

Length of intensive care unit stay following cardiac surgery: is it impossible to find a universal prediction model?

Yunita Widyastuti^{a,b}, Roar Stenseth^{c,d}, Alexander Wahba^{c,e}, Hilde Pleym^{d,f} and Vibeke Videm^{a,f,*}

^a Department of Laboratory Medicine, Children's and Women's Health, Norwegian University of Science and Technology, Trondheim, Norway

^b Department of Anaesthesiology and Reanimation, Faculty of Medicine Gadjah Mada University, Yogyakarta, Indonesia

^c Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway

^d Department of Cardiothoracic Anaesthesia and Intensive Care, St. Olavs University Hospital, Trondheim, Norway

^e Department of Cardiothoracic Surgery, St. Olavs University Hospital, Trondheim, Norway

^f Department of Immunology and Transfusion Medicine, St. Olavs University Hospital, Trondheim, Norway

* Corresponding author: Department of Immunology and Transfusion Medicine, St. Olavs University Hospital, 7006 Trondheim, Norway.
Tel: +47-72-573321; fax: +47-72-576426; e-mail: vibeke.videm@ntnu.no (V. Videm).

Received 3 March 2012; received in revised form 25 May 2012; accepted 31 May 2012

Abstract

OBJECTIVES: Accurate models for prediction of a prolonged intensive care unit (ICU) stay following cardiac surgery may be developed using Cox proportional hazards regression. Our aims were to develop a preoperative and intraoperative model to predict the length of the ICU stay and to compare our models with published risk models, including the EuroSCORE II.

METHODS: Models were developed using data from all patients undergoing cardiac surgery at St. Olavs Hospital, Trondheim, Norway from 2000–2007 ($n = 4994$). Internal validation and calibration were performed by bootstrapping. Discrimination was assessed by areas under the receiver operating characteristics curves and calibration for the published logistic regression models with the Hosmer-Lemeshow test.

RESULTS: Despite a diverse risk profile, 93.7% of the patients had an ICU stay <2 days, in keeping with our fast-track regimen. Our models showed good calibration and excellent discrimination for prediction of a prolonged stay of more than 2, 5 or 7 days. Discrimination by the EuroSCORE II and other published models was good, but calibration was poor (Hosmer-Lemeshow test: $P < 0.0001$), probably due to the short ICU stays of almost all our patients. None of the models were useful for prediction of ICU stay in individual patients because most patients in all risk categories of all models had short ICU stays (75th percentiles: 1 day).

CONCLUSIONS: A universal model for prediction of ICU stay may be difficult to develop, as the distribution of length of stay may depend on both medical factors and institutional policies governing ICU discharge.

Keywords: Risk prediction • Length of ICU stay • Cardiac surgery

INTRODUCTION

Over a period of time, mortality rates following cardiac surgery have been steadily decreasing. In parallel, however, patients undergoing cardiac surgery tend to be older, having more advanced disease and greater co-morbidity [1, 2]. Consequently, the risk of postoperative complications has increased, leading to more patients with a prolonged ICU stay. Prolonged ICU stays are associated with lower survival rates, as well as reduced quality of life [3]. They also lead to increased hospital costs and reduced ICU bed availability [4, 5]. Accurate prediction of length of stay in the ICU enables clinicians to provide patients with more reliable information for informed consent, enables better planning of treatment and gives better indications for allocation of resources. Furthermore, it allows for computation of risk-adjusted length of stay and comparison among and within institutions, e.g. after changes in routines [6, 7].

Even if a prolonged ICU stay is also related to other post-operative outcomes, it is therefore of interest in its own right, not as a surrogate for more specific complications. Thus, several models have been developed to predict prolonged ICU stay [4, 8–11]. Risk prediction models like the EuroSCORE, originally developed for mortality prediction, have also been evaluated for prediction of extended ICU stay [12–14]. The discriminatory ability of the EuroSCORE was poor to acceptable in predicting a length of stay of more than two days but, in one study, the calibration of the EuroSCORE was good [12–14]. To our knowledge, the EuroSCORE II that has recently been launched has not yet been evaluated for prediction of a prolonged stay in ICU [15].

Most models are only able to predict whether or not the patients will remain in the ICU for a predetermined number of days. An exception is the recently published predictive model developed by De Cocker and co-workers, based on preoperative variables, which can be used in the form of a risk index that correlates to the mean length of stay in the development

dataset [7]. This is denoted the 'Morbidity Defining Cardiosurgical index' or MDC-index. However, because the definition of a prolonged ICU stay varies between institutions, a model may not work well in other regions or hospitals that use a different definition. It is therefore of interest to evaluate how this scoring system works in other patient populations.

Our hypotheses were that a model using only preoperative variables would be able to predict the length of stay in the ICU, but that the addition of intraoperative variables would increase the accuracy. We also hypothesized that the model specifically developed to predict length of ICU stay would perform better than mortality-prediction models.

Our aims were twofold: first, to develop a model that could accurately predict length of ICU stay and, second, to compare our model with De Cocker's model, the EuroSCORE II, and with our previously published models for mortality (within 30 days postoperatively or during the current hospital stay) and for prolonged mechanical ventilation (more than 24 hours) [16, 17].

PATIENTS AND METHODS

From 2000–2007, preoperative and intraoperative data were consecutively collected from patients undergoing cardiac surgery at St. Olavs Hospital, Trondheim, Norway ($n=5029$). Quality control of the data was performed by senior anaesthesiologists during the course of treatment. After patient discharge, a single senior anaesthesiologist (RS) performed a final, independent, complete data check for all study patients. The EuroSCORE II, De Cocker's score and our group's mortality and prolonged ventilation model scores were calculated during this study. Based on clinical grounds and the literature, 21 preoperative variables and nine intraoperative variables were chosen to be included, to avoid over-fitting [18]. Definitions of variables are given in Table 1. Patients were discharged from the ICU as soon as vital functions and chest tube drainage were stable and all ventilatory or inotropic support (mechanical or inotropic drugs) was terminated. However, patients were not discharged from the ICU on the day of surgery. Continuous infusion of loop diuretics or low-dose noradrenaline infusion for compensation of peripheral vasodilation was accepted on the ward in a minority of cases. For patients who were transferred to the ICU of a local hospital, the total ICU stay in both hospitals was recorded for this study. Patients who had missing data and patients receiving dialysis preoperatively were excluded because dialysis patients, by protocol, were given a prolonged stay in the ICU for postoperative haemofiltration or dialysis. Thirty-four patients on preoperative dialysis and one patient with missing data were excluded, leaving 4994 eligible patients for model development.

Model development

All eligible patients were included in model development, which was performed with Cox proportional hazard regression modelling using the Design package in the 'R' statistical environment (version 2.13.1; R Foundation, <http://www.r-project.org>) [19]. The entire dataset was used for model development because data-splitting has been shown to reduce the predictive accuracy of the fitted model [19]. The outcome was coded as '1 = discharged from the ICU' and '0 = not discharged from the ICU', and patients who died prior to ICU discharge were thereby

censored. The time variable was time until discharge from the ICU or -until death (i.e. 'loss to follow-up'). A model containing only preoperative factors (Model I) was first developed. For development of Model II, which also included intraoperative variables, the variables shown in Table 1 were added to Model I, and the modelling strategy indicated below was repeated. Only patients who underwent cardiopulmonary bypass (CPB) were included for development of Model II ($n=4869$, i.e. 97.5%).

First, the full main effects model was fitted. The assumption of proportionality of the hazard was checked by log-minus-log plots, using SPSS software (version 19.0; SPSS Inc., Chicago, USA). For continuous variables, we also tested whether they could better be modelled using restricted cubic spline functions [19]. Predefined interactions were then tested, as suggested as the most efficient modelling strategy [18]. The predefined interactions in Model I were between age and degree of urgency, and between age and preoperative haemoglobin concentrations.

We then tested for overly-influential observations (multivariate outliers), defined as observations leading to a change in the regression coefficient of more than 0.2 standard errors, using the method based on the vector of score residuals in the Design package. Limited step-down, based on Akaike's information criterion (AIC), was performed after the final full model was obtained. To find more robust (or correct) estimates of the coefficients from the remaining significant variables from the step-down model, the model was fitted by bootstrapping ($n=400$ repetitions). By this method, the model is repeatedly fitted with step-down in a bootstrap sample and performance of each model is evaluated on the original sample. Hazard ratios and 95% confidence intervals were calculated from the bootstrapped coefficients. In this paper, the coefficient of each variable was negatively exponentiated to find the hazard ratio for stay (instead of discharge) to ease interpretation of the risk variables.

Model validation and calibration

Internal validation was done with bootstrapping ($n=400$ repetitions) by testing the final model on different random selections of patients from our sample. From this procedure, we estimated the optimism of the model if it were applied to a future dataset –also known as the shrinkage factor. A shrinkage factor above 0.85 is considered satisfactory. Calibration was also performed by bootstrapping ($n=400$ repetitions), resulting in bootstrap overfitting-corrected calibration curves of predicted vs observed probabilities of an ICU stay of more than 2, 5 or 7 days. The calibration plots were generated by dividing patients into 10 equally-sized groups along the range of predicted prolonged stay, and plotting the Kaplan-Meier estimate within each group against the mean predicted outcome in the same group [20]. The Hosmer-Lemeshow test was used to assess calibration of the previously published logistic regression models (EuroSCORE II, our mortality and prolonged ventilation models).

Accuracy, defined as the ability of the model to discriminate between two patients with different lengths of stay, was evaluated by receiver operating characteristics (ROC) curves. The area under the curve (AUC or c-statistic) was used to compare the discriminative ability of the models to predict an ICU stay of more than 2, 5, or 7 days. An AUC higher than 0.7 is considered acceptable and an AUC higher than 0.8 is considered good. The AUC plots were obtained using the SPSS software. The AUCs for the different

Table 1: Variable definitions and patient characteristics

Variables	Definition	Mean (95% CI) or percentage 'yes'
Age	Continuous variable (years)	66.1 (65.8–66.4)
Sex	male/female	74.3 / 25.7
BMI	Body mass index, continuous variable (kg/m ²)	26.7 (26.6–26.8)
Smoking	Current smoker or quit <6 months ago (yes/no)	45.3
Hypertension	Receiving medication or diastolic blood pressure >90 mmHg (yes/no)	48.7
Diabetes mellitus	Receiving medication (yes/no)	12.8
Previous myocardial infarction	History of myocardial infarction (yes/no)	45.8
Preoperative intra-aortic balloon pump (IABP)		0.6
Chronic heart failure	Based on history and clinical evaluation by an attending cardiologist (yes/no)	15.6
Pulmonary hypertension	Systolic pulmonary arterial pressure (PAP) >40 mmHg or mean PAP > 25 mmHg, echocardiography or catheterization (yes/no)	8.9
Left ventricular hypertrophy	Electrocardiography or echocardiography (yes/no)	20.2
Peripheral arterial disease	Aortic aneurysm or carotid stenosis or <i>claudicatio intermittens</i> (yes/no)	10.6
NYHA class	New York Heart Association classification, Class I or II vs Class III or IV	29.3% / 70.8%
Non-sinus rhythm	Electrocardiography (yes/no)	7.6
Chronic pulmonary disease	Use of bronchodilating agents or FEV<75% (yes/no)	13.9
Preoperative renal dysfunction	Serum creatinine >140 µmol/L (yes/no)	4.2
Previous syncope	(yes/no)	6.0
Previous cardiac surgery	(yes/no)	5.8
Preoperative haemoglobin concentration	Continuous variable (mg/dL)	13.7 (13.6–13.7)
Degree of urgency	0 = standard waiting list 1 = operation within 2 weeks 2 = operation within 24 h	54.3 40.1 5.6
Operation type	1 = Coronary artery bypass grafting or atrial septum defect correction 2 = Aortic valve replacement only, AVR and CABG combined, non-ischaemic mitral valve replacement/repair or aneurysm in the ascending aorta 3 = Dissection of the ascending aorta or ventricular septum rupture 4 = Miscellaneous*	70.4 21.2 1.9 6.5
Defibrillation	Defibrillation for ventricular fibrillation during surgery (yes/no)	15.8
Inotropic support	On clinical indication during surgery (yes/no)	23.5
Vasoconstrictor treatment	On clinical indication during surgery (yes/no)	76.1
Plasma transfusion	On clinical indication during surgery (yes/no)	8.8
Red blood cell transfusion	On clinical indication during surgery (yes/no)	15.0
Platelet transfusion	On clinical indication during surgery (yes/no)	11.8
Intraoperative bleeding	Bleeding >1000 mL (yes/no)	9.8
Fluid balance	Tertiles of fluid balance during surgery: 1 : ≤2600 mL 2 : 2600–3175 mL 3 : ≥3175 mL	33.4 33.3 33.3
CPB time	Time on cardiopulmonary bypass (per 10 min)	8.2 (8.0–8.3)
Thirty day-mortality rate		2.7%
EuroSCORE II	Mean of EuroSCORE II	2.8% (2.7%–2.9%)
ICU stay	≤2 days 3–6 days ≥7 days	4682 (93.7%) 198 (4%) 114 (2.3%)

AVR: aortic valve replacement; CABG: coronary artery bypass grafting; CI: confidence interval; FEV: forced expiratory volume.

*= mitral valve surgery in combination with CABG or AVR, AVR in combination with procedures other than CABG or operation for aneurysm of the ascending aorta, and other cardiac surgery like pericardiectomy and removal of cardiac tumours.

models were compared by DeLong's method [21] using SigmaPlot (version 11.0; Systat Software, Inc., San Jose, CA, USA).

To compare our model with De Cocker's model, the EuroSCORE II, our published mortality model and our prolonged ventilation model, we used variables from our dataset with the same definitions as theirs. Non-complete cases were excluded. The scores were then calculated in accordance with their respective logistic regression coefficients with constants or their respective Cox regression coefficients. To assess whether the models can be used for patient stratification, we calculated the observed median ICU stay according to three categories: low, intermediate and high risk, as obtained from the tertiles of each model. The median and 95th percentile of the observed ICU

stay in each group for each score were then plotted. We also compared the stratification ability of Model I and De Cocker's model in our patients, based on the MDC-index [7].

Data are given as mean ± 95% confidence interval for continuous variables, and as frequency (percentage) for categorical variables, unless otherwise is stated. *P*-values below 0.05 were considered statistically significant.

RESULTS

Patient characteristics, degrees of urgency of the operations, operation types, and intraoperative variables are given in Table 1.

Sixty-five patients (1.3%) died prior to discharge from the ICU. The predefined interactions were not significant ($P > 0.05$) and there were no overly-influential cases. Most of the patients (89.7%) were discharged during the first ICU day and the median stay was, correspondingly, 1 day. The longest stay was 65 days. On a year-to-year basis, the mean stay varied from 1.2 to 1.6 days with no obvious trend. Further information on ICU stays is given in Table 1.

Independent predictors of a prolonged ICU stay

Table 2 shows the preoperative and intraoperative variables that were independent predictors of a prolonged ICU stay. The preoperative model (Model I) included age, renal insufficiency, chronic pulmonary disease, peripheral arterial disease, chronic heart failure, left ventricular hypertrophy, previous cardiac surgery, preoperative intra-aortic balloon pump, degree of urgency and type of operation.

The intraoperative model (Model II) included renal insufficiency, chronic pulmonary disease, peripheral arterial disease, left ventricular hypertrophy, preoperative intra-aortic balloon pump, degree of urgency, type of operation, intraoperative inotropic support, intraoperative red cell transfusion, intraoperative platelet transfusion, intraoperative bleeding and CPB duration as predictors of a prolonged ICU stay (Table 2).

Calibration and discrimination

Figure 1A shows the calibration curve of Model I for prediction of a stay in the ICU of more than 2 days (i.e. observed vs

predicted probability). The model was well calibrated in lower-risk patients but showed slight underestimation in the highest-risk patients. The curves for a stay in the ICU of more than 5 or 7 days were comparable (data not shown). The curves for Model II were very similar (data not shown). The shrinkage factors for Model I, for stays of more than 2, 5 or 7 days, were 0.96, 0.97 and 0.96, respectively, indicating that the model should give accurate predictions in a future dataset.

ROC curves of Model I, Model II, EuroSCORE II, De Cocker's model and our previously published mortality and prolonged ventilation models showed excellent to acceptable ability to discriminate when predicting an ICU stay of more than 2, 5 or 7 days. Fig. 1B shows the curves for more than 2 days (other curves not shown). The AUC was significantly different between Model I (AUC = 0.824 (0.800–0.848)) and Model II (AUC = 0.862 (0.840–0.885), $P = < 0.001$), but not between Model I and our prolonged ventilation model (AUC = 0.815 (0.790–0.840), $P = 0.11$). However, Model I had a significantly larger AUC than the EuroSCORE II (AUC = 0.801 (0.776–0.826), $P = 0.008$), our published mortality model (AUC = 0.793 (0.767–0.820), $P < 0.0001$) and De Cocker's model (AUC = 0.752 (0.725–0.780), $P < 0.0001$). Our mortality model and the EuroSCORE II did not have significantly different AUCs ($P = 0.47$) but both had significantly higher AUCs than De Cocker's model ($P = 0.004$ and $P < 0.0001$, respectively).

However, when the scores for our patients from Model I, De Cocker's model, EuroSCORE II, our published mortality model and our prolonged ventilation model were sorted into three groups corresponding to a low, intermediate or high risk of a prolonged ICU stay, the median observed stay in all risk groups

Table 2: Cox regression models for length of ICU stay

Variable	Model I (n = 4994)				Model II (4869)			
	Coefficient	P-value	HR	95% CI (bootstrapped)	Coefficient	P-value	HR	95% CI (bootstrapped)
Age [†]	-0.003	0.016	1.003	1.003–1.004		NS		
Renal insufficiency	-0.280	0.0005	1.323	1.090–1.447	-0.285	<0.0001	1.330	1.198–1.461
Chronic pulmonary disease	-0.149	0.001	1.161	1.080–1.242	0.127	0.003	1.136	1.026–1.209
Peripheral arterial disease	-0.139	0.002	1.149	1.068–1.235	-0.152	0.001	1.165	1.077–1.252
Chronic heart failure	-0.164	0.0002	1.178	1.090–1.265		NS		
Left ventricle hypertrophy	-0.114	0.019	1.122	1.033–1.208	0.110	0.015	1.117	1.036–1.197
Previous cardiac surgery	-0.344	<0.0001	1.410	1.291–1.530		NS		
Preoperative intra-aortic balloon pump	-0.840	0.0001	2.316	2.079–2.554	-0.819	<0.0001	2.269	2.006–2.532
Degree of urgency								
Standard waiting list			1	Reference			1	Reference
Operation within 1 week	-0.082	0.013	1.086	1.035–1.137	-0.036	0.026	1.035	1.020–1.050
Operation within 24 h	-0.517	<0.0001	1.677	1.531–1.823	-0.361	<0.0001	1.435	1.302–1.568
Operation type								
CABG or ASD			1	Reference			1	Reference
Pure AVR, AVR and CABG, non-ischaemic MVR/R, or aneurysm of ascending aorta	-0.142	0.001	1.157	1.077–1.238	0.079	0.14	0.924	0.838–1.009
Dissection of ascending aorta, or VSR	-1.037	<0.0001	2.820	2.640–2.999	-0.348	0.021	1.417	1.207–1.626
Miscellaneous	-0.427	<0.0001	1.533	1.405–1.662	-0.136	0.043	1.145	1.009–1.281
Intraoperative inotropic support					-0.142	0.0002	1.152	1.094–1.210
Intraoperative red cell transfusion					-0.174	0.0001	1.190	1.107–1.274
Intraoperative platelet transfusion					-0.360	<0.0001	1.434	1.330–1.537
Intraoperative bleeding					-0.225	0.0003	1.252	1.138–1.367
CPB duration*					-0.028	<0.0001	1.028	1.020–1.036

ASD: atrial septal defect; AVR: aortic valve replacement; CABG: coronary artery bypass grafting; CPB: cardiopulmonary bypass; HR: hazard ratio; ICU: intensive care unit; MVR/R: mitral valve replacement or repair; NS: non-significant; VSR: ventricle septum rupture.

[†]Per year.

*Every 10 min.

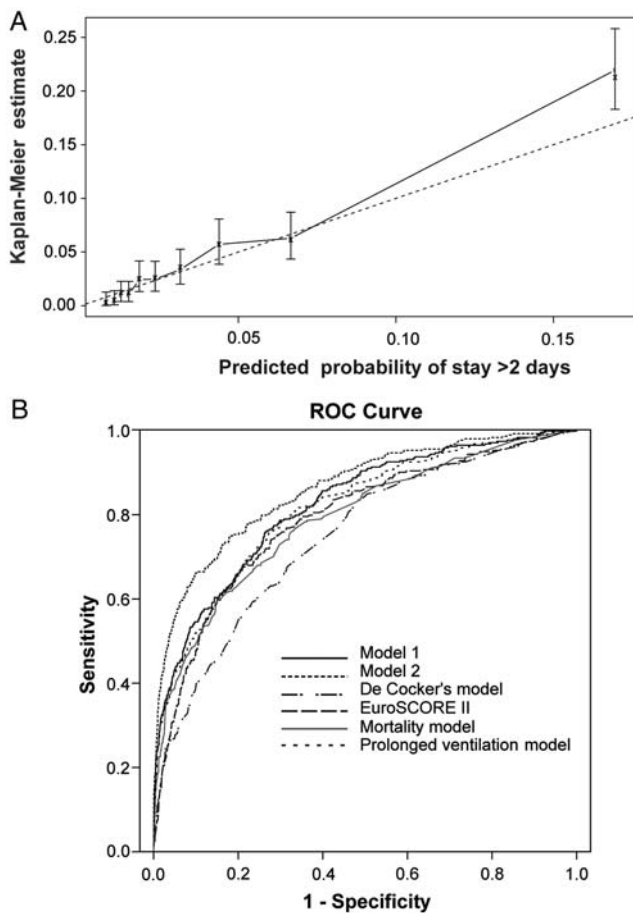


Figure 1: Model calibration and discrimination. (A) Calibration plot for Model 1 for an ICU stay of >2 days, bootstrapped ($n = 400$ repetitions) means with 95% confidence intervals. X-axis indicates predicted probability for each decile of patients; Y-axis gives corresponding Kaplan-Meier estimate. Dotted line indicates ideal calibration, i.e. predicted stay is identical to Kaplan-Meier estimate. (B) ROC curves for Model 1, Model 2, De Cocker's model [7], the EuroSCORE II [15], and the published models for mortality [16] and prolonged ventilation [17] from our group for an ICU stay >2 days.

for all models was 1 day (Fig. 2A). Only when comparing the 95th percentile, was there sufficient difference between the risk groups to discriminate between a high risk and a low or intermediate risk. Thus, none of the models could be used for prediction in individual patients because any level of the risk scores would correspond to most patients having a 1-day stay in the ICU.

Figure 2B (upper panel) shows the relationships between the patients who had ICU stays longer than 2, 5 or 7 days and the patients who died during the hospital stay. It is evident that far from all of the patients who had prolonged ICU stays finally died. For patients with an ICU stay longer than 2 days (i), the mortality rate was 23.1%; for patients with an ICU stay of more than 5 days (ii), it was 37.6%, and for patients having an ICU stay of over 7 days (iii), it was 46.0%. The overlap between patients who needed prolonged mechanical ventilation and those with a prolonged ICU stay was substantially larger (Fig. 2B, lower panel): 63.4% of the patients with an ICU stay of more than 2 days (i), 85.3% of those with an ICU stay of over 5 days (ii), and 84.9% of those with an ICU stay longer than 7 days (iii) needed prolonged ventilation. Neither the EuroSCORE II nor our published models for prolonged ventilation and mortality following cardiac surgery

were well-calibrated for prediction of an ICU stay of >2 days ($P < 0.0001$, Hosmer-Lemeshow test).

The distribution patterns of De Cocker's MDC-index in their study population and ours were relatively similar (Fig. 3A) [7]. To investigate whether De Cocker's MDC-index was useful for prediction of ICU stays in our patients, the observed mean stays were compared to the mean predicted stays in each group of the calculated MDC-index (Fig. 3B). The MDC-index gave a large overprediction in our patients.

DISCUSSION

In this study, we developed a preoperative- and an intraoperative model to predict length of stay in the ICU following cardiac surgery, using Cox regression because time to discharge is easily modelled with this approach. The models were well-calibrated. Importantly, however, despite good discrimination by our models as well as by several previously published scoring systems and models, none would be suitable for prediction in individual patients. This was because most patients were discharged during the first ICU day, independently of their risk level. We also found that the published MDC-index by De Cocker and co-workers gave large overpredictions of ICU stays in our study population [7].

Distributions of ICU stay data are usually right-skewed and several other authors also report a median of 1 day [10, 11, 13]. Our distribution had a shorter right tail than in these reports—corresponding to fewer patients with longer stays—as indicated by the mean lengths of stay that were 2.2 days [10], 1.9 days [11] and 1.8 days [13], as compared to 1.4 days in our population. However, mean values are not well suited to describe such skewed distributions and the use of means in the setting of ICU stay prediction may render it difficult to identify the problems related to prediction in individual patients, which are evident from Fig. 2A.

Prolonged ICU stay: too subjective an outcome measure?

The reason that De Cocker's MDC-index did not work well in our study population is probably not related to major differences in the patients themselves, since Fig. 3A shows largely similar distributions of the indices. Even so, many more of their patients had longer ICU stays than did ours, giving a median stay of 2 days and a mean stay of 5.5 days [7]. Thus, our data demonstrate that, in order for a score to be well-calibrated for use in a population other than the one in which it was developed, the policies on when to discharge the patient from the ICU must be comparable between the two institutions. This is consistent with the findings of a study comparing 14 published models to predict ICU stay, where only two models were well calibrated in the validation set [22].

We adopted a fast-track regimen in cardiac surgery in 1990, including a balanced intravenous/inhalational anaesthesia that permitted early extubation (following standard criteria and resulting in a median postoperative intubation time of 3 h) and mobilization out of bed the day after surgery. The staffing in the standard wards is sufficient that most patients could be discharged from the ICU during the first day and only 6.3% who suffered from the more serious complications remained in the

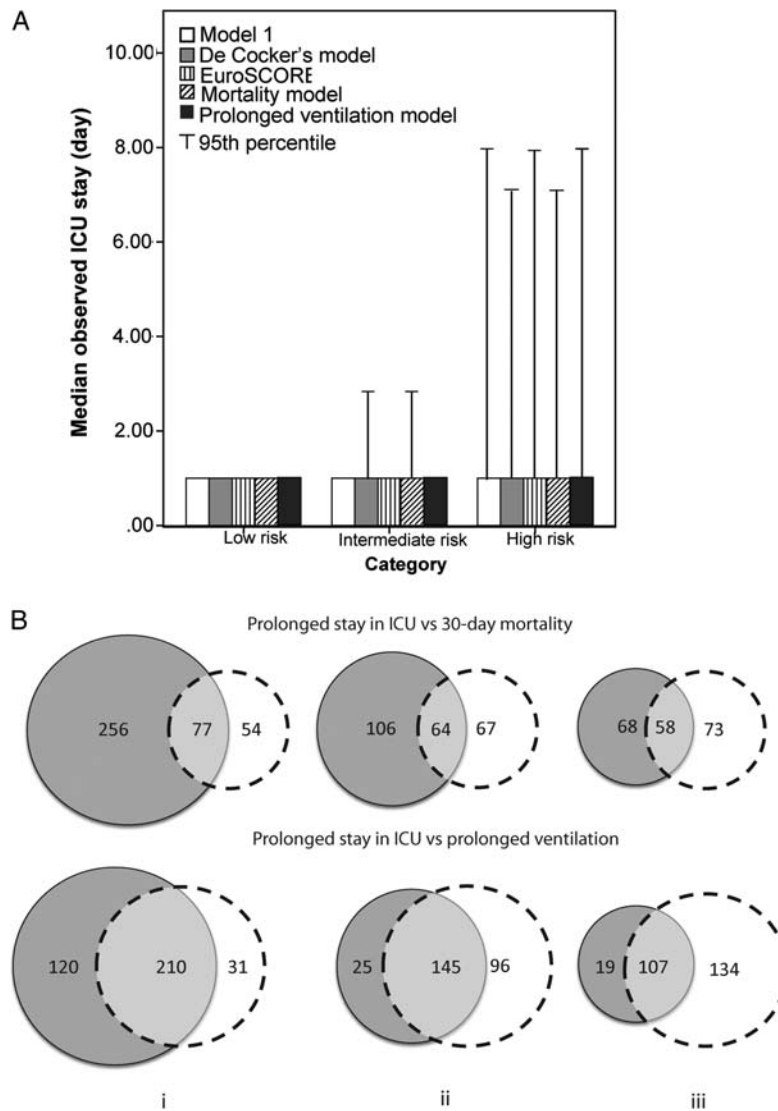


Figure 2: Predicted vs observed ICU stays and patient overlap. (A) Median observed ICU stay with 95th percentile for patients with calculated risk scores indicating low, intermediate, or high risk based on Model I, De Cocker's model [7], the EuroSCORE II [15], and the published models for mortality [16] and prolonged ventilation [17] from our group. For the low risk group, the 95th percentile was 1 day for all models. The 75th percentile of all risk groups was 1 day for all models. (B) Number of patients with an ICU stay of (i) >2 days, (ii) >5 days or (iii) >7 days in grey circles and number of patients who died (upper figure half) or needed prolonged mechanical ventilation (lower figure half) in circles with broken lines, indicating overlapping patients in light grey areas.

ICU for 3 days or longer. With a 30-day mortality rate of 2.7% and rates of some other important complications, such as cardiac dysfunction of 5.7% and prolonged ventilation of 4.9%, our results were comparable to those published by others [16, 17, 23]. Through this policy, more ICU beds are available to the patients who really need them and the number of operations performed can more easily meet demands. In other institutions, ICU capacity may be larger or the level of care in the standard wards may require the patients' condition to be better before it is advisable to transfer them there, leading to longer ICU stays. Thus, the observed patterns of ICU stay may be influenced by different policies for discharge. In this way, the situation for ICU stays seems to be somewhat parallel to that demonstrated for patterns of in-hospital stay by data from the USA and Britain, where different reimbursement models for cardiac surgery units in the two countries—as opposed to the type of patients or incidence of postoperative complications—were important for duration [1].

A prediction model for any outcome requires strict endpoint definitions in order to be useful. However, when the decision to discharge the patients from the ICU is most probably based on policy as well as medical criteria, it becomes difficult to make a good prediction model based solely on medical variables. Rather, length of ICU stay seems to be a somewhat subjective outcome, decided by a combination of medical and non-medical factors.

Most of the risk variables for a prolonged ICU stay identified in our population, are similar to those found by others [8–11]. Like De Cocker and co-workers, we found that the need for a preoperative intra-aortic balloon pump was the single most influential predictor [7]. With respect to intraoperative predictors, previous studies have identified duration of CPB as important [9, 24]. We also found that other intraoperative variables like inotropic support, transfusions and intraoperative bleeding of more than 1000 mL were significant.

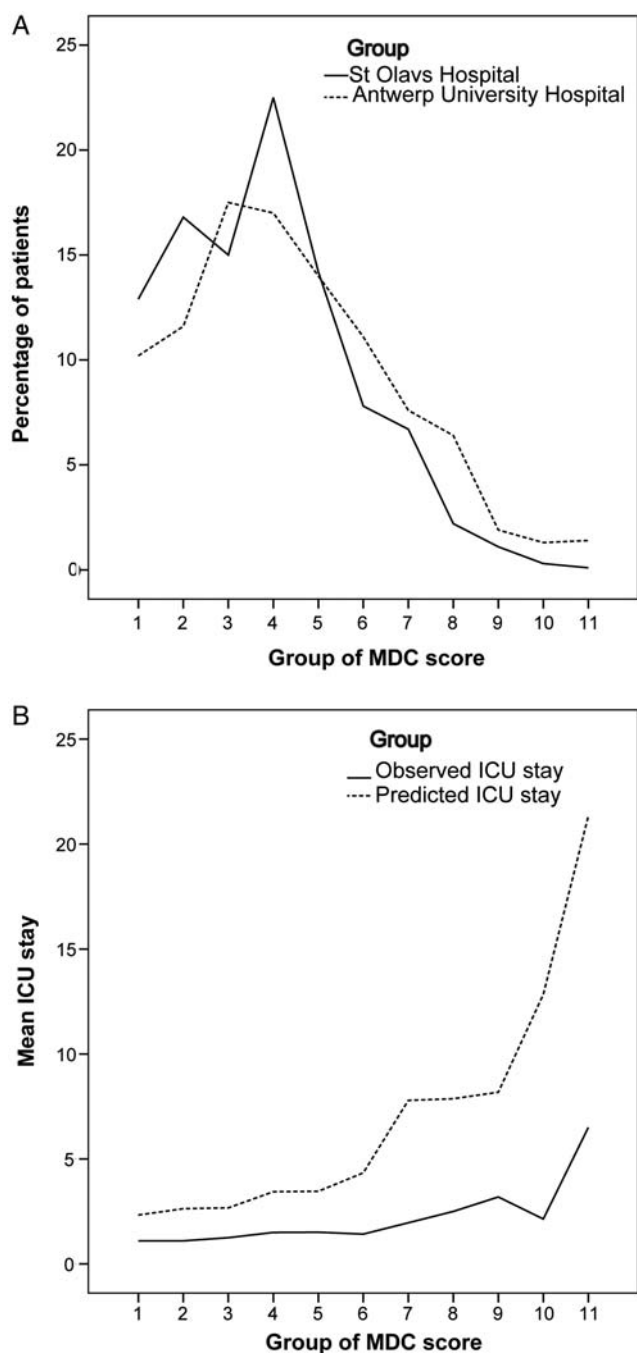


Figure 3: Comparisons with data from De Cocker and co-workers. The patients were grouped in 11 groups according to the published grouping of De Cocker's MDC-index. Data from [7] are re-used with permission from the authors and Oxford University Press on behalf of the European Association for Cardio-Thoracic Surgery. (A) Frequency of patients in the original study population at Antwerp University Hospital [7] and in our study population. (B) Observed and predicted ICU stay based on De Cocker's MDC-index [7].

Use of mortality scores to predict length of ICU stay

It is tempting to use a standard mortality score, which is often calculated as part of clinical routine, to predict the risk of other complications, including a prolonged ICU stay. Both the EuroSCORE II and our previously published mortality model showed good discrimination but had significantly lower AUCs than Models I and II

for ICU stay. This is probably explained by the fact that a substantial fraction of the patients who died did not overlap with the patients who had a prolonged ICU stay (Fig. 2B). This may be a consequence of the greatly reduced mortality rates from cardiac surgery over recent decades. Thus, even if several risk factors are common to both, these outcomes are sufficiently different that it seems unlikely that any mortality model will be sufficiently well calibrated for prediction of a prolonged ICU stay. This was confirmed by the Hosmer-Lemeshow tests.

On the other hand, our score for prolonged mechanical ventilation showed excellent discrimination, but still was not sufficiently well calibrated. As shown in Fig. 2B, most of the patients who had an ICU stay of more than 2, 5, or 7 days also underwent prolonged ventilation. Our percentages of overlap are comparable to the 72% previously found by Arabi and co-workers [4]. Another study also showed that mechanical ventilation at 24 h was a strong predictor of a long ICU stay [25]. Thus, scores developed to predict prolonged ventilation may be more relevant for prediction of prolonged ICU stays than mortality scores. Even so, the problem remains of varying criteria deciding the length of ICU stays in different hospitals.

Methodological considerations

Cox regression intuitively seems a better choice for development of a prediction model for ICU stay than logistic regression, which is mostly used for binary outcomes. Cox regression also permits censoring of patients who died in the ICU, thus taking into account that they would have had a prolonged stay if they were alive. However, individual prediction is easier from logistic regression models. Based on the distribution of ICU stays in our patients, we could have developed a logistic regression model to predict a stay of more than 2 days, where patients who died before discharge from the ICU could have been attributed to the group with a long ICU stay. However, the general applicability of such a model would be no better than for all the others tested, because of the above-mentioned subjectivity regarding ICU stay as an outcome.

Strengths and limitations of study

To our knowledge, this is the first study to assess the EuroSCORE II for prediction of length of ICU stay and to validate De Cocker's MDC-index in another population. However, our study clearly demonstrated that excellent discrimination is not a guarantee of accuracy of prediction, nor of the clinical usefulness of a predictive score.

We could not include left ventricular ejection fraction as an explanatory variable, due to many missing observations and the fact that the remaining measurements were performed using two different methods (catheterization or echocardiography). These methods yield somewhat different result and thus cannot be pooled.

A model based on a multi-centre database could potentially be more generally applicable than a single-centre model. However, this approach would not necessarily overcome the problems related to ICU stay as a subjective outcome.

CONCLUSIONS

Our data indicate that it may be difficult to develop a universal model for prediction of ICU stay, as the distribution of stay

durations may depend both on medical factors and institutional policies for ICU discharge. Models for prediction of prolonged mechanical ventilation may provide better approximations for a prolonged ICU stay than mortality models. However, good discrimination does not necessarily translate to good calibration or clinical usefulness of any model.

FUNDING

Yunita Widyastuti was supported by a scholarship from the Indonesian government.

Conflict of interest: none declared.

REFERENCES

- [1] Pintor PP, Colangelo S, Bobbio M. Evolution of case-mix in heart surgery: from mortality risk to complication risk. *Eur J Cardiothorac Surg* 2002;22:927–33.
- [2] Bridgewater B, Gummert J, Kinsman R, Walton P. *Towards global benchmarking: the Fourth EACTS Adult Cardiac Surgical Database Report 2010*. ISBN 1-903968-26-7. Henley-on-Thames: Dendrite Clinical Systems, 2010.
- [3] Lagercrantz E, Lindblom D, Sartipy U. Survival and quality of life in cardiac surgery patients with prolonged intensive care. *Ann Thorac Surg* 2010;89:490–95.
- [4] Arabi Y, Venkatesh S, Haddad S, Al Shimemeri A, Al Malik S. A prospective study of prolonged stay in the intensive care unit: predictors and impact on resource utilization. *Int J Qual Health Care* 2002;14:403–10.
- [5] Doering LV, Esmailian F, Laks H. Perioperative predictors of ICU and hospital costs in coronary artery bypass graft surgery. *Chest* 2000;118:736–43.
- [6] Austin PC, Rothwell DM, Tu JV. A comparison of statistical modeling strategies for analyzing length of stay after CABG surgery. *Health Serv Outcomes Res Methodol* 2002;3:107–33.
- [7] De Cocker J, Messaoudi N, Stockman BA, Bossaert LL, Rodrigus IE. Preoperative prediction of intensive care unit stay following cardiac surgery. *Eur J Cardiothorac Surg* 2011;39:60–7.
- [8] Ghotkar SV, Grayson AD, Fabri BM, Dihmis WC, Pullan DM. Preoperative calculation of risk for prolonged intensive care unit stay following coronary artery bypass grafting. *J Cardiothorac Surg* 2006;31:1–14.
- [9] Rosenfeld R, Smith JM, Woods SE, Engel AM. Predictors and outcomes of extended intensive care unit length of stay in patients undergoing coronary artery bypass graft surgery. *J Card Surg* 2006;21:146–50.
- [10] Janssen DP, Noyez L, Wouters C, Brouwer RM. Preoperative prediction of prolonged stay in the intensive care unit for coronary bypass surgery. *Eur J Cardiothorac Surg* 2004;25:203–7.
- [11] Atoui R, Ma F, Langlois Y, Morin JF. Risk factors for prolonged stay in the intensive care unit and on the ward after cardiac surgery. *J Card Surg* 2008;23:99–106.
- [12] Messaoudi N, De Cocker J, Stockman BA, Bossaert LL, Rodrigus IE. Is EuroSCORE useful in the prediction of extended intensive care unit stay after cardiac surgery? *Eur J Cardiothorac Surg* 2009;36:35–9.
- [13] Nilsson J, Algotsson L, Högglund P, Luhrs C, Brandt J. EuroSCORE predicts intensive care unit stay and costs of open heart surgery. *Ann Thorac Surg* 2004;78:1528–34.
- [14] Pitkänen O, Niskanen M, Rehnberg S, Hippeläinen M, Hynynen M. Intra-institutional prediction of outcome after cardiac surgery: comparison between a locally derived model and the EuroSCORE. *Eur J Cardiothorac Surg* 2000;18:703–10.
- [15] www.euroscore.org. euroSCORE II. Lisbon, 2011 (accessed 10/10/2011).
- [16] Berg KS, Stenseth R, Pleym H, Wahba A, Videm V. Mortality risk prediction in cardiac surgery: comparing a novel model with the EuroSCORE. *Acta Anaesthesiol Scand* 2011;55:313–21.
- [17] Widyastuti Y, Stenseth R, Pleym H, Wahba A, Videm V. Pre-operative and intraoperative determinants for prolonged ventilation following adult cardiac surgery. *Acta Anaesthesiol Scand* 2012;56:190–99.
- [18] Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
- [19] Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer, 2001.
- [20] Lynn J, Teno JM, Harrell FE Jr. Accurate prognostications of death. Opportunities and challenges for clinicians. *West J Med* 1995;163:250–57.
- [21] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach. *Biometrics* 1988;44:837–45.
- [22] Ettema RG, Peelen LM, Schuurmans MJ, Nierich AP, Kalkman CJ, Moons KG. Prediction models for prolonged intensive care unit stay after cardiac surgery: systematic review and validation study. *Circulation* 2010;122:682–89.
- [23] Widyastuti Y, Stenseth R, Berg KS, Pleym H, Wahba A, Videm V. Preoperative and intraoperative prediction of risk of cardiac dysfunction following open heart surgery. *Eur J Anaesthesiol* 2012;29:143–51.
- [24] Nakasuji M, Matsushita M, Asada A. Risk factors for prolonged ICU stay in patients following coronary artery bypass grafting with a long duration of cardiopulmonary bypass. *J Anaesth* 2005;19:118–23.
- [25] Higgins TL, McGee WT, Steingrub JS, Rapoport J, Lemeshow S, Teres D. Early indicators of prolonged intensive care unit stay: impact of illness severity, physician staffing, and pre-intensive care unit length of stay. *Crit Care Med* 2003;31:45–51.

eComment. Organism functionality parameters in predicting the length of intensive care unit stays?

Authors: Juan Bustamante¹, Sergio Canovas², J. Aurelio Sarralde³ and Eduardo Tamayo⁴

¹Department of Cardiovascular Surgery, Hospital Universitario La Princesa, Madrid, Spain

²Department of Cardiac Surgery, Hospital General Universitario de Valencia, Valencia, Spain

³Department of Cardiovascular Surgery, Hospital Universitario Valdecilla, Santander, Spain

⁴Department of Anaesthesiology and Reanimation, Hospital Clinico Universitario de Valladolid, Valladolid, Spain

doi:10.1093/icvts/ivs434

© The Author 2012. Published by Oxford University Press on behalf of the European Association for Cardio-Thoracic Surgery. All rights reserved.

We read the article by Widyastuti *et al.* [1] with great interest and it is true that cardiac surgery results have improved in accordance with improvements in cardiac techniques and perioperative care. This has led to the fact that the current profile of patients undergoing cardiac surgery is becoming an increasingly more elderly one with greater co-morbidity, thus implying an increase in surgical risk and, likewise, an increase in in-hospital stay, whether in the intensive care unit (ICU) or in total [2].

The currently most-used risk prediction scales have certain limitations and this means that, for specific groups of patients, they are not capable of a correct risk fit [3]. Moreover, it is difficult to know what the evolution of a patient who has been operated on will be, due to the large number of variables that may exert an influence on this aspect, and, therefore, how long the length of ICU stay will be.

The variables that measure the organism functionality parameters once the surgical procedure is over could be of great use in estimating subsequent evolution in conjunction with the preoperative risk scales. Many authors, and even various scales, have highlighted the importance of bearing in mind the parameters related with the inflammatory response and haemodynamic state of the patient, such as: hyperlactataemia, bicarbonate, heart rate, lactate and creatinine upon admittance to an ICU in order to establish predictive models that increase the precision of such an estimation [4, 5].

Another important aspect in analyzing immediate postoperative evolution is evaluating the development of complications during the surgical procedure (e.g., bleeding, acute myocardial infarction, inadequate myocardial protection, etc). These are going to be reflected in a series of metabolic alterations like: creatinine, haematocrit, troponin T (TnT), pH, bicarbonate (HCO₃⁻) or the ratio of partial pressure of arterial oxygen to the fraction of inspired oxygen (PaO₂/FiO₂), etc.

With regard to the intraoperative variables to which the authors refer, these are fundamentally centred on bleeding and the need for haemoderivatives [1]. Such variables may give rise to multi-collinear phenomena and it would therefore be advisable to carry out a check with an analysis of tolerance in order to determine