients).

The median mortality rate 1 year after ICU admission in the total population was 51% (range 16-77%, N=10), in the general ICU population 24% (range 16-44%, N=5), and in the general elderly and specific ICU populations (e.g. patients with a high length of ICU stay) 58% (range 57-67%, N=5). The (number of) determinants for which the long-term mortality was adjusted also varied widely between the studies (Table 5.2). The most common determinants were age, severity of illness, and co-morbidities. The severity of illness of the ICU population was expressed using different scales. Most of the studies (n=15) reported the severity of illness in terms of the acute physiology and chronic health evaluation (APACHE) II score (21-23,25,27-30,35), 4 studies in terms of the simplified acute physiology score (SAPS) II score (18-20,33), 2 studies in terms of the APACHE III score (24,32), and 1 study in terms of the acute physiology score (APS) of the APACHE II model (16). Two studies did not report on severity of illness as a determinant (26,31). The more rarely used determinants were the use of vasoactive drugs, race and Glasgow Coma Scale. Next to patient-related determinants to correct for case-mix differences, some studies also used organizational factors such as type of ICU/hospital, length of ICU/hospital stay, and ICU discharge destination. The number of determinants included in the analyses differed between the studies, with a minimum of 2 determinants (29) and a maximum of 13 determinants (25).

Table 5.2 shows the results of the quality assessment of the articles based on the 23 quality criteria in our checklist. The number of criteria that were met ranged from 11 (34, 41) to 19.5 (32) and the median number of criteria fulfilled was 16. Low scores in the quality assessment mainly concerned not describing the study attrition (e.g. reasons for loss to follow-up, description of attempts to collect data on patients who dropped out (if no attempts were undertaken then this criterion was not fulfilled), and description of characteristics of censored and non-censored patients). Furthermore, the criteria for description of excluded patients, description of proportion of patients with complete data and handling missing data, and checks on data consistency were also often not fulfilled.

Table 5.3: determinants used in the articles to adjusted the long-term mortality

Group	Author	Age	Severity of illness	Co-morbidities	Gender	Reason for ICU admission	LOS ICU/hospital	Mechanical ventilation	ICU admission type	Previous hospital or ICU admission	Blood values	Patient characteristic (race, income, marital status, residence)	Inotropic support/ vasoactive drug	Year of ICU admission	Type of hospital	Other	Total number of determinants
General	Williams et al.	+	+	+	+	+	-	-	+	-	-	-	-	+	-	-	7
	Nilsson et al.	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	5
	Wright et al.	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	3
	Keenan et al.	+	-	+	+	+	-	-	-	+	-	+	-	-	+	-	8
	Niskanen et al.	+	+	+	+	+	-	-	+	-	-	-	-	-	-	-	6
	Laupland et al.	+	+	+	+	+	-	-	-	-	+	-	-	-	-	+	6
	Rockwood et al.	+	+	+	-	-	+	-	-	+	-	-	-	-	-	-	5
	Ridley et al.	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	2
	Engoren et al.	+	+	-	+	+	+	+	+	-	+	-	-	-	+	+	13
Elderly	Wunsch et al.	+	-	+	-	-	-	-	-	+	-	-	-	-	-	+	4
	Somme et al.	+	+	+	+	-	+	+	-	-	-	-	-	-	-	+	7
	Boumentil et al.	+	+	+	+	-	-	+	-	-	-	-	-	-	-	-	5
	Sacanella et al.	+	+	+	-	-	+	-	-	-	-	-	-	-	-	+	5
	Dardaine et al.	+	+	+	+	-	-	-	-	+	-	+	-	-	-	-	6
Specific	Schneider et al.	+	+	+	+	+	-	-	+	-	-	-	-	+	-	-	7
	Hartl et al.	+	+	+	-	+	-	-	-	-	-	-	-	-	-	+	5
	Iribarren et al.	+	-	+	-	-	+	-	-	-	-	-	-	-	-	-	3
	Hofhuis et al.	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	4
	Carden et al.	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	3
	Friedrich et al.	+	+	+	-	-	+	+	-	+	-	-	+	-	-	-	7
	Grander et al.	+	+	+	-	+	+	-	-	-	+	-	-	-	-	-	6
	Babuin et al.	+	+	+	+	-	-	-	-	-	+	-	+	-	-	+	7
	Chelluri et al.	+	+	+	+	-	-	-	+	-	-	+	-	-	-	-	6
	Combes et al.	+	-	+	-	+	-	+	-	-	-	-	-	-	-	-	4
Number of studies that used the determinant		24	19	18	13	10	7	5	5	5	5	3	2	2	2	7	

Cohort study: Determinants in the Dutch ICU population

Data

From January 1st, 2007 to October 1st, 2010, 68,847 ICU patients not discharged to another hospital, fulfilling the APACHE IV inclusion criteria (10) and with complete data on all found determinants were registered in the NICE registry. These patients were admitted to one of the 81 Dutch ICUs included in the study and survived until hospital discharge. All included ICUs were mixed medical-surgical units located in university hospitals (n=7), teaching hospitals (n=27) or non-teaching hospitals (n=47). Of the 68,847 records, 48,107 (69.9%) records could be linked with the insurance claims database and were included in the analyses.

During the literature review numerous determinants were found which are categorized into 15 groups, as presented in table 5.3. The most commonly used determinants are collected as part of the NICE registry. However, some organizational determinants and patient characteristics such as race, residence, income, marital status, current smoking, some blood values (CRP value, troponin, protein C, and antithrombin), earlier hospital or ICU admission, number of surgical revisions, TISS, and OMEGA score are not available in the NICE registry.

Assessment of determinants

The determinants which were available in the NICE registry were inserted as such into the three Cox proportional hazard models. The reason for ICU admission was inserted in the model by using the main categories of the APACHE IV reasons for ICU admission classification (10), i.e., non-operative and postoperative cardiovascular, gastro-intestinal, genitourinary, hematological, metabolic, musculoskeletal, neurologic, respiratory, transplant, and trauma. The severity of illness was expressed in the APACHE IV Acute Physiology Score (APS). The determinants length and weight were modeled by inserting the Body Mass Index (BMI) of the patients in the Cox proportional hazard models. Table 5.4 gives the average and corresponding confidence interval of the hazard ratios of the determinants that had at least in six of the ten developmental sets a significant influence (i.e. significantly increased the AIC of the model) on the 3, 6 or 12 months mortality after hospital discharge. The most important determinants based on the hazard ratios were age, reason for ICU admission, and co-morbidities, especially the co-morbidities metastatic neoplasm and cirrhosis. The determinants that are selected after stepwise backward selection based on the AIC were mostly the same for the different end-points of follow-up. Source of admission has only an influence on the 3 months mortality after hospital discharge while gender has only influence on the 6 and 12 months mortality after hospital discharge.

Table 5.4: Hazard ratio of the determinants that were selected during the 10 fold cross validation design (assessed using a multivariate Cox model).

Determinant	Hazard ratio (confidence interval)		
	3 months after hospital discharge	6 months after hospital discharge	12 months after hospital discharge
Year of ICU admission	NS	NS	1.02 (1.01,1.02)
Admission type			
Elective surgery	Reference	Reference	Reference
Urgent surgery	1.33 (1.31,1.34)	1.08 (1.07,1.09)	0.95 (0.94,0.96)
Medical	2.17 (2.09,2.26)	1.80 (1.76,1.85)	1.62 (1.59,1.65)
Female gender	NS	0.87 (0.87,0.88)	0.86 (0.85,0.86)
Age			
<54	Reference	Reference	Reference
54-66	2.00 (1.95,2.04)	1.73 (1.70,1.76)	1.67 (1.65,1.69)
66-75	2.73 (2.66,2.79)	2.34 (2.30,2.37)	2.17 (2.14,2.19)
>75	4.58 (4.48,4.67)	3.80 (3.76,3.85)	3.46 (3.43,3.49)
Source of admission			
Floor same hospital	Reference	Reference	Reference
Other department of same hospital	0.90 (0.89,0.91)	NS	NS
Other hospital	1.03 (1.00,1.05)	NS	NS
Other admission source	0.85 (0.83,0.88)	NS	NS
Mechanical ventilation in first 24hrs of ICU	0.93 (0.92,0.94)	0.89 (0.88,0.90)	0.91 (0.90,0.91)
Use of vasopressors in first 24hrs of ICU	0.94 (0.93,0.95)	0.94 (0.93,0.95)	0.93 (0.92,0.93)
Physiology condition (APACHE IV APS Score)			
<25	Reference	Reference	Reference
25-38	1.32 (1.30,1.35)	1.26 (1.25,1.28)	1.14 (1.13,1.15)
38-57	1.59 (1.56,1.62)	1.44 (1.43,1.46)	1.28 (1.27,1.29)
>=57	1.89 (1.84,1.93)	1.67 (1.65,1.69)	1.41 (1.40,1.42)
Length of hospital stay in days			
<5.81	Reference	Reference	Reference
5.81-9.86	0.88 (0.86,0.89)	0.95 (0.94,0.95)	0.99 (0.98,1.00)
9.86-18.6	1.19 (1.18,1.21)	1.24 (1.22,1.26)	1.32 (1.30,1.33)
>=18.6	1.61 (1.58,1.64)	1.72 (1.70,1.75)	1.74 (1.71,1.76)
Maximal urea in first 24 hrs of ICU adm.			
<4.7	Reference	Reference	Reference
4.7-6.7	1.15 (1.13,1.17)	1.14 (1.13,1.16)	1.09 (1.08,1.10)
6.7-10.5	1.55 (1.53,1.58)	1.53 (1.51,1.54)	1.42 (1.41,1.44)
>=10.5	2.35 (2.30,2.39)	2.19 (2.17,2.22)	1.95 (1.94,1.97)
INR in first 24 hrs of ICU adm.			
<1.1	Reference	Reference	Reference
1.1-1.2	1.02 (1.01,1.03)	1.00 (0.99,1.00)	1.01 (1.00,1.02)
1.2-1.5	1.21 (1.20,1.22)	1.11 (1.10,1.12)	1.06 (1.05,1.07)
>=1.5	1.25 (1.24,1.27)	1.21 (1.20,1.21)	1.20 (1.19,1.20)
Minimal platelets in first 24 hrs of ICU adm.			
>=234	Reference	Reference	Reference
171-234	0.74 (0.73,0.75)	0.75 (0.74,0.76)	0.81 (0.80,0.82)
121-171	0.66 (0.65,0.67)	0.65 (0.65,0.66)	0.69 (0.69,0.70)
<121	0.70 (0.70,0.71)	0.67 (0.67,0.68)	0.69 (0.68,0.69)
Maximal creatinine in first 24 hrs of ICU			
<66	Reference	Reference	Reference
66-85	0.74 (0.73,0.74)	0.74 (0.73,0.75)	0.80 (0.80,0.81)
85-121	0.69 (0.68,0.70)	0.72 (0.71,0.72)	0.78 (0.77,0.79)
>=121	0.60 (0.60,0.61)	0.62 (0.61,0.62)	0.69 (0.68,0.70)

*NS: not selected during 10 fold cross validation design

Table 5.4: continued

Determinant	Hazard ratio (confidence interval)		
	3 months after hospital discharge	6 months after hospital discharge	12 months after hospital discharge
Minimal hematocrit in first 24 hrs of ICU			
>=0.35	Reference	Reference	Reference
0.3-0.35	1.06 (1.05,1.07)	1.15 (1.14,1.16)	1.18 (1.17,1.18)
0.26-0.3	1.10 (1.09,1.11)	1.19 (1.18,1.21)	1.22 (1.20,1.23)
<0.26	1.27 (1.26,1.29)	1.37 (1.35,1.38)	1.36 (1.34,1.37)
APACHE IV reason for ICU admission			
Cardiovascular Operative	Reference	Reference	Reference
Gastro-intestinal Operative	1.58 (1.55,1.60)	1.85 (1.82,1.88)	2.23 (2.21,2.26)
Genito-urinary Operative	1.92 (1.90,1.94)	2.36 (2.34,2.38)	2.90 (2.88,2.92)
Metabolic Operative	1.93 (1.64,2.23)	2.01 (1.77,2.25)	2.02 (1.87,2.17)
Musculoskeletal/skin Operative	2.59 (2.52,2.67)	2.73 (2.67,2.80)	2.81 (2.74,2.88)
Neurologic Operative	4.49 (4.42,4.56)	4.80 (4.74,4.86)	5.22 (5.15,5.28)
Respiratory Operative	1.90 (1.86,1.93)	2.15 (2.11,2.19)	2.55 (2.51,2.58)
Trauma Operative	2.77 (2.70,2.84)	2.91 (2.83,2.99)	2.74 (2.66,2.81)
Hematology Operative & Non Operative	1.42 (1.34,1.49)	1.91 (1.85,1.97)	1.99 (1.91,2.06)
Transplant Operative & Non Operative	0.11 (0.09,0.13)	0.39 (0.35,0.43)	0.73 (0.69,0.76)
Cardiovascular Non Operative	1.23 (1.19,1.28)	1.49 (1.45,1.52)	1.59 (1.56,1.61)
Gastro-intestinal Non Operative	1.00 (0.97,1.04)	1.22 (1.18,1.25)	1.36 (1.33,1.39)
Genito-urinary Non Operative	1.25 (1.20,1.29)	1.47 (1.42,1.51)	1.64 (1.60,1.67)
Metabolic Non Operative	1.37 (1.30,1.44)	1.52 (1.48,1.56)	1.63 (1.59,1.67)
Musculoskeletal/skin Non Operative	0.58 (0.52,0.64)	0.63 (0.58,0.69)	0.73 (0.67,0.79)
Neurologic Non Operative	1.67 (1.61,1.73)	1.70 (1.64,1.76)	1.64 (1.60,1.69)
Respiratory Non Operative	1.24 (1.19,1.29)	1.55 (1.50,1.60)	1.63 (1.60,1.66)
Trauma Non Operative	0.93 (0.89,0.98)	0.93 (0.88,0.97)	0.95 (0.92,0.99)
Co-morbidities			
Acute renal failure	0.91 (0.89,0.92)	0.87 (0.85,0.88)	0.86 (0.85,0.86)
COPD	1.37 (1.36,1.38)	1.29 (1.28,1.30)	1.30 (1.29,1.31)
Diabetes	1.13 (1.12,1.14)	1.11 (1.10,1.12)	1.14 (1.13,1.15)
Chronic renal insufficiency	1.37 (1.34,1.40)	1.36 (1.34,1.38)	1.33 (1.31,1.35)
Cirrhosis	2.64 (2.57,2.71)	2.39 (2.34,2.44)	2.17 (2.12,2.22)
Hematological malignancy	1.95 (1.90,2.00)	2.22 (2.18,2.27)	2.25 (2.21,2.28)
Metastatic neoplasm	3.00 (2.94,3.05)	3.06 (3.02,3.09)	2.77 (2.75,2.80)
Immunological insufficiency	1.18 (1.16,1.20)	1.16 (1.15,1.18)	1.22 (1.21,1.23)
Chronic respiratory insufficiency	1.46 (1.45,1.48)	1.39 (1.37,1.41)	1.41 (1.39,1.43)
Discharge destination			
Floor	Reference	Reference	Reference
Recovery/MC/Special care	1.07 (1.06,1.09)	1.04 (1.02,1.05)	0.97 (0.96,0.99)
Other discharge destination	3.05 (2.95,3.15)	2.43 (2.37,2.49)	2.03 (1.99,2.07)

Discussion

The literature review showed that crudely 76% of the general ICU population is still alive 1 year after ICU admission. In the literature, the most commonly used determinants for case-mix adjustment were age, severity of illness, co-morbidities, and gender. When applying the determinants found during the literature review to the Dutch ICU population, the three Cox proportional hazard models showed that the most important determinants for long-term mortality in the Dutch ICU population are age, reason for ICU admission and co-morbidities. Furthermore, this study showed that the selected determinants after stepwise backward selection based on the AIC were mostly the same for the different end-points of follow-up.

The differences in case-mix (e.g. demographic variables, severity of illness, diagnose, and ratio of medical and surgical patients), the determinants, and follow-up starting point (ICU admission, ICU discharge, or hospital discharge) of the included studies hampers firm conclusions on case-mix adjusted long-term mortality of ICU patients. Nevertheless, we attempted to summarize the unadjusted long-term mortality by Kaplan-Meier curves. Though, it should be stated that these results are an indication of the mortality rate in ICU patients but cannot be compared directly as they are not adjusted for a uniform set of case-mix parameters.

For our literature review we used the recommendations for reporting on long-term survival as presented in the review by Williams et al. (6) together with the quality criteria on prognosis studies in systematic reviews by Hayden et al. (7) to assess the quality of the included articles. Most of the studies score relatively well on the description of study participation, outcome measures, and analyses (with an exception for the description of the characteristics of the excluded patients). However, there are still great opportunities for improvement in the description of study attrition (i.e. the description of data completeness, handling missing values, and consistency checks on data) and prognostic factor measurement. By improving the description of the study attrition, the reliability of the study findings can be improved and methods leading to a low number of lost to follow-up can be assessed. A possible limitation of using these quality indicators is that they suffer a certain degree of subjectivity, we however endeavored to eliminate this subjectivity by applying the criteria independently by two reviewers.

In 2008 Williams et al. performed a single center study to identify determinants of long-term mortality of ICU patients (6). In this study age, co-morbidities, primary diagnose, and severity of illness were found as important determinants for long-term mortality. Williams et al. recommended comparing their results to other ICU populations to investigate the generalizability. Although the number of investigated determinants in their study was less and not based on a literature review, our results are comparable indicating that the identified determinants might be generalizable to other country settings. In our literature

review we also found studies that included determinants to correct for organizational factors such as type of ICU/hospital, length of ICU/hospital stay, and ICU discharge destination. This is no problem if the goal of a study is solely to assess the long-term mortality without comparison between different ICUs. However, in order to assess and to compare the quality of care between different ICUs using case-mix adjusted mortality, it is not desirable to correct for organizational factors as they might be part of the explanation for the identified differences in the quality of care. Our findings can be used to make a well-considered selection of the determinants to adjust the long-term mortality for benchmark purposes.

We were not able to analyze all determinants on their influence on long-term mortality as some determinants were lacking in the NICE registry, though most of the found determinants could be analyzed during the cohort study.

High creatinine and low platelets are often markers of acute deterioration of vital functions (e.g. in severe sepsis), which frequently have a poor short-term prognosis. It may be that there is some selection bias: patients with severe acute deterioration are only candidates for intensive treatments if there is no additional condition severely limiting life-expectancy. If these patients survive up to hospital discharge, long-term prognosis may be good which might explain the result that lower platelets and higher creatinine are protective. Furthermore, we showed that next to the APACHE IV APS score also the determinants urea, creatinine, and hematocrit were statistically significant predictors of long-term mortality. This is noteworthy as urea, creatinine, and hematocrit are all three part of the APACHE IV APS score, which might indicate that the APACHE IV model inadequately corrects for these blood values when predicting long-term mortality. The identified determinants for the long-term mortality of ICU patients are similar across the different endpoints of follow-up, though the influence of the determinants varies over the different endpoints. In most of the cases, the direction of the influence (i.e. increase or decrease of mortality risk) is similar over the different end-points. Only the direction of the influence of the determinants "urgent surgery and discharged to recovery", and "Medium Care or Special Care" changed over time, showing an increase of the mortality risk at 3 months after hospital discharge and a decrease of the mortality risk at 12 months after hospital discharge. Apparently once these patients survived the first 6 months after ICU admission the mortality risk decreases which might be explained by the fact that these patients have less co-morbidities compared to the reference patient group (e.g. the percentage of patients with metastatic neoplasm among urgent surgery patients is 4.4% and among elective surgery patients is 6.0%). The magnitude of the influence of the rest of the determinants found varied over the different end-points of follow-up, showing that some determinants have more influence on the mortality 3 months after hospital discharge (e.g. age and severity of illness) while other determinants have more influence

on the mortality 12 months after hospital discharge (e.g. length of hospital stay and reason for ICU admission). It should be noted, though, that the included determinants slightly differ between the different end-point of follow-up. As the influence of the determinants on the mortality at different end-points of follow-up has not been described earlier, these findings should be further investigated before they can be generalized to other country settings. Future studies should also focus on the implications of using the case-mix adjusted in-hospital mortality or the case-mix adjusted long-term mortality as quality indicator.

Conclusion

In the literature about 76% of the general ICU patients are still alive 1 year after ICU admission. It is difficult to compare the case-mix corrected long-term mortality of ICU patients among different studies due to differences in case-mix adjustment and outcome measures in terms of start-point and end-point of follow-up. The most commonly used determinants were age, severity of illness, and co-morbidities. After applying the determinants from the literature to the Dutch ICU population, the most important determinants for long-term mortality were age, reason for ICU admission, and several co-morbidities. The determinants that were selected for the different end-points of follow-up (i.e. 3, 6, and 12 months after hospital discharge) are comparable. Though, the magnitude and direction of the influence by these determinants differed between the follow-up end-points. Future studies should focus on the implications of using the case-mix adjusted in-hospital mortality or the case-mix adjusted long-term mortality as quality indicator used for benchmarking purposes.

References

1. de Vos M, Graafmans W, Keesman E, et al: Quality measurement at intensive care units: which indicators should we use? *J Crit Care* 2007; 22:267-274
2. Martin MC, Cabre L, Ruiz J, et al: Indicators of quality in the critical patient. *Med Intensiva* 2008; 32:23-32
3. Pronovost PJ, Berenholtz SM, Ngo K, et al: Developing and pilot testing quality indicators in the intensive care unit. *J Crit Care* 2003; 18:145-155
4. Dutch National Intensive Care Evaluation (NICE) foundation: Available online at: <http://www.stichting-nice.nl>.
5. Taori G, Ho KM, George C, et al: Landmark survival as an end-point for trials in critically ill patients--comparison of alternative durations of follow-up: an exploratory analysis. *Crit Care* 2009; 13:R128
6. Williams TA, Dobb GJ, Finn JC, et al: Long-term survival from intensive care: a review. *Intensive Care Med* 2005; 31:1306-1315
7. Hayden JA, Cote P, Bombardier C: Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006; 144:427-437

8. Knaus WA, Draper EA, Wagner DP, et al: APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13:818-829
9. Knaus WA, Wagner DP, Draper EA, et al: The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; 100:1619-1636
10. Zimmerman JE, Kramer AA, McNair DS, et al: Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34:1297-1310
11. Le Gall JR, Lemeshow S, Saulnier F: A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270:2957-2963
12. Roos LL, Wajda A: Record linkage strategies. Part I: Estimating information and evaluating approaches. *Methods Inf Med* 1991; 30:117-123
13. Akaike H: A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* 1974; 19:716-723
14. Ho KM, Knuiman M, Finn J, et al: Estimating long-term survival of critically ill patients: the PREDICT model. *PLoS One* 2008; 3:e3226
15. Ho KM, Dobb GJ, Knuiman M, et al: The effect of socioeconomic status on outcomes for seriously ill patients: a linked data cohort study. *Med J Aust* 2008; 189:26-30
16. Williams TA, Ho KM, Dobb GJ, et al: Effect of length of stay in intensive care unit on hospital and long-term mortality of critically ill adult patients. *Br J Anaesth* 2010; 104:459-464
17. Williams TA, Dobb GJ, Finn JC, et al: Determinants of long-term survival after intensive care. *Crit Care Med* 2008; 36:1523-1530
18. Boumendil A, Maury E, Reinhard I, et al: Prognosis of patients aged 80 years and over admitted in medical intensive care unit. *Intensive Care Med* 2004; 30:647-654
19. Combes A, Costa MA, Trouillet JL, et al: Morbidity, mortality, and quality-of-life outcomes of patients requiring ≥ 14 days of mechanical ventilation. *Crit Care Med* 2003; 31:1373-1381
20. Dardaine V, Constans T, Lasfargues G, et al: Outcome of elderly patients requiring ventilatory support in intensive care. *Aging (Milano)* 1995; 7:221-227
21. Iribarren-Diarasari S, Aizpuru-Barandiaran F, Munoz-Martinez T, et al: Health-related quality of life as a prognostic factor of survival in critically ill patients. *Intensive Care Med* 2009; 35:833-839
22. Sacanella E, Perez-Castejon JM, Nicolas JM, et al: Mortality in healthy elderly patients after ICU admission. *Intensive Care Med* 2009; 35:550-555
23. Carden GP, Graham JW, McLennan S, et al: Long-term outcome of long stay ICU and HDU patients in a New Zealand hospital. *Critical Care and Shock* 2008; 11:26-34
24. Babuin L, Vasile VC, Rio Perez JA, et al: Elevated cardiac troponin is an independent risk factor for short- and long-term mortality in medical intensive care unit patients. *Crit Care Med* 2008; 36:759-765
25. Engoren M, Arslanian-Engoren C: Long-term survival in the intensive care unit after erythrocyte blood transfusion. *Am J Crit Care* 2009; 18:124-131
26. Keenan SP, Dodek P, Chan K, et al: Intensive care unit admission has minimal impact on long-term mortality. *Crit Care Med* 2002; 30:501-507
27. Laupland KB, Zygun DA, Doig CJ, et al: One-year mortality of bloodstream infection-associated sepsis and septic shock among patients presenting to a regional critical care system. *Intensive Care Med* 2005; 31:213-219
28. Niskanen M, Kari A, Halonen P: Five-year survival after intensive care--comparison of 12,180 patients with the general population. *Finnish ICU Study Group. Crit Care Med* 1996; 24:1962-1967
29. Ridley S, Jackson R, Findlay J, et al: Long term survival after intensive care. *BMJ* 1990; 301:1127-1130
30. Wright JC, Plenderleith L, Ridley SA: Long-term survival following intensive care: subgroup analysis and comparison with the general population. *Anaesthesia* 2003; 58:637-642
31. Wunsch H, Guerra C, Barnato AE, et al: Three-year outcomes for Medicare beneficiaries who survive intensive care. *JAMA* 2010; 303:849-856
32. Chelluri L, Im KA, Belle SH, et al: Long-term mortality and quality of life after prolonged mechanical ventilation. *Crit Care Med* 2004; 32:61-69

33. Grander W, Dunser M, Stollenwerk B, et al: CRP Levels and Post-Intensive Care Unit Mortality in Non-Surgical Intensive Care Patients. *Chest* 2010;
34. Rockwood K, Noseworthy TW, Gibney RT, et al: One-year outcome of elderly and young patients admitted to intensive care units. *Crit Care Med* 1993; 21:687-691
35. Schneider CP, Fertmann J, Geiger S, et al: Long-term survival after surgical critical illness: the impact of prolonged preceding organ support therapy. *Ann Surg* 2010; 251:1145-1153
36. Somme D, Maillet JM, Gisselbrecht M, et al: Critically ill old and the oldest-old patients in intensive care: short- and long-term outcomes. *Intensive Care Med* 2003; 29:2137-2143
37. Friedrich JO, Wilson G, Chant C: Long-term outcomes and clinical predictors of hospital mortality in very long stay intensive care unit patients: a cohort study. *Crit Care* 2006; 10:R59
38. Hartl WH, Wolf H, Schneider CP, et al: Acute and long-term survival in chronically critically ill surgical patients: a retrospective observational study. *Crit Care* 2007; 11:R55
39. Hofhuis JG, Spronk PE, van Stel HF, et al: Quality of life before intensive care unit admission is a predictor of survival. *Crit Care* 2007; 11:R78
40. Nilsson G, Astermark J, Lethagen S, et al: Protein C levels can be forecasted by global haemostatic tests in critically ill patients and predict long-term survival. *Thromb Res* 2005; 116:15-24
41. Nilsson G, Astermark J, Lethagen S, et al: Protein C levels can be forecasted by global haemostatic tests in critically ill patients and predict long-term survival. *Thromb Res* 2005; 116:15-24

Chapter 6

Mortality after hospital discharge
in ICU patients

Sylvia Brinkman, Evert de Jonge, Ameen Abu-Hanna,
M. Sesmu Arbous, Dylan W. de Lange, Nicolette F. de Keizer

In press for publication in Critical Care Medicine, 2013



Abstract

Objectives: To assess the mortality risk of intensive care unit (ICU) patients after hospital discharge and compare it to mortality of the general Dutch population.

Design: Cohort study of ICU admissions from a national ICU registry linked to administrative records from an insurance claims database.

Setting: 81 Dutch ICUs.

Patients: 91,203 ICU patients who were discharged alive from the hospital between January 1st 2007 and October 1st 2010.

Interventions: None

Measurements and Main Results: The unadjusted observed survival was inspected by Kaplan-Meier curves. Mortality risk at 1, 2, and 3-year after hospital discharge was 12.5%, 19.3%, and 27.5% respectively. The 3-year mortality after hospital discharge in ICU patients was higher than the weighted average of the gender and age specific death risks of the general Dutch population (27.5% versus 8.2%). The 1-year mortality after hospital discharge was adjusted for case-mix differences by a set of determinants which showed a statistically significant influence on the outcome in a 10-fold cross validation. The elective and cardiac surgical patients have statistically significantly better mortality outcomes (adjusted hazard ratio (HR_{adj}) 0.73 and 0.28 respectively), whereas medical patients and patients admitted for cancer have statistically significantly worse mortality outcomes (HR_{adj} 1.41, 1.94 respectively) compared to other ICU patients. Urgent surgery patients and patients with a subarachnoid hemorrhage, trauma, acute renal failure or severe community-acquired pneumonia did not differ statistically from the other ICU patients after adjustment for case-mix differences.

Conclusions: In-hospital mortality underestimates the true mortality of ICU patients as the mortality in the first months after hospital discharge is substantial. Most ICU patients still have an increased mortality risk in the subsequent years after hospital discharge compared to the general Dutch population. The mortality after hospital discharge differs widely between ICU subgroups. Future studies should focus on the analysis of mortality after hospital discharge that is attributable to the former ICU admission.

Introduction

The increasing use of quality indicators to assess the clinical process performance and patient outcomes is an important issue in the health care debate, especially in a complex and expensive environment such as the intensive care unit (ICU) (1,2). Currently, the observed in-hospital mortality is commonly used to describe the outcome of ICU patients. The in-hospital mortality adjusted for case-mix is commonly used as quality indicator to compare the performance among different hospitals. Unfortunately, for several reasons, mortality may still be higher than expected for many months after hospital discharge. First, patients may have an increased mortality risk due to critical illness related disorders, such as weakness, immunological insufficiency or other co-morbidities. Also, patients may still be (moribundly) ill at hospital discharge, i.e. if they are discharged from one hospital to another, or to a palliative care facility. Thirdly, the critical illness and ICU admission may accelerate the underlying diseases (3). A more optimal way to estimate quality of ICU care, both from the point of view of health care institutions and the patient, is to consider the mortality sometime after hospital discharge. It can even be argued that assessing the mortality after hospital discharge is more important than the ubiquitously used in-hospital mortality as little has been achieved when patients die soon after hospital discharge. Information on mortality after hospital discharge could help clinicians to identify groups at elevated mortality risk after hospital discharge to identify interventions that improve their long-term mortality (4). Clinical registries commonly register patient data and mortality up until hospital discharge and are frequently used to monitor and analyze the quality of health care based on in-hospital mortality. The in-hospital mortality can be registered quite easily and reliably while complete registration of mortality after hospital discharge may be more challenging and time consuming. This partly explains why ICU and hospital mortality have been described more frequently than the long-term mortality. However, linking clinical databases with administrative databases, e.g. from insurance companies, offers the opportunity to assess long-term mortality.

The aim of this study was to assess (case-mix adjusted) mortality after hospital discharge in a large Dutch ICU population and compare it to mortality of the general Dutch population. In this study we focused on the total ICU population as well as on specific ICU subgroups.

Materials and Methods

Data

This study was a retrospective cohort study, comprising all ICU patients discharged from the hospital between January 1st 2007 and October 1st 2010. Data on this population were derived by linking the Dutch National Intensive Care Evaluation (NICE) registry (5) to a national administrative database of health insurance companies (insurance claims database of Vektis), covering 95% (in 2008) of the insured Dutch population (6). As a health insurance is compulsory for all citizens in the Netherlands, the claims database is representative for the total Dutch population.

The NICE registry contains demographic, physiological and clinical data of all consecutive ICU patients admitted to participating ICUs, including the Acute Physiology and Chronic Health Evaluation (APACHE) IV (7), chronic comorbidity and reason for ICU admission. During the study period, approximately 85% of all Dutch ICUs recorded data of all their admissions in the NICE registry. We used the APACHE IV score to correct the mortality 1-year after hospital discharge for the severity of illness at admission. Therefore, we could only include admissions fulfilling the inclusion criteria of the APACHE IV model (7). ICU patients discharged to other hospitals were excluded as their in-hospital outcome is unknown. The NICE registry includes data until hospital discharge, so the mortality after hospital discharge is unavailable. However, the mortality after hospital discharge is obtained by linkage to the insurance claims database of Vektis. The insurance claims database contains, amongst others, information on the vital status of patients. After the insurance claims database was linked to the NICE registry, the status of the patients (either alive or death) on January 1st 2011, and if relevant, the date of death, was extracted from the insurance claims database. The data used in this study 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. The data is officially registered according to the Dutch Personal Data Protection Act.

The records from the NICE registry and insurance claims database are anonymously linked by a deterministic linkage algorithm (8) that used the hospital of admission, gender, date of birth, ICU admission date, and ICU discharge date. In this method the variables in both databases must be exactly the same for a positive match. In the algorithm, the patients are first linked by using the separately declared ICU days of the insured patients, the remaining unlinked patients are further linked by using the declared hospitalization periods for complex interventions such as cardiac surgery and transplantations. If the ICU admission registered in the NICE database occurs in the declared hospitalization period, the records are linked. In the final linked dataset, patients' vital status was assessed on

January 1st, 2011 and patients were assumed to be alive if there was no date of death in the insurance claims database at that time. Case-mix characteristics of the linked and non-linked records of the NICE registry were compared using Student's t-tests for normally distributed data and Mann-Whitney tests for non-normally distributed data to evaluate potential bias due to incomplete linkage.

Mortality after hospital discharge

To assess patient mortality after hospital discharge we performed Kaplan-Meier and Cox proportional hazard analyses for the whole ICU population, for ICU subgroups based on admission type (elective surgery, urgent surgery, and medical (i.e. nonsurgical)), and based on reason for ICU admission (cardiac surgery, subarachnoid hemorrhage, acute renal failure, severe community-acquired pneumonia, cancer, and trauma). Choice for these subgroups was based on existing literature on long-term mortality (9-14), expert opinion, and the availability of sufficient (i.e. more than 500) ICU admissions that survived hospitalization during our study period. Definitions of the subgroups are given in box 6.1.

Box 6.1: explanation of the diagnostic subgroups.

<p><u>Cardiac surgery subgroup:</u> Patients with a post-operative cardiovascular APACHE IV reason for ICU admission. However, cardiac arrest and sepsis APACHE IV reason for ICU admission were excluded as these diagnoses can be post-operative as well as non-operative.</p> <p><u>Subarachnoid hemorrhage subgroup:</u> Patients with a post-operative or non-operative subarachnoid hemorrhage/intracranial aneurysm or a non-operative subarachnoid hemorrhage/arteriovenous malformation as APACHE IV reason for ICU admission.</p> <p><u>Acute renal failure subgroup:</u> Patients with acute renal failure as APACHE IV reason for ICU admission and/or acute renal failure in the 24 hours after ICU admission.</p> <p><u>Severe community-acquired pneumonia (sCAP) subgroup:</u> Patients with pneumonia (aspiration, bacterial, fungal, parasitic, viral or other pneumonia) as APACHE IV reason for ICU admission and hospitalized for maximal 2 days before ICU admission.</p> <p><u>Cancer subgroup:</u> Unplanned ICU admissions of patients with cancer (breast, colorectal, gastrointestinal, lung, urogenital, CNS, leukemia, malignant lymphoma, or other malignancy) as APACHE IV reason for ICU admission.</p> <p><u>Trauma subgroup:</u> Patients with a post-operative or non-operative trauma APACHE IV reason for ICU admission.</p>

The expected mortality for the general Dutch population was assessed by using gender and age specific death risks reported by the Dutch governmental institution called Statistics Netherlands (15). To assess these specific death risks, the ICU patients included in the analysis were categorized in 7 age groups (i.e. <40, 40-50, 50-60, 60-70, 70-80, 80-90, >90 years) in which the percentages of females and males were calculated. According to the percentages of patients in each age group and the corresponding percentage of females/males, the weighted average of the death risks of the general Dutch population was assessed. The weighted average 1-year mortality risk of the general Dutch population was compared to the 1-year mortality risk of the ICU population.

In the literature the long-term mortality of ICU patients is adjusted for various determinants. The most commonly used determinants are age, severity of illness, and comorbidities (16-22). Based on an extensive list of determinants that were reported at least once in the literature we performed a 10 fold cross validation to identify the determinants that have a significant influence on the outcome of our ICU population and therefore should be included in the Cox proportional hazard model. During these analysis only the ICU patients with complete data on all determinants will be included. We compared the case-mix adjusted mortality 1-year after hospital discharge between the ICU subgroups by calculating the adjusted hazard ratios (HR_{adj}) and corresponding 95% confidence intervals (CI) for the ICU subgroups with the whole ICU population (excluding the subgroup of interest) as reference.

All statistical analyses were performed using PASW statistics 18 and SAS 9.2.

Results

Data

From January 1st, 2007 to October 1st, 2010, 149,566 patients not discharged to another hospital and fulfilling the APACHE IV inclusion criteria (9) were discharged from one of the 81 Dutch ICUs included in the study. All participating ICUs are mixed medical-surgical units located in university hospitals (n=7), teaching hospitals (n=27) or non-teaching hospitals (n=47). Of the 149,566 records, 108,295 (72.4%) records could be linked with the insurance claims database of Vektis. Table 6.1 shows the demographics of the linked and non-linked records, showing that the non-linked records have a higher proportion of elective surgery (especially cardiac surgical) patients and subsequently a lower in-hospital mortality. Of the 108,295 patients, 11,225 (10.4%) patients died on the ICU and another 5,867 (5.4%) patients died in the hospital after ICU discharge, resulting in a total of 17,092 (15.8%) in-hospital deaths. The remaining 91,203 patients who survived hospitalization were included in the analysis of mortality after hospital discharge. Of these included patients 17,113; 25,236; 28,736; and 20,118 patients received ICU care in the year 2007, 2008, 2009, and 2010 respectively. Table 6.2 shows the in-hospital mortality of the total ICU population and of the diagnostic subgroups and shows the characteristics of the ICU

patients that survived hospitalization in which the mortality after hospital discharge is assessed.

Table 6.1: Demographics of the patients in the linked and non-linked NICE dataset.

	Linked NICE dataset	Non-linked NICE dataset
Number of admissions	108,295	41,271
IC mortality (%)	10.4	6.8*
In-hospital mortality %	15.8	10.9*
Male %	58.6	61.1*
Admission type %		
Medical	40.3	29.8*
Urgent surgery	15.5	13.2*
Elective surgery	44.2	57.0*
Age mean (sd)	63.6 (15.9)	62.9 (15.8)*
APACHE IV score	51 (36-72)	48 (34-65)*
Diagnostic subgroups %		
Cardiac surgery	24.2	40.3*
Severe community-acquired pneumonia	4.7	3.2*
Subarachnoid hemorrhage	1.0	1.3*
Renal	6.6	4.8*
Cancer	14.4	10.9*
Trauma	3.8	3.8

* Statistically significant difference based on a p -value < 0.05.

Table 6.2: Demographics of (survived) ICU patients in the diagnostic subgroups.

	Total	Cardiac surgery	sCAP**	SAH***	Renal	Cancer	Trauma
Number of patients	108,295	26,234	5,102	1,057	7,167	15,625	4,135
ICU mortality %	10.4	3.2	19.4	26.1	38.9	3.4	7.7
In-hospital mortality %	15.8	5.0	28.2	30.9	48.5	7.8	11.6
Number of patients in survival analyses	91,203	24,911	3,662	730	3,692	14,405	3,654
ICU length of stay in days median (25%-75%)	0.98 (0.78-2.35)	0.92 (0.79-1.73)	3.71 (1.58-7.80)	3.25 (0.99-7.96)	3.83 (1.66-9.38)	0.92 (0.80-1.59)	1.07 (0.73-2.82)
Hospital length of stay in days median (25%-75%)	10.1 (6.0-19.0)	8.0 (6.0-13.0)	14.0 (9.0-23.0)	18.0 (11.0-28.0)	20.6 (12.0-37.0)	12.0 (8.0-19.0)	11.0 (6.0-21.0)
Male %	58.9	71.6	59.1	33.4	58.7	58.6	65.1
One or more chronic diagnose* %	24.7	23.4	49.8	8.2	38.0	19.4	12.3
Admission type %							
Medical	35.2	0.0	98.0	68.8	67.8	2.2	52.0
Urgent surgery	14.7	10.9	1.1	15.2	19.3	6.0	34.6
Elective surgery	50.1	89.1	0.9	16.0	12.9	91.8	13.4
Age mean (sd)	62.3 (16.1)	66.7 (10.9)	63.5 (15.0)	55.7 (12.9)	65.6 (14.3)	65.6 (12.4)	53.7 (22.9)
APACHE IV score median (25-75%)	47 (34-63)	46 (36-57)	65 (50-81)	37 (27-52)	79 (64-99)	43 (33-54)	40 (26-57)

* chronic diagnoses are diabetes, COPD, cirrhosis or respirator insufficiency.

sCAP: severe community-acquired pneumonia. *SAH: subarachnoid hemorrhage

Mortality after hospital discharge

At 1, 2, and 3 years after hospital discharge respectively 21,128 (23.2%), 49,737 (54.5%), and 74,825 (82.0%) patients were censored because they were not yet discharged long enough from the hospital at the follow-up endpoint. Thus, 70,075 (76.8%) patients could be followed for 365 days after hospital discharge. Table 6.3 shows the demographics of censored and non-censored patients after 1 year of follow-up, showing small differences.

Table 6.3: Demographics of the censored and non-censored patients.

	Non-censored patients	Censored patients
Number of admissions	70,075	21,128
Male %	59.1	58.2*
Admission type %		
Medical	34.3	38.2*
Urgent surgery	14.5	15.6*
Elective surgery	51.2	46.2*
Age mean (SD)	62.4 (16.0)	62.2 (16.3)
APACHE IV score median (25-75%)	47 (34-62)	48 (34-64)*

* Statistically significant difference based on a p -value < 0.05

The observed mortality 1, 2, and 3 year after hospital discharge in the total ICU population was 12.5%, 19.3%, and 27.5% respectively. The 1, 2, and 3 year mortality of the general Dutch population according to the weighted average of the gender and age specific death risks was 2.4%, 5.1% and 8.2% respectively (15). In figures 6.1 and 6.2 the unadjusted mortality after hospital discharge of the general Dutch population, the total ICU population and of the ICU subgroups are shown in Kaplan-Meier curves. Medical admissions have a higher mortality after hospital discharge compared to surgical admissions. Among the different diagnostic subgroups the patients admitted for acute renal failure have the highest 1-year mortality after hospital discharge. However, the patients admitted for cancer have the highest 3-year mortality after hospital discharge. The highest risk of death after hospital discharge appears to be in the first three months. This phenomenon is emphasized in patients with subarachnoid hemorrhage in which the mortality risk three months after hospital discharge is 5.4% and the additional mortality in the subsequent 9 months is only 2.6%.

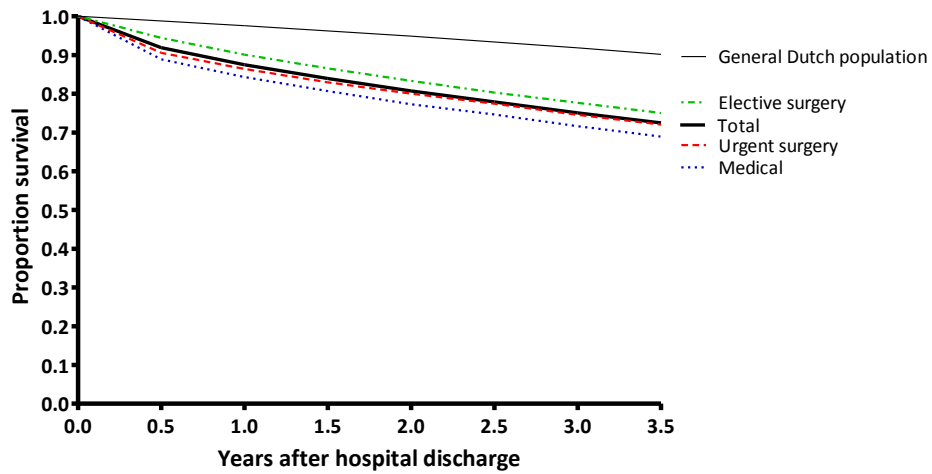


Figure 6.1: Kaplan Meier curves for total ICU patients and subgroups based on admission type

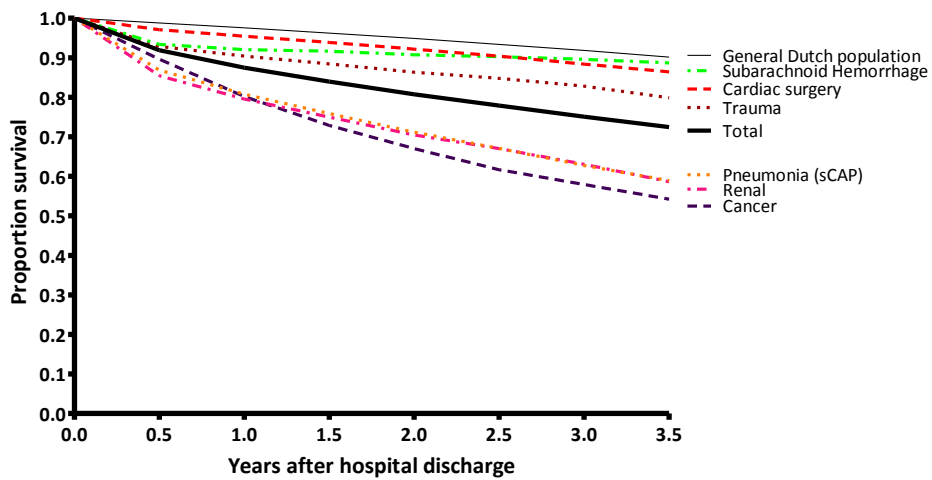


Figure 6.2: Kaplan-Meier curves of total ICU population and subgroups based on admission diagnose

Most of the ICU subgroups have an increased mortality risk during the 3-years after hospital discharge compared to the general Dutch population. However, six months after hospital discharge the diagnostic subgroups cardiac surgery and trauma have a comparable mortality risk and subarachnoid hemorrhage seems to have a decreased mortality risk compared to the general Dutch population. Strikingly, the survival curves for patients with sCAP, cancer and acute renal failure continue to diverge from the general Dutch population. However, these crude survival figures are not adjusted for case-mix.

The Cox proportional hazard model which is used to calculate the case-mix adjusted mortality 1-year after hospital discharge included the following determinants: age, gender, mechanical ventilation during the first 24 hours of ICU admission, length of hospital stay, physiological condition expressed as the acute physiology score (APS) according to the APACHE IV model, the co-morbidities diabetes, COPD, cirrhosis, and chronic respiratory insufficiency (oxygen at home, positive pressure ventilation at home, the so-called “NYHA IV of respiratory disease”), year of ICU admission, platelets, INR, hematocrit, and discharge destination. During the analysis 59,157 ICU patients with complete data on all determinants could be included. The cox model showed that particularly the presence of cirrhosis and chronic respiratory insufficiency leads to a statistically significant higher 1-year mortality after hospital discharge. In figure 6.3 the adjusted hazard ratios for the 1-year mortality after hospital discharge of the subgroups are compared to the whole ICU population, excluding the patients in the subgroup of interest as the reference group. This figure shows that the elective surgical and cardiac surgical patients have statistically significantly better mortality outcomes (HR_{adj} 0.73 and 0.28 respectively) than the rest of ICU patients. In contrast, medical patients and patients admitted for cancer have statistically significantly worse mortality outcomes (HR_{adj} 1.41, 1.94 respectively) compared to the rest.

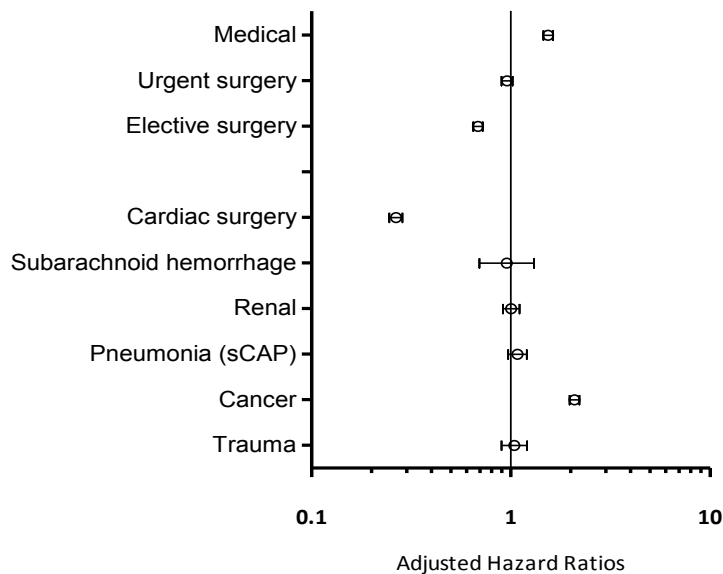


Figure 6.3: Adjusted Hazard Ratio (HR_{adj}) for the 1-year mortality after hospital discharge for different subgroups compared to the total ICU population excluding the subgroup of interest as reference group.

Discussion

This study shows that the 3-year mortality risk after hospital discharge in ICU patients is higher than the mortality risk of the weighted average of gender and age specific death risks of the general Dutch population (27.5% versus 9.8%), indicating that ICU patients who survived the hospital still have an increased risk of dying in the subsequent years. However, the mortality risk after hospital discharge differs among ICU subgroups. The cardiac surgery, subarachnoid hemorrhage, and trauma patients have a relative lower unadjusted observed mortality after hospital discharge compared to the patients with sCAP, acute renal failure and cancer. Generally, the highest post-hospital mortality of ICU patients is in the first three months after hospital discharge. Since we excluded the patients that were discharged to other hospitals, the higher mortality risk in the first three months cannot be explained by the in-hospital mortality of patients that were transferred to another hospital. The mortality of ICU patients after hospital discharge can be at least partially explained by the additional mortality associated with the ICU admission or by the pre-existing co-morbidities. Not unexpectedly, the Cox proportional hazard model showed that co-morbidities have statistically significant influence on mortality after hospital discharge. This is consistent with findings of Azoulay *et al.* (3). The cardiac surgery, subarachnoid hemorrhage, and trauma patients probably have less co-morbidities compared to the sCAP, acute renal failure and cancer patients partly explaining the differences found in the unadjusted observed mortality after hospital discharge of these diagnostic subgroups.

These results were not unexpected as previous publications already showed that the unadjusted mortality after hospital discharge is substantial. Our results are similar to the study by Keenan *et al.* in which the unadjusted 1-year mortality after hospital discharge in a Canadian ICU population (n=27,103) was 10.9% (23). In a study by Iribarren-Diarasari *et al.* the unadjusted 1-year mortality after hospital discharge in a Spanish ICU population was higher than in our study, namely 21.2% (24). However, the sample size of that study was small (n=283) and might be not generalizable to the general Spanish ICU population. In a study by Williams *et al.* the unadjusted 1-year mortality after hospital discharge in an Australian population (n=19,921) was lower, namely 5.4% (21). This apparent difference can, at least partially, be explained by the smaller proportion of patients admitted after cardiac surgery in our study (27%) compared to the Australian study (44%). However, the overall 1-year mortality of the total ICU population is an oversimplification. There are huge apparent differences between ICU subgroups. We defined specific ICU subgroups of which various 1-year mortalities have been reported with either higher (9,10,12) or lower (11,14) mortality risks than the mortality risks we have found for those subgroups. However, the direct comparison of different studies investigating long-term mortality of ICU patients is difficult. Firstly, the case-mix of ICU populations may markedly differ, and that explains

some of the differences between the studies. This is illustrated by the subgroups of patients with sCAP and acute renal failure. Their unadjusted 1-year mortality is impressively increased in comparison to the general ICU-population (see figure 6.2). However, after adjustment for several determinants the hazard ratio of these subgroups is no longer statistically significantly different from the total ICU population (see figure 6.3). Therefore, direct comparison of Kaplan-Meier curves without adjustment may lead to wrong conclusions. Second, the starting-point of the follow-up (ICU admission, ICU discharge or hospital discharge) and its end-point differs among the previously reported studies. We have chosen to focus on the outcome after hospital discharge and used the hospital discharge as the starting-point of follow-up as the short-term outcome of ICU patients (e.g. ICU and hospital mortality) is already extensively described.

Although we used a large dataset, this study is subject to some limitations. In our data the reasons of death are not known and might be unrelated to the reason of ICU admission. However, if the cause of death is unrelated to the former ICU admission it is likely that these cases are evenly distributed between the ICU and non-ICU patients and thus have no effect on our conclusions concerning the difference between ICU subgroups and between the ICU and the general Dutch population. In this study the patients that were readmitted to the ICU during the same hospitalization period were excluded. However, patients that were admitted more than once to the ICU during different hospitalization periods were not excluded. Furthermore, in this study 28.6% of the NICE registry records could not be linked with the insurance claims database. This could lead to some bias in the selection of included patients (table 1). The non-linked records mainly concerned the cardiac surgery patients, suggesting that the overall mortality of the total ICU population might have been lower if these patients could have been included. We used deterministic linkage which overall produces a low number of false positive links (25), meaning that the linked dataset is reliable. Vice versa, it also explains the rather high percentage non-linked records as deterministic linkage can miss matches due to errors in the linking variables (false negative links). Furthermore, the insurance claims database of Vektis covers 95% of the Dutch insured population, meaning that 5% of the insured admitted ICU patients may not be present in the insurance claims database. The non-linked patients are predominantly cardiac surgery patients and the amount of patients with sCAP, acute renal failure and cancer are somewhat lower in the non-linked group. It is, therefore, safe to assume that inclusion of these patients would decrease the overall 1-year mortality and strengthen the already found differences between the subgroups with higher mortality risks. Of course, the linkage could be greatly improved if a social security code of the patients would have been registered in both databases. Unfortunately, such identifier was not available in the Netherlands during the study period due to existing Dutch privacy protection rules. In some Scandinavian countries, however, the social security code is already being used to assess long-term mortality (26). Yet, the strength of this

observational study is its enormous size and our ability to correct for various known determinants. We were able to show that the excess of mortality in certain subgroups was more related to co-morbidities than to the direct influence of the diseases themselves.

Conclusion

In the general ICU population the mortality after hospital discharge is substantial and much higher than the weighted average of gender and age specific death risks of the general Dutch population. The mortality after hospital discharge differs widely between ICU subgroups, though the highest risk of death after hospital discharge occurs in the first three months. Reporting the ubiquitously used in-hospital mortality may hence lead to an important underestimation of the true mortality of ICU patients. Future studies should focus on the analysis of mortality after hospital discharge that is attributable to the former ICU admission and on ways to improve long-term mortality.

References

1. de Vos M, Graafmans W, Keesman E, et al: Quality measurement at intensive care units: which indicators should we use? *J Crit Care* 2007; 22:267-274
2. Pronovost PJ, Berenholtz SM, Ngo K, et al: Developing and pilot testing quality indicators in the intensive care unit. *J Crit Care* 2003; 18:145-155
3. Azoulay E, Adrie C, de Lassence A, et al: Determinants of postintensive care unit mortality: a prospective multicenter study. *Crit Care Med* 2003; 31:428-432
4. Angus DC, Clermont G, Kramer DJ, et al: Short-term and long-term outcome prediction with the Acute Physiology and Chronic Health Evaluation II system after orthotopic liver transplantation. *Crit Care Med* 2000; 28:150-156
5. Dutch National Intensive Care Evaluation (NICE) foundation: Available online at: <http://www.stichting-nice.nl>. Last access at Juli 24,2012
6. Vektis: Available online at: <http://www.vektis.nl>. Last access at Juli 24,2012
7. Zimmerman JE, Kramer AA, McNair DS, et al: Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34:1297-1310
8. Roos LL, Wajda A: Record linkage strategies. Part I: Estimating information and evaluating approaches. *Methods Inf Med* 1991; 30:117-123
9. Hein OV, Birnbaum J, Wernecke KD, et al: Three-year survival after four major post-cardiac operative complications. *Crit Care Med* 2006; 34:2729-2737
10. Kolh P, Kerzmann A, Honore C, et al: Aortic valve surgery in octogenarians: predictive factors for operative and long-term results. *Eur J Cardiothorac Surg* 2007; 31:600-606
11. Kroschinsky F, Weise M, Illmer T, et al: Outcome and prognostic features of intensive care unit treatment in patients with hematological malignancies. *Intensive Care Med* 2002; 28:1294-1300
12. Rabbat A, Chaoui D, Montani D, et al: Prognosis of patients with acute myeloid leukaemia admitted to intensive care. *Br J Haematol* 2005; 129:350-357
13. Ulvik A, Kvale R, Wentzel-Larsen T, et al: Multiple organ failure after trauma affects even long-term survival and functional status. *Crit Care* 2007; 11:R95
14. Ulvik A, Wentzel-Larsen T, Flaatten H: Trauma patients in the intensive care unit: short- and long-term survival and predictors of 30-day mortality. *Acta Anaesthesiol Scand* 2007; 51:171-177

15. Centraal Bureau voor de Statistiek (CBS): Available online at: <http://statline.cbs.nl/StatWeb>. Last access at Juli 24,2012
16. Chelluri L, Im KA, Belle SH, et al: Long-term mortality and quality of life after prolonged mechanical ventilation. *Crit Care Med* 2004; 32:61-69
17. Engoren M, Arslanian-Engoren C: Long-term survival in the intensive care unit after erythrocyte blood transfusion. *Am J Crit Care* 2009; 18:124-131
18. Grander W, Dunser M, Stollenwerk B, et al: CRP Levels and Post-Intensive Care Unit Mortality in Non-Surgical Intensive Care Patients. *Chest* 2010;
19. Ho KM, Knuiman M, Finn J, et al: Estimating long-term survival of critically ill patients: the PREDICT model. *PLoS One* 2008; 3:e3226
20. Schneider CP, Fertmann J, Geiger S, et al: Long-term survival after surgical critical illness: the impact of prolonged preceding organ support therapy. *Ann Surg* 2010; 251:1145-1153
21. Williams TA, Dobb GJ, Finn JC, et al: Determinants of long-term survival after intensive care. *Crit Care Med* 2008; 36:1523-1530
22. Williams TA, Ho KM, Dobb GJ, et al: Effect of length of stay in intensive care unit on hospital and long-term mortality of critically ill adult patients. *Br J Anaesth* 2010; 104:459-464
23. Keenan SP, Dodek P, Chan K, et al: Intensive care unit admission has minimal impact on long-term mortality. *Crit Care Med* 2002; 30:501-507
24. Iribarren-Diarasari S, Aizpuru-Barandiaran F, Munoz-Martinez T, et al: Health-related quality of life as a prognostic factor of survival in critically ill patients. *Intensive Care Med* 2009; 35:833-839
25. Tromp M, Ravelli AC, Bonsel GJ, et al: Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol* 2011; 64:565-572
26. Niskanen M, Kari A, Halonen P: Five-year survival after intensive care--comparison of 12,180 patients with the general population. Finnish ICU Study Group. *Crit Care Med* 1996; 24:1962-1967

Chapter 7

Prediction of long-term mortality in ICU patients: Model validation and assessing the effect of using the in-hospital versus long-term mortality on benchmarking

Sylvia Brinkman, Ameen Abu-Hanna, Evert de Jonge,
Nicolette F. de Keizer

Submitted



Abstract

Objectives: To analyze whether the Acute Physiology And Chronic Health Evaluation (APACHE) IV model can be used for reliable mortality prediction at fixed time points (3, 6, 12 months) after ICU admission and to analyze the influence of using these endpoints on benchmarking.

Design: Cohort study of ICU admissions from a national ICU registry linked to administrative records including long-term outcome from an insurance claims database.

Setting: 44 Dutch mixed ICUs.

Patients: 83,555 ICU patients admitted between January 1st 2008 and October 1st 2011.

Interventions: None

Measurements and Main Results: Four APACHE IV models were customized to predict the in-hospital mortality and the mortality at 3, 6, and 12 months after ICU admission. Their performance was assessed by bootstrapping. The APACHE IV model's performance was best when applied to the prediction of the in-hospital mortality (Area Under the receiver operating characteristic Curve (AUC)=0.88, Brier score=0.09) and declined slightly over time reaching an AUC of 0.82 and a Brier score of 0.14 when it was applied for mortality prediction 12 months after ICU admission. The SMR of 7 ICUs significantly increased and the SMR of 7 ICUs significantly decreased when using the mortality 3 months after ICU admission instead of the in-hospital mortality.

Conclusions: The customized APACHE IV model can be used for the prediction of the in-hospital mortality as well as for the mortality at 3, 6, and 12 months after ICU admission. The endpoint chosen influences the SMR of ICUs suggesting that the in-hospital mortality might be influenced by differences in discharge policies. Therefore, the SMR based on mortality at a longer term fixed time point after ICU admission should preferably be used as quality indicator for benchmarking purposes. Further research should focus on which follow-up period gives the best insight in the true quality of ICU care.

Introduction

The increasing use of quality indicators to quantify process and patient outcomes is important in the health care debate, especially in complex and expensive environments such as in intensive care units (ICU) (1,2). Currently, the in-hospital mortality is the most commonly used outcome for ICU patients. As patients' case-mix is highly correlated with mortality risk, the in-hospital mortality has to be corrected for case-mix differences before it can be applied for quality of care assessment and for benchmarking of ICUs. The Standardized Mortality Ratio (SMR), which is the ratio of the observed and the case-mix adjusted predicted in-hospital mortality, is frequently used as a quality indicator, where $SMR=1$ serves as a reference point for "average" quality among the ICUs. Several prognostic models have been developed for predicting case-mix adjusted in-hospital mortality risk, such as the most recently developed Acute Physiology And Chronic Health Evaluation (APACHE) IV model (1). However, this case-mix adjusted in-hospital mortality ratio may still be influenced by other factors for which the prognostic models do not adjust such as differences in discharge policies. For instance hospitals that have a relatively high percentage of patients discharged to another hospital or to a palliative care facility might have a relatively low observed in-hospital mortality, even after adjusting for patient case-mix differences. Therefore, comparison among ICUs is more fair when based on the SMR using mortality on a longer term fixed time point after ICU admission (e.g. 3 months) which is not influenced by discharge policies. Linking clinical databases with administrative databases, e.g. from insurance companies, offers the opportunity to easily assess the long-term outcomes (2). It is, however, unclear whether current prognostic models for prediction of the case-mix adjusted in-hospital mortality, when customized, are suitable for predicting the long-term mortality.

The aim of this study is to analyze whether the APACHE IV model can be used to reliably predict the mortality at fixed time points after ICU admission and to analyze the influence of using the mortality at 3, 6, or 12 months after ICU admission instead of the ubiquitously used in-hospital mortality on the SMR rank position of Dutch ICUs.

Method

This study used a dataset of ICU patients admitted between January 1st, 2008 and October 1st, 2011. Data from the Dutch National Intensive Care Evaluation (NICE) registry (3) were linked to a national administrative database of health insurance companies (insurance claims database of Vektis) (4) containing the state of the patients (either alive or death) on January 1st 2012, and if applicable the date of death. The linkage of these databases is more extensively described elsewhere (5). To avoid biased results due to incomplete

linkage we only included the ICUs of which the data could be linked for more than 80% of the ICU admissions.

The NICE registry contains demographic, physiological and clinical data of all consecutive ICU patients admitted to participating ICUs, including all variables of the APACHE IV model (1). During the study period, approximately 90% of all Dutch ICUs recorded data of all their admissions in the NICE registry. The 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. The NICE initiative is officially registered according to the Dutch Personal Data Protection Act.

As we are interested in the performance of the APACHE IV model to predict the long-term mortality, we only included ICU admissions fulfilling the inclusion criteria of the APACHE IV model (1). The APACHE IV model uses a distinct set of predictor variables to estimate the probability of the in-hospital mortality for patients admitted after coronary artery bypass grafting (CABG model) which differ from the set of predictors used for patients admitted for other reasons (non-CABG model). In this study the APACHE IV non-CABG model will be used for long-term risk prediction of the in-hospital mortality and the mortality at 3, 6, or 12 months after ICU admission. Before the APACHE IV model is applied for risk prediction the model is customized on the Dutch non-CABG ICU patients. For this study, four customized APACHE IV models were developed for in-hospital mortality, and mortality at 3, 6, or 12 months after ICU admission as the dependent variables in four logistic regression models. In each model the independent variables are the predictors included in the original APACHE IV model (i.e. age, admission type, reason for ICU admission, acute physiology score, acute renal failure, COPD, cirrhosis, AIDS, immunological insufficient, metastatic neoplasm, and hematological malignancy). The 445 APACHE IV reasons for ICU admission are categorized in 118 groups based on their weight in the original APACHE IV model. The performance of the four customized models is assessed by measures of discrimination, calibration, and accuracy. To measure the discrimination of the models, the Area Under the Receiver Operating Characteristic Curve (AUC) was used (6). The Brier score (7) was used to assess overall accuracy. Calibration plots of the customized models are used to evaluate the models' calibration. These performance measures are more extensively described elsewhere (8) but it is important to note that a higher AUC and a lower Brier score indicate better performance. To assess whether the APACHE IV model can be used to reliably predict the mortality at fixed time points after ICU admission, we used the following criteria; the AUC should be higher than 0.80 (9), the Brier score should be smaller than 0.25 (10) and the calibration plots should not show large deviation. The performance of the four customized APACHE IV models and the confidence interval around them were calculated using the bootstrap method (11). In this method each APACHE IV model was customized on each of 1000 bootstrap samples and the

performance was assessed on the original dataset. To compare the performance of each of the three models based on the long-term mortality with that of the in-hospital mortality model, the performance differences (in AUCs and Brier scores) between the models on all the 1000 bootstrap samples were obtained. When the 95% confidence interval of these differences did not contain zero the differences were declared statistically significant (at the 0.05 level).

The influence of using the mortality at 3, 6, or 12 months after ICU admission instead of the in-hospital mortality on the SMR of the included ICUs was assessed by calculating the SMRs (i.e. the in-hospital, 3, 6, and 12 months after ICU admission SMR) for each ICU. The SMRs were calculated in 1000 bootstrap samples of which the median SMR values along with their 95% empirical confidence intervals for each ICU were obtained. An SMR below 1 implies less observed deaths than expected (i.e. better performance) by the prognostic model and a SMR above 1 implies more observed deaths than expected by the prognostic model (i.e. lower performance). A difference between a SMR based on the in-hospital mortality and the SMRs based on the mortality at 3, 6, or 12 months after ICU admission was considered statistically significant if the 95% confidence interval of the difference did not contain zero.

All statistical analyses were performed using the statistical environment R version 2.14.1 and PASW statistics 18.

Results

From January 1st, 2008 to October 1st, 2011, 87,192 non-CABG patients fulfilling the APACHE IV inclusion criteria (9) were admitted to one of the 44 Dutch ICUs included in the study. All participating ICUs were mixed medical-surgical units located in university hospitals (n=3), teaching hospitals (n=14) or non-teaching hospitals (n=29). In total 83,555 (95.8%) patients could be linked to the insurance claims database and were included in the analyses. Table 7.1 shows the demographics of the non-linked and linked ICU population. In the linked ICU population, which is included in the analysis, respectively 0, 5,862, and 18,556 patients were censored at 3, 6, and 12 months after ICU admission as their follow-up period after ICU admission was not long enough.

In table 7.2 the performance of the four customized APACHE IV models is given. The performance of the APACHE IV model is best for the prediction of the in-hospital mortality (AUC=0.88, Brier score=0.09). The performance declines slightly over time with an AUC of 0.82 and a Brier score of 0.14 when it was applied for mortality prediction 12 months after ICU admission. Also the calibration plots show slightly more deviation from the diagonal line in the high risk group when the APACHE IV model is used for the long-term mortality (Figure 7.1).

Table 7.1: Demographics of the linked and non-linked non-CABG ICU admissions

	Linked ICU population	Non-linked ICU population
Number of ICU admissions	83,555	3,637
In-hospital mortality %	15.7	11.6
3 months after ICU admission mortality %	20.0	Not applicable
6 months after ICU admission mortality %	22.9	Not applicable
12 months after ICU admission mortality %	26.6	Not applicable
Hospital length of stay in days median (25%-75%)	10.0 (5.0-20.0)	8.0 (3.0-15.0)
ICU length of stay in days median (25%-75%)	1.1 (0.8-3.1)	0.9 (0.6-2.0)
Age, mean (sd)	63.5 (16.4)	56.3 (19.4)
Male %	56.3	59.5
Ventilated %	44.0	42.1
One or more chronic diagnose* %	33.5	23.2
Admission type %		
Medical	48.9	52.3
Urgent surgery	16.0	18.0
Elective surgery	35.1	29.7
APACHE IV score median (25-75%)	54.0 (37.0-77.0)	46.0 (29.0-69.0)
APACHE IV probability median (25-75%)	8.1 (2.6-26.9)	4.6 (1.5-17.9)
ICU discharge destination		
Destination in the same hospital %	90.0	85.9
Range over the hospitals	(80.5-100.0)	(55.2-100.0)
Destination outside the hospital %	8.0	11.7
Range over the hospitals	(0.0-19.5)	(0.0-44.8)
Destination not reported %	2.0	2.4

* chronic diagnoses are immunological insufficiency, AIDS, hematological malignancy, metastatic neoplasm, cirrhosis, cardiovascular insufficiency, respirator insufficiency, COPD, chronic dialysis, renal insufficiency

Table 7.2: Performance of the four customized APACHE IV models

APACHE IV model	AUC	Brier score
In-hospital mortality	0.88 (0.88-0.88)	0.09 (0.09-0.09)
mortality 3 months after ICU admission	0.85 (0.85-0.85)*	0.11 (0.11-0.11)*
mortality 6 months after ICU admission	0.84 (0.84-0.84)*	0.13 (0.13-0.13)*
mortality 12 months after ICU admission	0.82 (0.82-0.82)*	0.14 (0.14-0.14)*

* statistically significant difference when compared to the in-hospital mortality model.

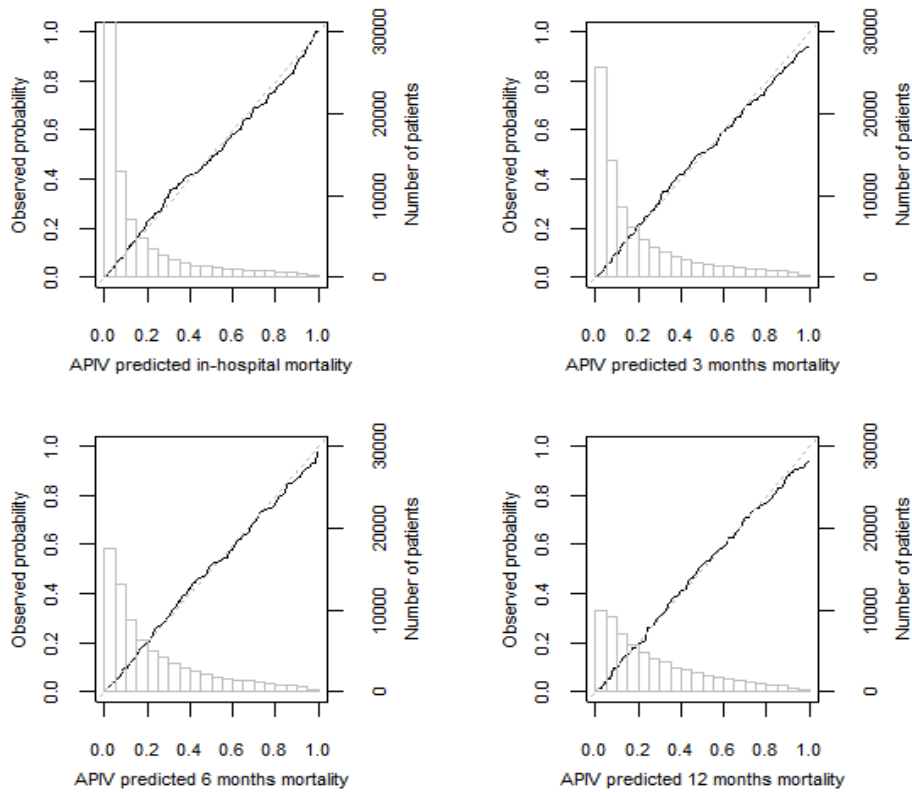


Figure 7.1: Calibration plots of the APACHE IV model customized on the in-hospital mortality and the mortality at 3, 6, and 12 months after ICU admission. The histograms show the distribution of the number of patients per predicted risk group.

The effect of benchmarking on in-hospital mortality or long-term mortality is assessed in the 44 included ICUs. The SMR of 7 ICUs (i.e. ICU 1, 3, 7, 8, 14, 19, and 31) significantly increased (i.e. lower performance) and the SMR of 7 ICUs (i.e. ICU 21, 25, 30, 38, 40, 41, and 43) significantly decreased (i.e. better performance) when using the 3 months after ICU admission SMR instead of the in-hospital SMR. Not only the SMR but also the SMR rank position of ICUs changed when using the in-hospital mortality or the mortality on a fixed time point after ICU admission (Figure 7.2). For instance ICU 8 has been ranked at position 8 (with a 95% confidence interval of 1–26) when the in-hospital mortality was used and was declined (i.e. lower performance) to rank position 33 (with a 95% confidence interval of 11–43) when the mortality at 3 months after ICU admission was used. In contrast, ICU 41 has been ranked at position 41 (33–44) when the in-hospital mortality was used and improved (i.e. better performance) to rank position 21 (13–31) when the mortality at 3 months after ICU admission was used.

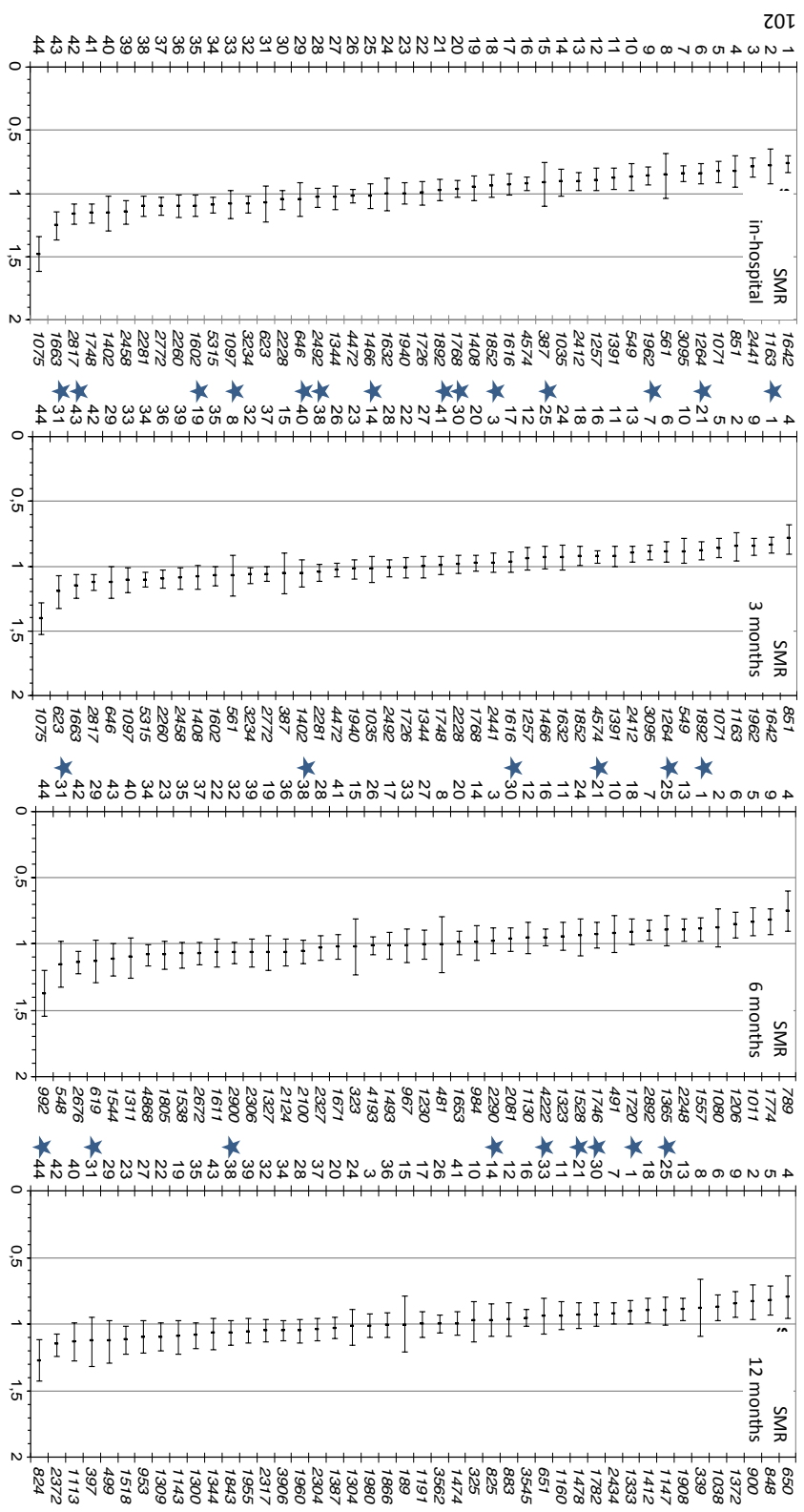


Figure 7.2: The median SMR and the associated confidence interval (CI) for each ICU according to the APACHE IV model customized for the in-hospital mortality and the mortality at 3, 6, and 12 months after ICU admission. The right y axis of each figure represents the number of ICU admissions, and the left y axis represents the ICU number, an asterix indicates that the SMR is statistically significantly changed in comparison to the in-hospital SMR

Discussion

This study shows that the customized APACHE IV model, originally developed for the prediction of the in-hospital mortality of ICU patients still performs well when applied for mortality prediction at 3, 6, and 12 months after ICU admission. The performance of the APACHE IV model is best when it is applied for the prediction of the in-hospital mortality, which is expected as the original model is specifically developed for the in-hospital mortality prediction and as short-term prediction is easier than long-term prediction. However, the APACHE IV model still performs well when it is applied for long-term mortality prediction.

We used calibration plots to assess the calibration of the prognostic models and not the frequently used Hosmer-Lemeshow \hat{C} -statistic (12) as this statistic has serious limitations when it is used for comparison of different models. Zhu *et al.* (13) showed that the Hosmer-Lemeshow \hat{C} -statistic is sensitive to the sample size of a population and to arbitrarily small changes in the cutoff points. In general, models validated in a large sample and/or with the ability to provide very low (or high) risks will tend to have a worse Hosmer-Lemeshow \hat{C} -statistic than a model validated in a small sample size and/or providing a wider range of probabilities.

Previous studies have focused on the development of new prognostic models for long-term mortality predictions (14,15). The SUPPORT model developed by Knaus *et al.* (15) predict the 6 months mortality after hospitalization for seriously ill adult patients (also non-ICU patients) and had an AUC of 0.82 which is comparable with our result of 0.84. The PREDICT model developed by Ho *et al.* (14) predicts the long-term mortality of ICU patients up to 15 years after ICU admission and had an AUC of 0.76.

We have shown that the SMR of ICUs is influenced by the duration of follow-up (e.g. in-hospital mortality or mortality at 3, 6, or 12 months after ICU admission). Previous studies showed that the use of different prognostic models can have an influence on the SMR rank position (16,17). This can also partly explain the differences in SMR rank position found in our study as the used customized APACHE IV models can slightly differ due to difference in the weight of the included variables. Our results suggest that the ubiquitously used case-mix adjusted in-hospital mortality may well be influenced by other factors such as differences in discharge policies. On average 8.0% of the ICU patients are discharged to a medical facility. In ICU 41, which showed the largest improvement in rank position 3 months after ICU admission (i.e. better performance), 4.6% of the admission were discharged to a facility outside the hospital. In contrast, ICU 8 showing the largest (though not significant) decline in rank position 3 months after ICU admission (i.e. lower performance), 18.1% of the admissions were discharged to a facility outside the hospital. These results suggest that the case-mix adjusted in-hospital mortality might still be influenced by discharge policies. Therefore, the SMR based on case-mix adjusted mortality

at a longer term fixed time point after ICU admission should preferably be used as quality indicator for benchmarking purposes. Further research might focus on which fixed time point after ICU admission gives the best insight in the true quality of care at ICUs.

The strength of our study is that we used the existing APACHE IV model which is already available in many hospitals and already has been externally validated to specific populations such as the Dutch ICU population (8). Although we used a large dataset, this study is subject to some limitations. In our data the reasons for death are unknown and might be unrelated to the reason of ICU admission. Furthermore, in this study we used data of ICUs of which at least 80% of the admissions could be linked to the assurance claims database to avoid selection biased. We used deterministic linkage which overall produces a very low number of false links (18). At the same time, deterministic linkage can miss matches due to errors in the linking variables (false negative links), explaining that 4.2% of the admissions could not be linked with the insurance claims database. In the patient populations in which the 3, 6, and 12 months after ICU admission could be assessed respectively 4.2, 4.8, and 4.0% of the records were non-linked. We believe it is safe to assume that the potential bias due to non-linkage most likely influence the performance of the four customized models evenly and will have no influence on our conclusions.

Conclusion

The customized APACHE IV model can be used for the reliable prediction of the in-hospital mortality of ICU patients as well as the mortality at 3, 6, and 12 months after ICU admission. The SMR of ICUs is influenced by the duration of follow-up (e.g. in-hospital mortality or mortality at 3, 6, or 12 months after ICU admission) suggesting that the ubiquitously used case-mix adjusted in-hospital mortality is likely influenced by differences in discharge policies. Therefore, the SMR based on case-mix adjusted mortality at a longer term fixed time point after ICU admission should preferably be used as quality indicator for benchmarking purposes. Further research should focus on which fixed time point after ICU admission gives the best insight in the true quality of care.

References

1. Zimmerman JE, Kramer AA, McNair DS, et al: Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34:1297-1310
2. Brinkman S, de Jonge E, Abu-Hanna A, et al: The use of linked registries to assess long-term mortality of ICU patients. *Stud Health Technol Inform* 2012; 180:230-234
3. Dutch National Intensive Care Evaluation (NICE) foundation: Available online at: <http://www.stichting-nice.nl>, Last access at 12 december 2012.
4. Vektis: Available online at: <http://www.vektis.nl>, Last access at 12 december 2012.
5. Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, et al: Determinants of mortality after hospital discharge in intensive care patients; Literature review and Dutch cohort study. *Crit Care Med* 2012; in press
6. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29-36
7. Hilden J, Habbema JD, Bjerregaard B: The measurement of performance in probabilistic diagnosis. III. Methods based on continuous functions of the diagnostic probabilities. *Methods Inf Med* 1978; 17:238-246
8. Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, et al: External validation of Acute Physiology and Chronic Health Evaluation IV in Dutch intensive care units and comparison with Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II. *J Crit Care* 2011; 26:105-108
9. Salter K, Jutai JW, Teasell R, et al: Issues for selection of outcome measures in stroke rehabilitation: ICF Participation. *Disabil Rehabil* 2005; 27:507-528
10. Biancari F, Vasques F, Mikkola R, et al: Validation of EuroSCORE II in patients undergoing coronary artery bypass surgery. *Ann Thorac Surg* 2012; 93:1930-1935
11. Efron B: Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation . Available online at: <http://www.jstor.org/stable/2288636>, Last access at 12 december 2012..
12. Hosmer DW, Hosmer T, Le Cessie S, et al: A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997; 16:965-980
13. Zhu BP, Lemeshow S, Hosmer DW, et al: Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. *Crit Care Med* 1996; 24:57-63
14. Ho KM, Knuijman M, Finn J, et al: Estimating long-term survival of critically ill patients: the PREDICT model. *PLoS One* 2008; 3:e3226
15. Knaus WA, Harrell FE, Jr., Lynn J, et al: The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults. Study to understand prognoses and preferences for outcomes and risks of treatments. *Ann Intern Med* 1995; 122:191-203
16. Bakhshi-Raiez F, Peek N, Bosman RJ, et al: The impact of different prognostic models and their customization on institutional comparison of intensive care units*. *Crit Care Med* 2007; 35:2553-2560
17. Brinkman S, Abu-Hanna A, van der Veen A, et al: 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 standardized mortality ratios of intensive care units. *Crit Care Med* 2012; 40:373-378
18. Tromp M, Ravelli AC, Bonsel GJ, et al: Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol* 2011; 64:565-572

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 increased mortality was not statistically significant and probably caused by pre-existing comorbidities 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 of 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 \hat{C} -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 \hat{C} -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 \hat{C} -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 the 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 publically due to the increased interest in the transparency of hospital mortality. In other countries such as the UK the hospital mortality figures are already publically 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 publically 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 publically 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

1. Dutch National Intensive Care Evaluation (NICE) foundation: Available online at: <http://www.stichting-nice.nl>, last access at 12-12-2012.
2. National Medical Registration (LMR): Available online at: http://www.swov.nl/uk/research/kennisbank/inhoud/90_gegevensbronnen/inhoud/lmr.htm, last access at 12-12-2012.
3. Vektis: Available online at: <http://www.vektis.nl>, last access at 12-12-2012.
4. Le Gall JR, Lemeshow S, Saulnier F: A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270:2957-2963
5. HSMR: Available online at: <http://www.hsmr.nl/index.php/hsmr>, last access at 12-12-2012.
6. Zimmerman JE, Kramer AA, McNair DS, et al: Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34:1297-1310
7. Knaus WA, Draper EA, Wagner DP, et al: APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13:818-829
8. Lilford R, Pronovost P: Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ* 2010; 340:c2016
9. Netherlands Society of Intensive Care (NVIC) guidelines: Available online at: http://www.nvic.nl/richtlijnen_geaccordeerd.php, last access at 12-12-2012.
10. Steyerberg EW: *Clinical Prediction Models A Practical Approach to Development, Validation, and Updating*. New York, Springer, 2009
11. Zhu BP, Lemeshow S, Hosmer DW, et al: Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. *Crit Care Med* 1996; 24:57-63
12. Minne L, Eslami S, de Keizer N, et al: Effect of changes over time in the performance of a customized SAPS-II model on the quality of care assessment. *Intensive Care Med* 2012; 38:40-46

Chapter 8

13. Harrison DA, Brady AR, Parry GJ, et al: Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med* 2006; 34:1378-1388
14. Le Gall Jr, Neumann A, Hemery F, et al: Mortality prediction using SAPS II: an update for French intensive care units. *Crit Care* 2005; 9:R645-R652
15. Tromp M, Ravelli AC, Bonsel GJ, et al: Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol* 2011; 64:565-572

**Summary
&
Samenvatting**



In recent years, measuring the quality of health care and quantitative comparison of performance between health institutions has received growing attention by healthcare professionals as well as policy makers. Quality indicators, such as in-hospital mortality rates, which is believed to reflect the quality of care are frequently used to get insight in the quality of health care. However, differences in patient case-mix (i.e. specific characteristics such as diagnosis or age) influence mortality rates of hospitals. To date, several prognostic models have been developed to predict the mortality rates based on the case-mix of the patients. These mortality predictions are used to calculate the quality indicator Standardized Mortality Ratio (SMR, i.e. the ratio of the observed and predicted mortality). The SMR can be used for performance comparison across different institutions. Though, when the performance of a prognostic model is poor the case-mix adjusted mortality rates are unreliable and the quality assessment based on the prognostic model is meaningless. The quality indicator SMR is commonly based on the in-hospital mortality, however, it can be argued whether the long-term mortality is more important and therefore should be used for quality assessment.

This thesis elaborates on these topics by first addressing the performance of different prognostic models currently used for case-mix adjustment (chapter 2 and 3) and the application of a prognostic model to assess differences between the SMR of Intensive Care Unit (ICU) patients admitted during office hours and off hours (chapter 4). Second, the long-term outcomes of ICU patients is described and the effect of using the long-term mortality instead of the in-hospital mortality on the quality indicator SMR is assessed (chapter 5 to 7).

Chapter 2 described a study that compared the performance and robustness of a model based on administrative data (customized hospital standardized mortality ratio (cHSMR)) and a model based on clinical data (customized Simplified Acute Physiology Score II (cSAPS II)) in the Dutch intensive care unit population. The study used data from the clinical National Intensive Care Evaluation (NICE) registry linked to administrative records from the Dutch National Medical Registration. This linked dataset consisted of ICU patients from 55 Dutch intensive care units admitted between 2005 and 2008. The performance of both models is expressed by discrimination, accuracy, and calibration (Area Under the receiver operating characteristic Curve (AUC), Brier score, Hosmer-Lemeshow \hat{C} -statistic, and calibration plots). These performance measures showed that the cSAPS II outperforms the cHSMR in the Dutch intensive care unit population. Additionally, the robustness of the two models was assessed by simulating changes in the population's severity of illness and analyzing the effect on the SMR of ICUs. This simulation showed that the SMR of ICUs is influenced by the severity of illness in both models, but the cSAPS II model showed more robustness. This implies that comparing institutions based on SMRs can be unfavorable for those with a more severely ill intensive care unit population.

Chapter 3 focused on the performance of the most recently developed Acute Physiology and Chronic Health Evaluation (APACHE) IV prognostic model to predict the in-hospital mortality in comparison to the performance of the widely applied older prognostic models, i.e. the APACHE II and Simplified Acute Physiology Score (SAPS) II models. The study used NICE data of ICU patients from 59 Dutch ICUs admitted between 2006 and 2009. Performance measures of discrimination, accuracy, and calibration (AUC, Brier score, R-squared, Hosmer-Lemeshow \hat{C} -statistic, and calibration plots) were calculated using bootstrapping. Additionally, the standardized mortality ratios (SMRs) were calculated. The original APACHE IV showed good discrimination and accuracy but poor calibration. The overall discrimination and accuracy of the customized APACHE IV were statistically better and the overall Hosmer-Lemeshow \hat{C} -statistic was inferior to those of the customized APACHE II and SAPS II, but these differences were small in perspective of clinical use. The study showed that the three models have comparable capabilities for benchmarking purposes after customization. Main advantage of APACHE IV is the large number of reasons for admission which enable subgroup analysis. The APACHE IV coronary artery bypass grafting (CABG) model has a good performance in the Dutch ICU population and can be used to complement the three models.

Chapter 4 evaluated whether the in-hospital mortality is associated with ICU admission time. The study used NICE data of ICU patients admitted between 2002 and 2008. In this study "office hours" was defined as 08.00-22.00 during weekdays and 09.00-18.00 during weekend days. The weekend was defined as Saturday from 00.00 hours until Sunday 24.00 hours. The in-hospital mortality varied over time but was consistently higher than expected outside office hours and lower during office hours. There was no significant difference in mortality between different weekdays but in-hospital mortality increased slightly on Friday. During the weekend the in-hospital mortality was increased in comparison to the rest of the week. The study showed that the in-hospital mortality appears to be increased outside office hours and during the weekends, even when adjusted for severity of illness expressed as the APACHE II score, reason for admission, admission type, age, and gender.

Chapter 5 focused on the determinants of mortality in ICU patients after hospital discharge. In this study first a literature review to evaluate determinants of long-term mortality was conducted. Second, a Dutch cohort study to assess the influence of the found determinants on the 3, 6, and 12 months mortality after hospital discharge in the Dutch ICU population was conducted. Pubmed and EMBASE were searched on English written articles published between 1966 and 2011 during the literature review resulting in 24 articles that were included. These 24 articles were summarized on patient and study

characteristics, methods, results and determinants used for case-mix adjustment. Additionally, the quality of the included articles was assessed using a checklist for studies on long-term mortality. The determinants used for case-mix adjustment differed widely between the studies. The cohort study was conducted in ICU patients from 81 Dutch ICUs discharged alive from the hospital between 2007 and 2010. The cohort study showed that age, reason for ICU admission, and co-morbidities were associated with the mortality 3, 6, and 12 months after hospital discharge. However, the magnitude and direction of the influence of these determinants differed for the different endpoints. The most commonly used determinants in the literature were comparable to the most important determinants found in the Dutch cohort study.

Chapter 6 described the mortality risk of ICU patients after hospital discharge and compare it to mortality of the general Dutch population. The study used data from the clinical National Intensive Care Evaluation (NICE) registry linked to administrative records from an insurance claims database. This linked dataset consisted of ICU patients from 81 Dutch ICUs discharged alive between 2007 and 2010. The unadjusted observed survival was presented by Kaplan-Meier curves. Mortality risk at 1, 2, and 3-years after hospital discharge was 12.5%, 19.3%, and 27.5% respectively. The mortality after hospital discharge differs widely between ICU subgroups. The elective and cardiac surgical patients have statistically significantly better case-mix adjusted mortality outcomes, whereas medical patients and patients admitted for cancer have statistically significantly worse case-mix adjusted mortality outcomes compared to other ICU patients. This study showed that the in-hospital mortality underestimates the true mortality of ICU patients as the mortality in the first months after hospital discharge was substantial. Most ICU patients still have an increased mortality risk in the subsequent years after hospital discharge compared to the general Dutch population.

Chapter 7 evaluated whether the APACHE IV model can be used for reliable mortality prediction at fixed time points (3, 6, and 12 months) after ICU admission and analyzed the influence of using these different endpoints on benchmarking. The study used data from the clinical National Intensive Care Evaluation (NICE) registry linked to administrative records from an insurance claims database. This linked dataset consisted of ICU patients from 44 Dutch ICUs discharged alive between 2008 and 2011. Four APACHE IV models were customized to predict the in-hospital mortality and the mortality at 3, 6, and 12 months after ICU admission. Performance was described using measures of discrimination, accuracy, and calibration (AUC, Brier score, and calibration plots). Although the APACHE IV model's performance was best when applied to the prediction of the in-hospital mortality and declined slightly over time the customized APACHE IV model can be used for the prediction of mortality at 3, 6, and 12 months after ICU admission. The SMR

of 7 ICUs significantly increased and the SMR of 7 ICUs significantly decreased when using the mortality 3 months after ICU admission instead of the in-hospital mortality. This suggest that the in-hospital mortality might be influenced by differences in discharge policies. Therefore, the SMR based on mortality at a fixed longer term endpoint after ICU admission should preferably be used as quality indicator for benchmarking purposes.

In de afgelopen jaren is er steeds meer belangstelling gekomen voor het meten van en verantwoording afleggen over de kwaliteit van de gezondheidszorg. In die context worden de prestaties van verschillende zorginstellingen met elkaar vergeleken. Om inzicht te krijgen in de kwaliteit van de gezondheidszorg worden vaak kwaliteitsindicatoren zoals sterftcijfers gebruikt. Sterftcijfers worden echter beïnvloedt door verschillen in patiënten case-mix. Dat wil zeggen dat de specifieke kenmerken van de opgenomen patiënten zoals de diagnose, ernst van ziekte en leeftijd de sterftcijfers kunnen verhogen of verlagen. Om sterftcijfers tussen zorginstellingen vergelijkbaar te maken zijn er verschillende prognostische modellen ontwikkeld die de verwachte sterfte voorspellen op basis van de case-mix van de opgenomen patiënten. Deze verwachte sterfte wordt vervolgens gebruikt voor de berekening van de 'Standardized Mortality Ratio' (SMR), wat een veel gebruikte kwaliteitsindicator is. De SMR is de ratio tussen de geobserveerde ongecorrigeerde sterfte en de verwachte sterfte op basis van de case-mix. Momenteel wordt de SMR veel gebruikt om de kwaliteit tussen verschillende zorginstellingen te vergelijken, waaronder ook Intensive Cares (ICs).

Indien een prognostisch model echter onvoldoende corrigeert voor de case-mix van de opgenomen patiënten dan is de verwachte sterfte onbetrouwbaar en is de kwaliteitsbeoordeling op basis van de SMR zinloos. Daarnaast wordt tot op heden de kwaliteitsindicator SMR meestal gebaseerd op de ziekenhuissterfte. Het is echter discutabel of de lange termijn sterfte niet even belangrijk of zelfs belangrijker is dan de ziekenhuissterfte en daarom gebruikt dient te worden voor de kwaliteitsbeoordeling.

Dit proefschrift gaat dieper in op deze onderwerpen door eerst te analyseren hoe goed verschillende prognostische modellen de case-mix gecorrigeerde verwachte sterfte kunnen voorspellen (hoofdstuk 2 en 3). Vervolgens wordt met een prognostisch model geëvalueerd of er een relatie is tussen de SMR van IC patiënten die zijn opgenomen tijdens of buiten kantooruren (hoofdstuk 4). Tot slot wordt de lange termijn sterfte van IC patiënten beschreven en wordt onderzocht wat het effect op de SMR is van het gebruik van de lange termijn sterfte in plaats van ziekenhuissterfte (hoofdstuk 5 tot 7).

Hoofdstuk 2 beschrijft de validatie en vergelijking van de prestaties en robuustheid van twee prognostische modellen die de verwachte sterfte voorspellen. Het betreft het gerecalibreerde hospital standardized mortality ratio (cHSMR) model dat administratieve gegevens gebruikt voor case-mix correctie en het gerecalibreerde Simplified Acute Physiology Score (cSAPS) II model dat klinische IC gegevens gebruikt voor case-mix correctie. De studie gebruikt klinische gegevens uit de Nationale Intensive Care Evaluatie (NICE) registratie gekoppeld aan administratieve gegevens uit de Nederlandse Landelijke Medische Registratie. Deze gekoppelde dataset bestaat uit patiënten van 55 Nederlandse ICs opgenomen tussen 2005 en 2008. Om de twee modellen te valideren en te vergelijken werden maten van discriminatie, nauwkeurigheid en calibratie gebruikt. Deze maten zijn

Area Under the receiver operating characteristic Curve (AUC), Brier score, Hosmer-Lemeshow \hat{C} -statistiek en calibratie grafieken. Op basis van deze validatie maten bleek dat binnen de Nederlandse IC populatie het cSAPS II model de sterfte beter voorspelt dan het cHSMR model. De robuustheid van de twee prognostische modellen werd daarna beoordeeld door de ernst van ziekte van de IC populatie met een simulatie studie te veranderen en de invloed hiervan op de SMR te analyseren. Hieruit bleek dat bij beide modellen de SMR wordt beïnvloed door de ernst van de ziekte, maar dat het effect groter was bij het cHSMR model. Dit houdt in dat bij het vergelijken van verschillende ICs op basis van de SMR de resultaten ongunstig beïnvloed worden voor ICs met een meer ernstig zieke patiëntenpopulatie.

Hoofdstuk 3 vergelijkt het discriminerend en calibrerend vermogen en de nauwkeurigheid van het Acute Physiology and Chronic Health Evaluation (APACHE) IV prognostisch model met die van twee oudere en breed toegepaste prognostische modellen, het APACHE II en SAPS II model. De studie maakte gebruik van gegevens uit de NICE registratie van patiënten van 59 Nederlandse ICs opgenomen tussen 2006 en 2009. De SMR werd vervolgens op basis van de verschillende modellen berekend. Het originele (niet gerecalibreerde) APACHE IV model had een goede discriminatie en nauwkeurigheid, maar een slechte calibratie. De discriminatie en nauwkeurigheid van het gerecalibreerde APACHE IV model was statistisch significant beter maar de calibratie was slechter dan die van de gerecalibreerde APACHE II en SAPS II modellen. De validatie verschillen waren echter klein en klinisch gezien niet relevant. De studie toonde aan dat de validiteit van de drie modellen vergelijkbaar is voor de Nederlandse IC populatie en dat alle drie de modellen gebruikt kunnen worden voor IC kwaliteitsmetingen. De uitgebreide classificatie van IC opnameredenen die gebruikt wordt voor case-mix correctie is een duidelijke meerwaarde van het APACHE IV model hetgeen het mogelijk maakt om analyses te doen voor specifieke IC subgroepen. Bovendien heeft het APACHE IV model, in aanvulling is op de oudere modellen, de mogelijkheid om ook de verwachte sterfte van cardio-chirurgische patiënten te voorspellen.

Hoofdstuk 4 evalueert of ziekenhuissterfte geassocieerd is met het tijdstip van opname op de IC. De studie maakte gebruik van gegevens uit de NICE registratie bestaande uit IC patiënten die opgenomen waren tussen 2002 en 2008. Een IC opname tijdens kantooruren werd in deze studie gedefinieerd als een opname tussen 8:00-22.00 uur op weekdays en tussen 9:00-18.00 uur op weekenddagen. Het weekend werd hierbij gedefinieerd als zaterdag van 00.00 uur tot en met zondag 24.00 uur. De ziekenhuissterfte varieerde gedurende de IC opnametijd en was hoger dan verwacht buiten kantooruren en lager dan verwacht tijdens kantooruren. Er was geen statistisch significant verschil in ziekenhuissterfte tussen de verschillende weekdays, maar de sterfte was licht gestegen

voor opnamen op vrijdag. De studie toonde aan dat de ziekenhuissterfte verhoogd is buiten kantooruren en in het weekend, ook na correctie voor ernst van de ziekte (uitgedrukt als de APACHE II score), reden van opname, opname type, leeftijd en geslacht.

Hoofdstuk 5 richt zich op de determinanten (voorspellers) van sterfte na ziekenhuisontslag van IC patiënten. Voor deze studie is eerst een literatuuronderzoek naar de determinanten van lange termijn sterfte uitgevoerd. Daarna werd een cohort studie uitgevoerd naar de invloed van deze gevonden determinanten op de sterfte 3, 6 en 12 maanden na ziekenhuisontslag. Voor het literatuuronderzoek werden de databases van Pubmed en EMBASE doorzocht op Engels geschreven artikelen gepubliceerd tussen 1966 en 2011. Dit resulteerde in 24 artikelen die zijn opgenomen in het onderzoek en zijn samengevat op basis van kenmerken van de patiënt populatie, studiedesign, methoden, resultaten en determinanten die gebruikt zijn voor case-mix correctie. Daarnaast werd de kwaliteit van de artikelen beoordeeld met behulp van een vooraf gedefinieerde checklist voor studies naar lange termijn uitkomsten. De determinanten die gebruikt werden voor case-mix correctie varieerde sterk tussen de studies. Voor de cohort studie werden gegevens uit de NICE registratie gebruikt van patiënten van 81 Nederlandse ICs die tussen 2007 en 2010 levend uit het ziekenhuis werden ontslagen. De cohort studie toonde aan dat de determinanten leeftijd, reden voor IC opname en de aanwezigheid van bepaalde co-morbiditeiten bleken een associatie te hebben met de sterfte op 3, 6 en 12 maanden na ziekenhuisontslag. De grootte en richting van de invloed van deze determinanten verschildte per eindpunt. De meest gebruikte determinanten uit het literatuuronderzoek waren vergelijkbaar met de belangrijkste determinanten gevonden in de cohort studie.

Hoofdstuk 6 beschrijft de sterfte van IC patiënten na ontslag uit het ziekenhuis en vergelijkt dit met de algemene Nederlandse bevolking. De studie gebruikte NICE gegevens die gekoppeld waren aan administratieve gegevens uit een overkoepelend verzekeringsdatabase. Deze gekoppelde dataset bestond uit patiënten van 81 Nederlandse ICs die tussen 2007 en 2010 levend uit het ziekenhuis werden ontslagen. De geobserveerde sterfte werd weergegeven met behulp van Kaplan-Meier curves. Het risico op sterfte binnen 1, 2 en 3 jaar na ziekenhuisontslag was respectievelijk 12,5%, 19,3% en 27,5%. De sterfte na ziekenhuisontslag verschildte sterk tussen IC subgroepen op basis van opnameredenen. Electieve en cardio chirurgische patiënten hadden een statistisch significant lagere case-mix gecorrigeerde sterfte 1 jaar na ziekenhuisontslag, terwijl medische patiënten en patiënten opgenomen voor kanker een statistisch significant hogere case-mix gecorrigeerde sterfte 1 jaar na ziekenhuisontslag hadden in vergelijking met andere IC patiënten. Deze studie toonde aan dat ziekenhuissterfte de werkelijke sterfte van IC patiënten onderschat aangezien de sterfte in de eerste maanden na ziekenhuisontslag aanzienlijk was. De meeste IC patiënten hadden in vergelijking met de

algemene Nederlandse bevolking een verhoogd sterfterisico in de eerste jaren na ziekenhuisontslag.

Hoofdstuk 7 evalueert of het APACHE IV model gebruikt kan worden voor een valide voorspelling van de lange termijn sterfte op vaste eindpunten (3, 6 en 12 maanden) na de IC opname. Daarnaast werd het effect op de SMR geanalyseerd wanneer er gebruik werd gemaakt van de sterfte op vaste eindpunten na IC opname in plaats van de ziekenhuissterfte. De studie gebruikte NICE gegevens die gekoppeld waren aan administratieve gegevens uit een overkoepelende verzekeringsdatabase. Deze gekoppelde dataset bestond uit IC patiënten van 44 Nederlandse IC's die levend werden ontslagen tussen 2008 en 2011. Het APACHE IV model werd gerecalibreerd op ziekenhuissterfte en op de sterfte 3, 6 en 12 maanden na IC opname, resulterend in vier verschillende APACHE IV modellen. Maten van discriminatie, nauwkeurigheid, en calibratie (AUC, Brier score en calibratie grafieken) werden gebruikt om de verschillen in validatie van de vier APACHE IV modellen te meten. Het APACHE IV model presteerde het best wanneer het werd toegepast voor de voorspelling van de ziekenhuissterfte. De prestatie werd iets minder in de loop van de tijd maar het gerecalibreerde APACHE IV blijft voldoende goed voor het voorspellen van de sterfte 3, 6 en 12 maanden na IC opname. Zeven ICs hadden een statistisch significant verhoogde SMR en zeven ICs hadden een statistisch significant verlaagde SMR bij het gebruik van de sterfte 3 maanden na IC opname in plaats van de ziekenhuissterfte. Dit suggereert dat de ziekenhuissterfte beïnvloedt kan zijn door verschil in ontslagbeleid. Indien een SMR als kwaliteitsindicator wordt gebruikt bij het vergelijken van verschillende ICs heeft een SMR gebaseerd op lange termijn sterfte de voorkeur.

Dankwoord



Met veel plezier heb ik mijn promotieonderzoek uitgevoerd. Dit had echter niet gekund zonder de hulp van alle mensen die meedachten bij het opzetten, uitvoeren en vastleggen van het onderzoek. Ik ben hen allen erg dankbaar.

Allereerst wil ik mijn beide promotoren en copromotor bedanken. Beste Evert, met je kritische kijk heb je al mijn onderzoeken weten te verbeteren tot een mooi resultaat. Bedankt voor het delen van jouw inzichten en voor je inzet en enthousiasme die je toonde tijdens mijn promotie. Beste Ameen, van copromotor werd je mijn promotor. Bedankt dat het altijd lukte om een plekje te vinden in je agenda om even te brainstormen over hoe we de data moesten gaan analyseren. Jouw inbreng bij het toepassen en beschrijven van de juiste methoden en het oplossen van de lastige R puzzels was onmisbaar. Lieve Nicolette, als copromotor ben jij het meest betrokken geweest bij mijn promotie. Ik was altijd verbaasd over de snelheid waarmee je mijn stukken van waardevolle opmerkingen en commentaar had voorzien. Bedankt voor je goede begeleiding. Jouw harde werken en enthousiasme voor de wetenschap heeft me altijd gemotiveerd. Ook mijn coauteurs wil ik bedanken voor hun waardevolle inbreng tijdens mijn promotie. Alle leden van de promotiecommissie bedank ik voor hun tijd om mijn proefschrift te beoordelen. Dit proefschrift zou er niet zijn zonder de vele mensen die zich inzetten voor de NICE registratie. Het NICE bestuur en de NICE-KIKkers wil ik bedanken voor de fijne samenwerking. Ook de NICE deelnemers wil ik bedanken, zonder hun inzet bij het vastleggen van de gegevens was mijn promotie niet mogelijk.

Ik kijk terug op een leuke tijd op de KIK waar ik mijn promotie heb uitgevoerd. Bedankt voor jullie gezelligheid tijdens de lunch, KIK uitjes en borrels. De FAM-meiden, Ferishta, Anneke en Moniek, wil ik hier graag in het bijzonder noemen. Ik ben erg blij dat jullie mijn vriendinnen zijn geworden en heb altijd genoten van onze “diepgaande” gesprekken tijdens de lunch en andere uitjes. Ik hoop dat we de FAMS-uitjes blijven plannen! Daarnaast wil ik mijn paranimfen bedanken. Lieve Moniek, jij stond altijd klaar met je leuke weetjes en wijze adviezen en zorgde voor wat vrolijke afleiding. Bedankt voor de leuke Skype gesprekken en de gezellige autoritjes. Lieve Suzanne, bedankt dat je altijd voor me klaar staat. Ik ben blij met onze hechte band en blij dat je naast me wilt staan tijdens mijn verdediging.

Ik wil mijn familie en vrienden bedanken. Lieve Frans, helaas kan jij er niet meer bij zijn. Ik had je tijdens mijn verdediging heel graag op de voorste rij, zo trots als een pauw, willen zien zitten. Je wordt nog steeds heel erg gemist. Lieve Rob en Betty, zonder jullie was ik nooit zover gekomen. Bedankt dat zowel Dennis als ik altijd bij jullie terecht kunnen. Lieve broers, Maurice en Stefan, bedankt dat ik altijd op jullie kan rekenen. Suzanne, je bent al genoemd als paranimf, maar hier kan je niet ontbreken om ook als mijn zus genoemd te worden. Lieve Liese, mijn vrolijke kleine meid, ik geniet elke dag van je! Lieve Dennis, we zijn al jaren samen en genieten inmiddels samen van onze dochter. Ik hoop dat we samen nog vele mooie momenten mogen meemaken.

Bedankt, Sylvia

**Curriculum Vitae
&
List of publications**



Sylvia Brinkman werd op 24 oktober 1980 geboren in Utrecht. Na het behalen van haar HAVO diploma aan de werkplaats in Bilthoven in 1998 begon zij met de studie diergezondheidszorg aan de Hogere Agrarische School in Den Bosch. Na het behalen van haar diploma in 2003 is zij begonnen aan de studie dierwetenschappen aan de Universiteit van Wageningen waar ze in 2005 haar masterdiploma heeft behaald.

Ze is vervolgens gaan werken als datamanager bij het Julius Centrum in Utrecht. In 2007 is ze begonnen als datamanager en promovenda bij de Nationale Intensive Care Evaluation (NICE) registratie op de afdeling Klinische Informatiekunde van het Academisch Medisch Centrum in Amsterdam. Hier heeft zij onder leiding van Nicolette de Keizer haar promotieonderzoek dat beschreven is in dit proefschrift uitgevoerd. Na het afronden van haar promotieonderzoek zet Sylvia Brinkman haar werkzaamheden als post-doc voort bij de NICE registratie.

Gedurende het promotietraject heeft Sylvia de volgende cursussen gevolgd en presentaties gegeven:

Cursussen	Jaar	Werklast in uren
Scientific writing in English for publication	2008	24
NIHES regressie analyse	2008	36
HSMR conferentie	2008	5
Basic biostatistics	2008	36
Advanced biostatistics	2009	36
Better use of PubMed	2009	16
Multivariable model building	2009	24
WEON propensity score	2009	5
R cursus	2009	8
CME: introductory course on epidemiology	2010	24
Klinische epidemiologie en biostatistiek	2010	36
Advanced topics in clinical epidemiology	2010	36
NIHES survival analysis	2010	36
Project management	2012	24
Literatuurbesprekingen	2007-2012	200
Presentaties		
Quality and safety congress in Amsterdam "Prognostic models to compare the quality of ICUs general vs IC specific models"	2010	24
MIE congress in PISA "The use of different quality registries to assess long-term mortality of ICU patients"	2012	24
NICE training "scoringssystemen binnen NICE" (4x per jaar)	2007-2012	60

List of publications

- Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, de Jonge E, de Keizer NF. Determinants of mortality after hospital discharge in intensive care patients; Literature review and Dutch cohort study. Crit Care Med. 2013 [in press].
- Brinkman S, de Jonge E, Abu-Hanna A, Arbous MS, de Lange DW, de Keizer NF. Mortality after hospital discharge in ICU patients. Crit Care Med. 2013 [in press].
- Brinkman S, de Jonge E, Abu-Hanna A, de Lange DW, de Keizer NF. The use of linked registries to assess long-term mortality of ICU patients. Stud Health Technol Inform. 2012;180:230-4.
- Brinkman S, Abu-Hanna A, van der Veen A, de Jonge E, de Keizer NF. 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 standardized mortality ratios of intensive care units. Crit Care Med. 2012 Feb;40(2):373-8.
- Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, de Jonge E, Bosman RJ, Peelen L, de Keizer NF. External validation of Acute Physiology and Chronic Health Evaluation IV in Dutch intensive care units and comparison with Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II. J Crit Care. 2011 Feb;26(1):105.e11-8.
- van der Wal G, Brinkman S, Bisschops LL, Hoedemaekers CW, van der Hoeven JG, de Lange DW, de Keizer NF, Pickkers P. Influence of mild therapeutic hypothermia after cardiac arrest on hospital mortality. Crit Care Med. 2011 Jan;39(1):84-8.
- Kuijsten HA, Brinkman S, Meynaar IA, Spronk PE, van der Spoel JI, Bosman RJ, de Keizer NF, Abu-Hanna A, de Lange DW. Hospital mortality is associated with ICU admission time. Intensive Care Med. 2010 Oct;36(10):1765-71.